MANAGING SIMULTANEOUS MOBILITY OF IP HOSTS

K. Daniel Wong, Ashutosh Dutta, Ken Young
Telcordia Technologies
Red Bank and Morristown, NJ
and
Henning Schulzrinne
Columbia University
New York, NY

ABSTRACT

Since triangular routing in Mobile IP (MIP) is undesirable, MIP with Route Optimization (MIP-RO) and MIP with Location Registers (MIP-LR), all use binding updates that are sent directly to a Correspondent Host. Session Initiation Protocol (SIP) based mobility management also uses direct binding updates between a Mobile Host and a Correspondent Host. However, this makes these protocols (except the basic MIP) vulnerable to the simultaneous mobility problem, i.e. the special case when both end hosts are mobile and move at about the same time. In this paper, we analyze the simultaneous mobility problem and propose new ways for MIP-LR and SIP to handle simultaneous mobility using a common approach stemming from a generalized solution.

I. INTRODUCTION

IP mobility management protocols based on Mobile IP with Location Registers (MIP-LR [1]) and Session Initiation Protocol (SIP [2]) have certain benefits for quasi-static ad hoc networks. We have previously shown some schemes [3][4] that can provide more survivable, robust and efficient solutions than the traditional Mobile IP (MIP [5]) scheme. These properties make these schemes very attractive for tactical military communications networks and other networks where robustness and survivability are critical. However, SIP and MIP-LR currently do not handle simultaneous mobility well.

Simultaneous mobility is the special case when two communicating end hosts are mobile and both move at about the same time. Although it may not be typical (more typical would be the case when one of the two hosts moves and the other remains stationary during that time), it would happen once in a while and must be handled properly by mobility protocols. We define the *simultaneous mobility problem* to be the problem of losing a binding update from one Mobile Host because it is sent to a previous address of the other Mobile Host that is also moving at around the same time. Note that the disruption caused by the simultaneous mobility problem goes beyond the typical disruption caused by non-simultaneous mobility. Older protocols like MIP handle simultaneous mobility adequately, because of non-mobile home agents. Therefore, in this paper, we analyze the simultaneous mobility problem for SIP-based and MIP-LR based mobility management schemes, and propose solutions using a common approach. Our solutions are designed to impose minimal changes on the existing protocols while efficiently dealing with the simultaneous mobility problems.

As a matter of scope, in this paper the kind of mobility of interest is terminal mobility (rather than other notions of mobility like personal mobility, service mobility, etc.) of end hosts (it may be assumed that router mobility would be handled by other means, e.g. ad hoc routing protocols in conjunction with auto-configuration protocols). Both pre-session and mid-session mobility are considered. We focus on layer 3 handoffs, i.e. where IP address changes are involved.

We focus on situations where the handoff rate of a Mobile Host is "typical" enough that consecutive handoffs *of the same Mobile Host* are non-overlapping. We do not focus on situations of overlapping consecutive handoffs of the same Mobile Host, where one handoff has not completely finished before the next begins, e.g. there has not been enough time after the acquisition of an IP address for binding updates to reach their destination networks. The reasons for our focus:

- The problems encountered with overlapping consecutive handoffs are not so much a problem of simultaneous mobility as a problem with excessive handoff rate. Whether the Correspondent Host is mobile or fixed, there will be very severe problems when the Mobile Host changes its IP address before binding updates for its previous IP address have even arrived at their destinations.
- For the foreseeable future, the extreme case of handoffs rates high enough for overlapping consecutive handoffs is highly improbable.

Hence, as shown in Figure 5, consecutive handoffs of a Mobile Host are assumed to be non-overlapping , and we focus rather on overlap of handoffs of *different* Mobile Hosts, i.e. simultaneous mobility.

As far as we know, the existing literature does not provide analysis of simultaneous mobility when MIP-related protocols or SIP are used. Reference [7] extends a TCP migration mobility protocol to handle simultaneous mobility, but there are significant differences between the TCP migration schemes (where mobility is handled at the transport layer) and MIP-related or SIP. We are not aware of solutions to the simultaneous mobility problem for SIP and MIP-LR having been proposed or analyzed before, although Reference [8] proposes techniques that could be useful for dealing with simultaneous mobility (the main subject of Reference [8], though, is fast handoffs for SIP mobility). This paper is organized as follows. We briefly describe the protocols in Section II, analyze the problems with simultaneous mobility in Section III, and propose solutions in Section IV. Section V concludes the paper with some discussion.

II. THE PROTOCOLS

Mobile IP (MIP [5]) enables Mobile Hosts to retain IP connectivity when roaming in foreign networks, while still using their permanent IP address for identification. When roaming, Mobile Hosts acquire temporary care-of-addresses for routing purposes. MIP requires registration messages to be exchanged with a Home Agent on a Mobile Host's home network whenever the Mobile Host moves between subnets in foreign networks, to

This material is based upon work supported by DARPA

ensure that the Home Agent can tunnel packets destined for the Mobile Host to the right foreign subnet.

With MIP, all packets sent to the Mobile Host have to pass through the Home Agent, making the Home Agent a single point of failure. This is a serious problem for networks that require high robustness and survivability. MIP-LR provides an efficient approach compared to MIP by using replicated databases (Home Location Registers) that replace Home Agents. In MIP-LR a Correspondent Host queries a Mobile Host's Home Location Register (HLR) for the location of the Mobile Host, and then can send packets directly to the foreign network. For efficiency, we have an implementation of MIP-LR that can send packet un-encapsulated to the foreign network [6]. Unlike the Home Agent, an HLR need not necessarily be located in the home network, and it can be replicated for survivability. When the Mobile Host moves, it sends a binding update to its HLR, as well as directly to all known Correspondent Hosts. MIP-LR mobility is illustrated in Figure 1(b).

SIP is designed to manage real-time sessions, e.g. packet-switched voice and video sessions. Mobility of the end hosts is handled very naturally by SIP, using existing SIP signaling mechanisms [2]. For example, to initiate a session, the SIP INVITE message is sent by the initiating party to the other party. The extension of SIP to handle mid-session mobility specifies that when one of the two parties moves, it sends a re-INVITE to the other party, informing it of its new location (e.g. its new IP address). Figure 1(a) shows this signaling for SIP mobility. In addition to the re-INVITE sent directly to the Correspondent Host, the Mobile Host also registers its presence in the new network with a SIP server in its home network. This allows other potential Correspondent Hosts to find the Mobile Host.

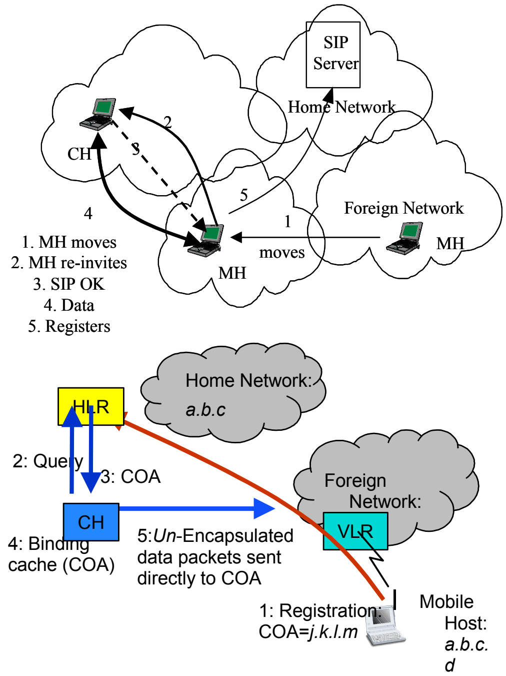

Figure 1: (a) SIP Mobility; (b) MIP-LR Mobility

III. PROBLEMS OF SIMULTANEOUS MOBILITY

In the literature, the terms "pre-session mobility" and "mid-session mobility" have meaning only for SIP-based mobility management, not MIP or its variants. In this paper, we extend the terms to the context of MIP and its variants as follows: pre-session mobility is mobility that happens when the Mobile Host is unaware of any active Correspondent Hosts and has no active bindings for other Mobile Hosts, while mid-session mobility is any mobility that is not pre-session. The reason these definitions are useful is because simultaneous mobility problem deals with sending of binding updates directly between Mobile Hosts. We therefore rule out problems with simultaneous mobility during pre-session mobility for all the mobility schemes of interest, since there is no Correspondent Host for the Mobile Host to update or be updated by. However, we will discuss an exception in Section B, where simultaneous mobility *during* session initiation may cause problems.

A. Mid-Session Scenarios

MIP does not have a problem with simultaneous mobility. By design, Correspondent Hosts are unaware of the mobility of Mobile Hosts. The Mobile Host's Home Agent functions as an anchor point for the Mobile Host. No matter where the Mobile Host moves, packets for it always go first to its home network for interception and tunneling by its Home Agent. If it turns out that the Correspondent Host is also mobile, it will also have a Home Agent, and packets from the Mobile Host will similarly be intercepted and tunneled to the appropriate network by its Home Agent. Since both Home Agents are stationary and can always be reached through IP routing, simultaneous mobility does not present a problem to MIP.

Two of the most important enhancements to the basic MIP are Mobile IP with Route Optimization (MIP-RO [9]) and Mobile IPv6 (MIPv6 [10]). MIP-RO uses binding updates transmitted from a Mobile Host's Home Agent to Correspondent Hosts, to allow direct routing of packets from Correspondent Host to Mobile Host. In addition, it includes the capability for the previous network to forward packets (all packets, including binding updates) to the new network. As this is one of the solution approaches for the simultaneous mobility problem, we defer further discussion on MIP-RO until Section IV.B. MIPv6 also has binding updates sent directly to Correspondent Hosts for route optimization, but these come from the Mobile Host, rather than the Home Agent as in MIP-RO. MIPv6 is currently still work in progress, so it is unclear any problems it may have with simultaneous mobility.

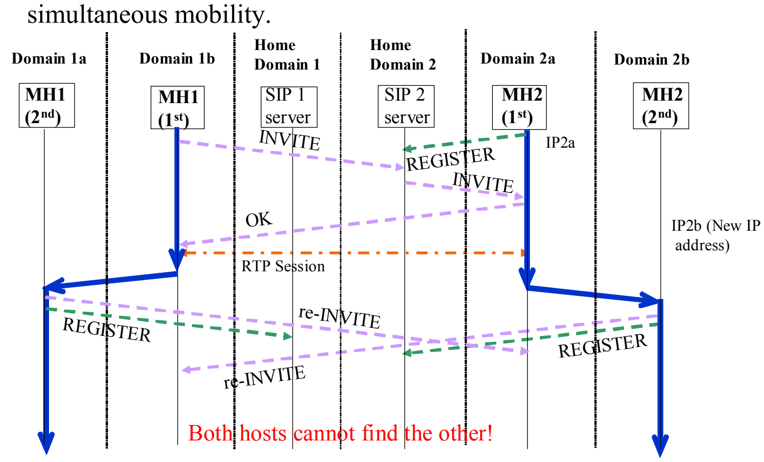

Figure 2: Mishandling of Simultaneous Mobility by SIP

In the basic SIP mobility scheme, when a Mobile Host moves to a new network, it sends a re-INVITE message to the Correspondent Host, as well as a REGISTER message to its home network SIP server. There are two options for the path of the re-INVITE, i.e. it could go through the inbound proxy of the Correspondent Host, or it could go directly to the Correspondent Host. This second option suffers from major problems with simultaneous mobility. Clearly, if the Correspondent Host moves at the same time, the re-INVITE may be lost (and similarly, the re-INVITE from the Correspondent Host to the Mobile Host could also be lost). As the home network SIP servers for both hosts are stationary and always reachable through IP routing, the *registrations* aren't affected by the simultaneous mobility. One might suppose that both hosts could now obtain the new location of the other party through the SIP servers, analogous to how the Home Agent provides such information in the MIP, MIP-RO and MIPv6 cases. However, in the MIP cases, the Home Agent quickly discovers that the Correspondent Host needs the updated binding information, because data packets from the Correspondent Host are routed to the home network and intercepted by the Home Agent. In the SIP case, on the hand, the Correspondent Host may not immediately contact the SIP server. Instead, the re-INVITE may time-out, and may be tried again several more times to the wrong network by virtue of its built-in retransmission mechanism. The crucial difference is that the data path and signaling path are separate in the case of SIP, unlike that of MIP. The problem is shown in Figure 2. For simplicity, automatic retransmissions of lost re-INVITE messages are not shown, and neither are messages like the ACK that should acknowledge the OK.

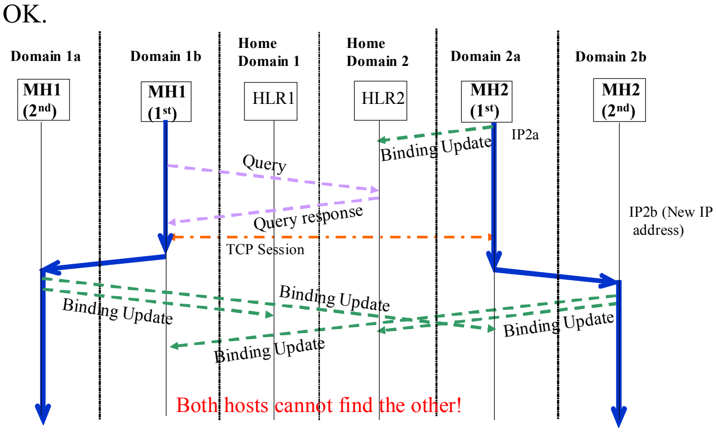

Figure 3: Mishandling of Simultaneous Mobility by MIP-LR

The problems that MIP-LR faces with simultaneous mobility are analogous to the problems faced by SIP with simultaneous mobility when SIP re-INVITEs are sent directly to Correspondent Hosts, and it is shown in Figure 3. When a Mobile Host moves to a new network, it sends a MIP-LR update message to the Correspondent Host, as well as a MIP-LR update message to its Home Location Register (HLR). Clearly, if the Correspondent Host moves at the same time, the direct MIP-LR update may be lost (and similarly, the direct MIP-LR update from the Correspondent Host to the Mobile Host could also be lost). Since the HLRs for both hosts are stationary and always reachable through IP routing, the registrations should not be affected by the simultaneous mobility. However, like in the SIP case, the Correspondent Host may not immediately query the HLR and both hosts will be using out-dated IP addresses for the other. From the similarity of the problems faced by SIP and MIP-LR, it might be expected that a common approach could be taken to solve their problems.

B. *Special Cases*

Simultaneous mobility of two end hosts is not an issue for pre-session mobility. However, during a transition before a session setup is complete, simultaneous mobility may present problems. As shown in Figure 5, it may so happen that one of the signaling messages (INVITE, OK, ACK) does not reach the other party and gets lost, despite the SIP server keeping an up-to-date registration for both parties. Figure 4 shows an example where the call does not get completed because of the timing of the movement by the CH and MH.

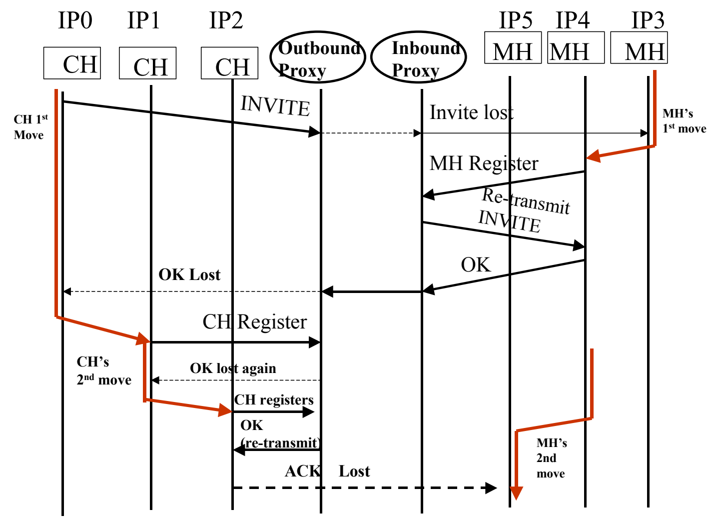

Figure 4: Call Flow of Call fail scenario

C. *Analyzing the Simultaneous Mobility Problem*

To analyze the likelihood of occurrence of the simultaneous mobility problem, we consider a scenario with two Mobile Hosts, A and B. We break down what happens in a typical handoff of host A:

- the Mobile Host selects a new network base station or access point
- the Mobile Host begins to reconfigure the wireless interface, and at some point in time, *e,* cannot be reached any longer at the old IP address
- after a short time γ, the Mobile Host has finished reconfiguring its wireless interface, and is reachable at its new IP address
- connectivity from B to A is restored, as a binding update reaches B directly (or reaches something like a Home Agent) from A

During this handoff, what is the time interval during which host A is vulnerable to losing binding updates that are sent to it by B, i.e. vulnerable to the simultaneous mobility problem? Although it may at first appear to be only vulnerable for time γ, the period of vulnerability may in fact be significantly longer. Suppose it takes the binding update to reach from B to A, α units of time, and from time *e* it takes another length of time, β, for A's binding update to reach B. Then if B changes IP address and transmits its binding update any time from e-α to e+β, the binding update will be addressed to A's old IP address and therefore be lost. Therefore, the vulnerability interval is in fact $\alpha+\beta$. Figure 5 illustrates the point. The subscripts *k-1, k, k+1* are an index to the handoffs made by A.

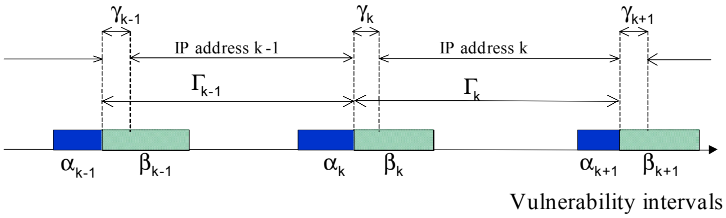

Figure 5: Intervals in which Mobile Host is vulnerable to lost binding updates

The inter-handoff intervals, Γ_{k-1}, Γ_k, Γ_{k+1}, can be modeled as Poisson distributed with mean λ. Let the probability that any particular handoff suffers from the simultaneous mobility problem be P_0, and the probability that at least one out of the N handoffs in a given session (e.g. voice session between two SIP end hosts), suffers from the simultaneous mobility problem be P_N. Then $P_N = 1-(1-P_0)^N$ and

$$P_0 \approx \frac{E[\alpha+\beta]}{\lambda}$$

where E[ ] denotes expected value.

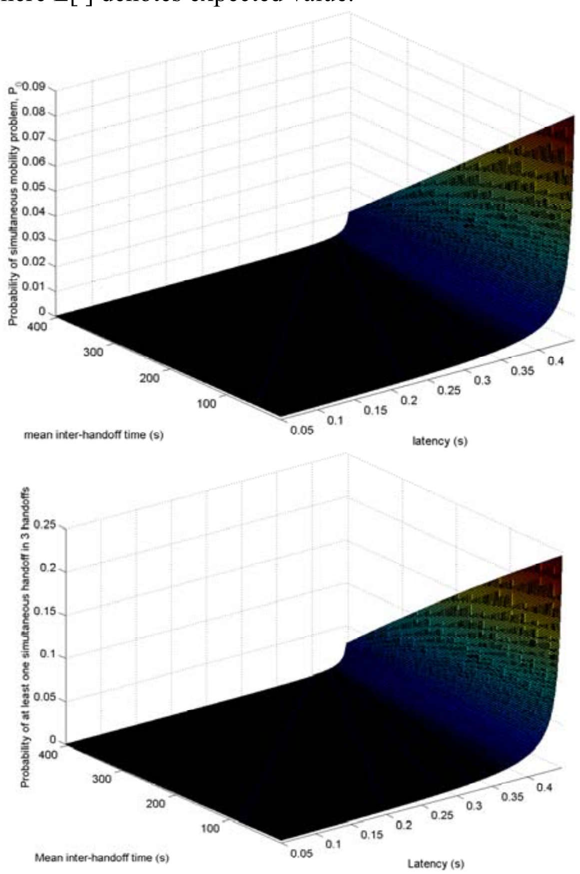

Figure 6: (a) Plot of P_0 against latency and mean handoff time (b) plot of P_3

Based on measurements, $E[\alpha+\beta]$ ranges from 50 ms to 500 ms, while λ may range from 5 s (movement at vehicular speeds across pico-cells of a few hundred meters in diameter) to 500 s or more (larger cells, slower speeds, non-linear movement pattern). We plot for approximately this range of $E[\alpha+\beta]$ and λ in Figure 6. As expected, the highest probability of simultaneous mobility is when latency is largest and average inter-handoff time smallest.

The effect of the simultaneous mobility problem could be quite significant. Without fixing the problem, the binding updates of both Mobile Hosts would never reach the other host, and so the connection would be lost.

For the case of simultaneous mobility during session initiation signaling, as discussed in Section III.B, the probability of failure also depends upon the mobility rate of the mobiles. From our lab measurements, it takes about 200-300 ms to complete the whole session initiation signaling sequence. A complete registration will take about 150 ms.. Hence, the probability of simultaneous mobility occurring during session initiation signaling is non-trivial.

IV. SOLUTIONS

We have explained, in Section III.A, how MIP-LR and SIP sending direct re-INVITEs (i.e. not through SIP proxies) perform poorly in the face of simultaneous mobility. Therefore, the solutions must involve stationary network elements, e.g. making signaling go through SIP servers, and extending the functionality of MIP-LR HLRs. Possible solutions are examined in Sections IV.A, IV.B, and IV.C. Each solution has drawbacks, so we propose a new class of solutions to the mid-session simultaneous mobility problem, in Section IV.D, and outline a solution for simultaneous mobility during session initiation signaling, in Section IV.E.

A. Use of Mobile-Originated Lost Message Retransmission

SIP has an in-built retransmission capability, where messages are retransmitted after a time-out if acknowledgement is not received. During mid-session mobility, a re-INVITE may get lost even if it goes through the SIP server that keeps the most recent registration status of the destination. However, SIP allows for automatic retransmissions of INVITEs (including re-INVITEs) by SIP UAs if a response (OK message) is not received within a specified time. Stateful SIP servers could also retransmit (re-)INVITEs, as seen in Figure 7 (the RTP translators will be soon explained).

One problem with timer-based retransmissions is that significant latency could be added to the handoff when messages are lost, as when simultaneous mobility occurs.

B. Use of Forwarding Mechanisms from Previous Network

With MIP-RO, to fix the triangular routing problem of MIP, packets from the Correspondent Host to the Mobile Host may bypass the Home Agent. When a Home Agent receives packets for a Mobile Host, it sends a Binding Update to the Correspondent Host (the source of the packets). This allows the Correspondent Host to send packets directly to the Mobile Host. After a Mobile Host moves, Correspondent Hosts will continue to send packets to its previous care-of address. However, the Foreign Agent in the previous network will do two things to ensure that the Correspondent Host is updated and the packets are not lost. Firstly, it will send a Binding Warning message to the Home Agent so the Home Agent can issue a Binding Update to the Correspondent Host. Secondly, it will forward packets to the new Foreign Agent, assuming the Mobile Host has used the Previous Foreign Agent Notification extension in its Registration in the new network, and the new Foreign Agent has updated the previous Foreign Agent as requested.

In SIP mobility, one tool for handling the simultaneous mobility problem is that node in the previous network, like an RTP (Real-time Transport Protocol) translator [8], can help redirect the traffic to the new address, analogous to how a foreign agent in the previous network redirects traffic in MIP-RO. Figure 7 shows an example, where the retransmissions allow the signaling to eventually be successful, while the RTP translator reduces the disruption in the data traffic flow while waiting for the retransmissions to complete. Note that the RTP translator only forwards data traffic from the previous network, so we propose (in Section IV.D) a forwarding element for SIP signaling as well.

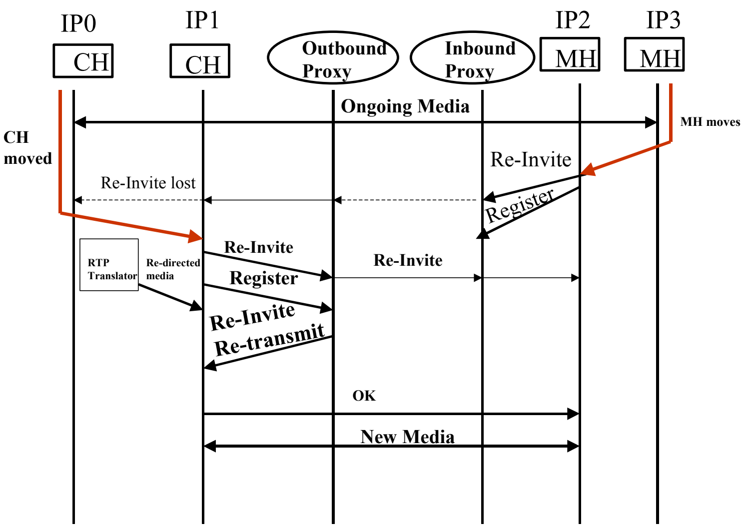

Figure 7: Successful Mid-session Mobility Flow with server-assisted re-transmission

C. *Use of Simultaneous Bindings*

Suppose a Mobile Host can have more than one valid IP address. This is sometimes referred to as "simultaneous mobility bindings", and should not be confused with the simultaneous mobility problem. In particular, if the previous IP address and new IP address can both be used to reach the Mobile Host during the time around a handoff, that may help solve the simultaneous mobility problem. Binding updates sent to the previous IP address would arrive correctly. However, this is not a universally applicable solution, for the following reasons:

- Simultaneous mobility bindings may not be supported by the Home Agent in MIP
- The radio network technology needs to be able to support multiple concurrent IP addresses for the wireless interface(s). It is too much to require this for solving simultaneous mobility.

Since our solution must work for any radio network technology, use of simultaneous bindings is not a satisfactory solution.

D. *Proposed Solutions*

We observe that binding updates are lost in a simultaneous mobility situation because senders do not have the most up-to-date binding for the destination. Therefore, we introduce two abstract concepts: stationary binding update proxies and stationary location proxies. Stationary binding update proxies act on behalf of a Mobile Host to ensure that binding updates for the Mobile Host's current configuration are sent to their destinations properly. Stationary location proxies, on the other hand, act on behalf of a Mobile Host as repositories of the latest configuration, and can be queried for such information. Both these proxies have the following properties:

- They are stationary, i.e. not mobile. Otherwise, if they move at the wrong times, something akin to the simultaneous mobility problem returns.
- They are abstract, functional elements, and could be implemented in the same host (forming a stable rendezvous point), and even in existing network elements, but don't have to be. Our design principle relies on re-using the existing functionality as much as possible, thus we try to implement the required functionality with a minimum of changes to the existing protocols and network elements.

We now introduce a generalized solution using the stationary location proxy and stationary binding update proxy. Immediately after a handoff, the Mobile Host sends a binding update to its stationary location proxy. It then sends a list of Correspondent Hosts to its stationary binding update proxy, together with the new IP address, after each handoff. The stationary binding update proxy is then responsible to obtain the latest address of the Correspondent Host from the Correspondent Host's stationary location proxy, and to initiate binding updates to the correct address. To reduce update latency, direct binding updates should also be sent by the Mobile Host to the last known address of each Correspondent Host. To make the simultaneous mobility solution even more robust, we introduce a third abstract function, an Interceptor in the previous network from which the Mobile Host just moved. The function of the Interceptor is to intercept packets destined to the Mobile Host that arrive in the network from which it just moved, and forward them to the new network.

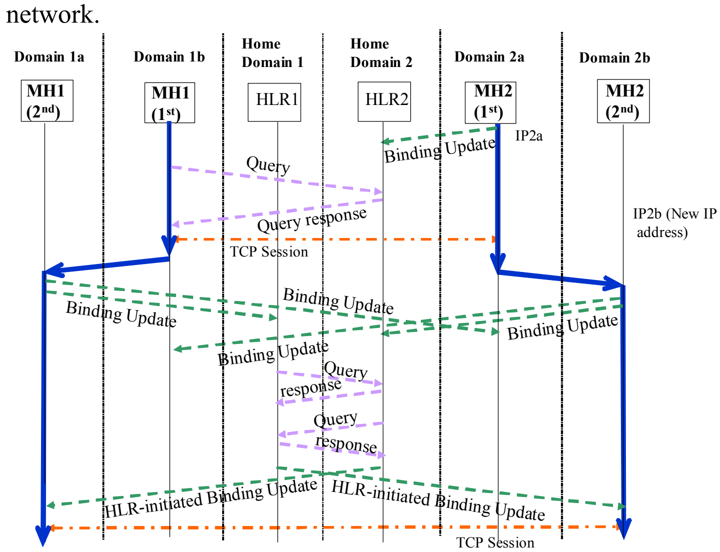

Figure 8: MIP-LR with HLR-initiated Binding Updates

We illustrate this solution using MIP-LR as an example. As a matter of principle, specific to our solution for SIP and MIP-LR, we try to re-use existing network elements and functionality as much as possible, augmented as necessary. In MIP-LR, the HLR already performs the role of stationary location proxy, and may be enhanced to also act as stationary binding update proxy, since it anyway obtains the current binding information as part of MIP-LR updating after each handoff. The scenario from Figure 3 is repeated in Figure 8, except that now the binding update sent by a Mobile Host to its HLR has a list of Correspondent Hosts, and the HLRs are enhanced as discussed. Note that the HLR-initiated binding updates are new. Sending these is part of the stationary binding update proxy functionality added to the HLR. Although the direct binding updates are lost, as in Figure 3, the HLR-initiated binding updates arrive at the correct destinations. In a special case, a HLR queries another HLR for the location of a Mobile Host just before the binding update arrives. In this case, it is desirable that the queried HLR send a revised response without the querying HLR having to query again. Therefore, the HLRs should treat these HLR-to-HLR queries differently from normal queries, and be prepared to revise their response, if updated bindings are available, for some period of time (e.g. 1 s) after receiving a query.

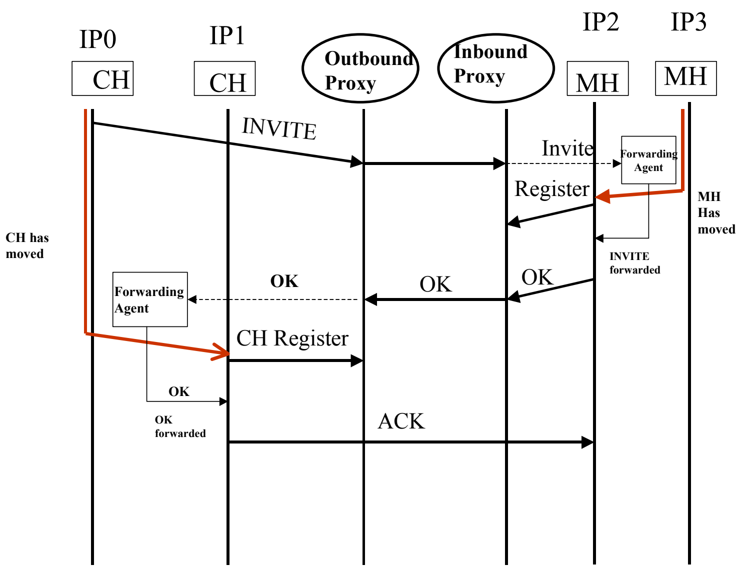

Figure 9: successful completion of session initiation signaling

For SIP mobility, the home network SIP server already has stationary location proxy functionality, and can take on stationary binding update proxy responsibilities for incoming re-INVITEs. One solution is to have the SIP server immediately re-transmit the re-INVITE upon receiving a REGISTER message from the destination of a pending re-INVITE (upon receiving the REGISTER message, it knows that the previous re-INVITE it sent may be lost), rather than waiting for a time-out. Thus, it addresses one criticism of using timer-based retransmissions as discussed in Section IV.A. Other than this pro-active retransmission, the solution is similar to what is shown in Figure 7. An alternative solution is to just use retransmissions based on timers, but to introduce a new network function (Interceptor function) in the previous network to intercept signaling and send to the correct location of the Mobile Host. This Interceptor function would be parallel to the RTP translator function of forwarding RTP (data) traffic. The flow is similar to that in Figure 9, except that re-INVITE is sent, instead of INVITE. Further study is required to examine how to control the Interceptor so that it can forward the UDP control packets.

E. *Handling Simultaneous Mobility during Session Initiation*

It turns out that our 2nd solution for the mid-session simultaneous mobility case (using the Interceptor function) also works for simultaneous mobility during session initiation. Figure 9 shows the solution.

V. DISCUSSION AND CONCLUSIONS

In this paper, we have identified the problems related to simultaneous mobility for SIP based mobility and MIP-LR schemes. Basically, the separation of data traffic paths from signaling paths in both schemes results in a problem that MIP manages to avoid. However, SIP and MIP-LR have advantages over MIP, especially in networks where survivability and robustness are critical, such as tactical military ad hoc network environments. Therefore, we have analyzed the problems associated with SIP and MIP-LR based mobility scheme and present viable solutions.

Since the simultaneous mobility problem could cause serious problems like dropped sessions, the proposed solutions may be considered and implemented in a scenario where two communicating hosts are mobile.

VI. REFERENCES

[1] R. Jain, T. Raleigh, D. Yang, L.F. Chang, C. Graff, M. Bereschinsky and M. Patel, “Enhancing Survivability of Mobile Internet Access using Mobile IP with Location Registers,” *IEEE Infocom 1999,* March 1999.

[2] 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.

[3] A. Dutta, K.D. Wong, J. Burns, R. Jain, A. McAuley, K. Young and H. Schulzrinne, “Realization of Integrated Mobility Management Protocol for Ad-Hoc Networks,” *IEEE Milcom 2002*, Anaheim, CA, USA, October 2002.

[4] K.D. Wong, A. Dutta, J. Burns, R. Jain, K. Young and H. Schulzrinne, “A Multi-layered Mobility Management Scheme for Auto-configured Wireless Networks,” submitted to *IEEE Wireless Communications Magazine*, January 2003.

[5] C. Perkins et al., “IP Mobility Support for IPv4,” *IETF RFC 3344,* August 2002.

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

[7] S. Tilak and N. B. Abu-Ghazaleh, “A Concurrent Migration Extension to an End-to-End Mobility Architecture", *Mobile Computing and Communication Review*, Volume 5, Number 3, July 2001.

[8] A. Dutta, S. Madhani, W. Chen, O. Altintas and H. Schulzrinne, “Optimized Fast-handoff Scheme for Application Layer Mobility Management,” *ACM International Conference on Mobile Computing (and Mobile Computing and Communication Review),* November 2002.

[9] C. Perkins and D.B. Johnson, “Route Optimization in Mobile IP”, *IETF* work in progress draft-ietf-mobileip-optim-11.txt, September 2001.

[10] D. Johnson, et al., “Mobility Support in IPv6,” *IETF* work in progress draft-ietf-mobileip-ipv6-21.txt, February 2003.