Mobile IP Telephony: Mobility Support of SIP

Melody Moh[1] Gregorie Berquin Yanjun Chen

Moh@mathcs.sjsu.edu berquin_g@hotmail.com chen1175@hotmail.com

Department of Mathematics and Computer Science

San Jose State University

San Jose, CA 95192-0103

***Abstract—The* Internet has recently become the most important, most popular way of communication. A significant new feature of Internet is the support of telephony. Two main competing signaling protocol standards have been developed for this purpose: H.323, proposed by the ITU, and SIP, proposed by the IETF. Another new feature for the Internet is the support of terminal mobility. The IETF efforts in this field have resulted in the Mobile IP standard. Recent, Liao has shown that Mobile IP doesn't provide fast enough handoffs to support voice communications. Liao has also demonstrated how H.323 may be extended to offer a solution. In this paper we address several major issues for supporting mobility on SIP. We detail the mechanisms for extending SIP to support location management, registration of roaming mobiles, and handoffs. The mechanisms are either extensions of existing SIP, or based on Mobile IP, and thus may be readily employed over the Internet. We believe that the work presented here is an important step towards supporting mobile telephony over the Internet.**

1. INTRODUCTION

Both the telephony and Internet businesses are coming to a new era. The marriage of the two, Internet telephony, is receiving increasing interest. Meanwhile, host mobility is also becoming important because of the recent blossoming of laptop computers and the high desire to have continuous network connectivity anywhere the host happens to be. It is desirable to integrate IP host mobility features to Internet telephony so that the IP host currently on the phone conversation can be on the move from network to network.

Internet telephony, also known as voice over IP (VoIP) or IP telephony, is the process of transmitting voice information, which is traditionally transmitted over public switched telephone network (PSTN), over packet networks – the Internet. It unites the telephony and data worlds, permitting phone calls, fax calls, and voice traffic to be transmitted over corporate enterprise networks, Intranets, and the Internet.

Internet telephony requires the communication partners to find each other and to signal to the other party they desire to communicate. This is defined as Internet telephony signaling, which establishes calls, manages calls, and exchanges endpoint capabilities in the call level. Currently there are two competing protocols emerged for signaling and controls for Internet telephony. One is ITU Recommendation H.323 [18], the other is the IETF Session Initiation Protocol (SIP) [5]. H.323 embraces the more traditional circuit-switched approach to signaling based on the ISDN Q.931 protocol and earlier H-series recommendations, and SIP favors the more lightweight Internet approach based on HTTP [15]. While both protocols are similar in terms of providing a rich set of functionality and services, neither of them supports IP host mobility in their current version. In order to provide mobile Internet telephony service, new approaches are needed. There has been some work done by Liao in mobile extensions to H.323 [6, 7].

In this paper, we describe several approaches for extending SIP to provide mobility support, including registration for roaming users, location management, and handoff handling. This is a continuation of our on-going work in wireless/mobile networks [8, 9, 10].

In the rest of this paper, we provide some related background knowledge on SIP and mobile IP in Section 2. Section 3 is devoted to our solutions to mobility support in SIP. Finally we give some conclusion remarks in Section 4.

2. BACKGROUND

To better understand how mobility extensions can be added to SIP, we will first briefly explain the basic SIP protocol functionality and operation. More specifically, we examine advanced personal mobility services and multi-party call services supported by SIP. Essential parts of the Mobile IP mechanism are also introduced. All these are closely related to the mobility support of SIP that will

[1] Supported in part by NSF (grant NCR-9714700), 3Com Technology Development Center, and Laverage Design Acceleration Co.

be discussed in the next section.

2.1 Overview of SIP

As described in SIP, callers and callees are identified by SIP addresses. When making a SIP call, a caller first locates the appropriate server and then sends a SIP request. The most common SIP operation is the INVITATION. Instead of directly reaching the intended callee, a SIP request may be redirected or may trigger a chain of new SIP requests by proxies. Users can register their location(s) with SIP servers.

SIP servers can act in two different modes - as proxy servers or as redirect servers. SIP proxy servers forward requests to the next hop, SIP server, or user-agent within an IP cloud. Redirect servers inform their clients of the address of the requested server and allow for the client to contact that server directly. Any number of hops can be traversed until the final destination for the request is found. A typical example of SIP redirect and proxy servers can be found in [4].

SIP defines six different method types and 37 headers. These methods and headers, when combined together, allow for complete control over a multimedia call session while limiting complexity. Here we briefly introduce each of the method. INVITE method is used to invite a user to a call and establishes a new connection. ACK is used to facilitate reliable message exchange for invitations. BYE terminates a connection between two users in a call or to decline an invitation. CANCEL terminates a request or the search for a user. REGISTER conveys information about a user's location to a SIP server. Finally, OPTIONS method is used to solicit information about SIP servers' capabilities.

It is important to notice that current version of SIP supports user mobility readily by proxying and redirecting requests to the user's current location. A callee may move between a number of different end systems over time. These locations can be dynamically registered with the SIP server. Users can register their current location by using registration services. The REGISTER request allows a client to let a proxy or redirect server know at which address(es) it can be reached.

SIP enables a wide variety of multi-party conferencing scenarios, which include multicast conferences, bridged conferences, and full-mesh conferences. Here we will just describe how SIP support Multi-point Control Unit (MCU) services, since it is essential to provide mobility support. More specifically, MCU can be extended to solve hand-off issue; details will be described in Section 3. The MCU service may be described as follows: Initially, users A and B have established a point-to-point connection in a MCU based conference. Later A wants to add C to join the conversation. A can do so by issuing an INVITE C request and indicating the address of the MCU in the Also header. The invitee (C) then contacts the MCU using the same session description and requests to be added to the call, just like a normal two-party call [14].

2.2 Mobile IP

In this section, we give an overview of Mobile IP. The mechanisms that will be used to provide SIP mobility support will be described in detail. Mobile IP is a proposed standard protocol that builds upon the IP by making mobility transparent to applications and higher level protocols like TCP. Its goal is to provide the ability of a host to stay connected to the Internet regardless of its location.

The main concern regarding mobility on the Internet is the fact that the IP address is usually representative of the location of the computer, in addition of being the endpoint identifier. The IP address can't conserve these two functions in mobile networks, as the location can change any time. One solution could be to give a new IP address to the node at each point of attachment. But it would imply that all IP clients established on the node would stop working each time there is a handoff, as they refer to the IP address. The IETF solution is to use two IP addresses, one constant address, the endpoint identifier, and one temporary address giving the location of the mobile. This second address is called care-of-address.

The endpoint identifier is relevant of the home network of the mobile. In this network, there is a home agent that keeps track of the location of all the mobile nodes originating from this network. In each foreign network, the foreign agent keeps track of the mobiles currently visiting the network and cooperates with the home agents to deliver the datagrams to the mobiles. The foreign agents and the home agents are collectively known as mobility agents.

When a mobile moves in a new network, it has to discover the new foreign agent. For this purpose, mobility agents broadcast periodically advertisements to make themselves known. These advertisement messages are extensions of the ICMP router advertisement messages. The mobile has the possibility to solicit an agent advertisement if it doesn't want to wait for the next one. Upon the reception of the agent advertisement, the node knows if it is in its home network or a foreign network. In the former case, its behavior is the same as the one of any other node. In the latter case, it has to obtain a care-of-address. This address is usually given by the agent advertisement. It can also be obtained using DHCP (Dynamic Host Control Protocol) [3] or PPP (Point-to-Point Protocol). Then, it has to register each new care-of-address with the home agent. This registration lasts only a finite amount of time. The mobile has to re-register after its expiration. The registration procedure must be secure. If somebody sends a registration request for another node, it is enough to disrupt all traffic destined to this node.

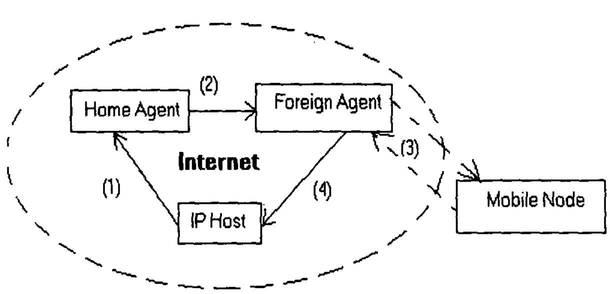


Fig 1. Triangle routing (tunneling)

The first routing scheme proposed by IETF for Mobile IP relied on *tunneling* (Fig 1). The packets sent to the mobile were directed to the home network. The home agent attracted them and encapsulated them (using either IP-within-IP encapsulation or minimal encapsulation) before sending the resulting packets to the mobile care-of-address. The foreign agent received and detunneled these packets before sending them to the mobile node.

A more efficient routing scheme has since been designed: *route optimization*. The idea in *route optimization* is that the routes to the mobile nodes would be improved if they hadn't to go by the home agent. It relies on an updated mobility binding provided to all peer entities communicating with the mobile node. It enables them to send directly the packets to the mobile care-of-address.

Route optimization allows smoother handoff than tunneling. With tunneling, the handoff is handled by the home agent. The mobile registers its new care-of-address with the home agent that will then tunnel the packets to this address. As the home agent is often far away, this process may last hundreds of milliseconds, during which megabits of data may be dropped. With route optimization, handoff may be achieved without any intervention from the home agent. This is performed under the authority of the mobile node. When the mobile moves from a cell to another, it tells its new foreign agent to contact its previous foreign agent to update the mobile binding. The problem here is about security. The previous foreign agent needs to be sure that the mobile node effectively requires handoff. However, the mobile and the foreign agent don't share any secret allowing authentication. A registration key has to be created. In this purpose, the foreign agent enters in communication with the home agent. They then both agree on a key that will be transmitted to the mobile. This process has to be performed just after the registration of the mobile with its new foreign agent, so that it's ready to use when a handoff occurs. The main drawback for *route optimization* is that most nodes in IPv4 are not able to understand this protocol. This feature has been introduced in IPv6 so that *route optimization* is now fully integrated into the protocol and *triangle routing* has been eliminated.

3. Solutions to Mobility Management in SIP

When a user initiates a PSTN phone call to an IP telephone on the Internet, the IP host may be currently on the move from networks to networks. To provide host mobility in Internet telephony, roaming should be supported and handoff should be performed upon crossing a subnet boundary that causes the IP address of the mobile phone to change.

For SIP to support mobile terminals, we need to target at three major issues:

- Foreign server discovery / registration
- Location management
- Handoffs

Mobile IP already deals efficiently with most of these issues; thus one option for mobility management is to use Mobile IP. However handoffs in Mobile IP are not performed fast enough for voice communications [5]. Some features of SIP could help to solve this problem. In the same way, special features of SIP could be used for assisting location management.

3.1 Registration/location update

When a mobile arrives in a new network (new cell), it has to discover a new BS, register with the new BS, and update its location with the location server. There is no such service in the current version of SIP, but Mobile IP may be used to provide the solution. In Mobile IP, the Foreign Agents periodically broadcast ICMP advertisement messages to signal their existence. The mobiles listen to these messages and engage a registration procedure with the Foreign Agent, obtaining a new temporary address (care-of-address). It then has to update its location with its Home Agent. This update process must comprise an authentication procedure. This is so because, if the location is updated without the mobile asking for (or aware of) it, the mobile would effectively be isolated from the network, and the communications for the mobiles could be received by another node (the one with the new temporary address corresponding to the entry for the mobile in the Home Agent location management table).

To provide security to this registration procedure, Mobile IP uses a set of security procedures called IPSec. SIP could run over Mobile IP and use these registration/location update/authentication mechanisms. Creating similar services in SIP to avoid using Mobile IP would be of limited use since the Mobile IP mechanisms work quite well.

With mobility comes the fact that the end-point identifier is no longer relevant of the location of the mobile. The solution offered by Mobile IP is to use two different addresses, one constant being the end-point identifier and one temporary address indicating the location: the care-of-address. This care-of-address is provided to the mobiles when they arrive in a new cell.

The mapping of the end-point identifier and the care-of-address is managed by the mobility agents in Mobile IP. We will see in the next part how it can be implemented in SIP.

3.2 Location Management

We consider 3 possible schemes for location management. In the first approach, one can make use of location management mechanism in Mobile IP by running SIP over Mobile IP. In the second approach, the SIP feature that supports people mobility is used. In the third approach, an independent location management service for terminal mobility is used. In the following three subsections, we describe and analysis each of these solutions.

A. Mobile IP location management service

The location management service in Mobile IP is distributed. There are mobility agents that administrate the mobiles and keep track of their location. These mobility agents are located in the BS and are of two kinds: the home agents and the foreign agents. Each mobile is registered with a home agent, in its home network, and is administered by a foreign agent in foreign networks. The home agent keeps track of the mobile location. When a mobile roams in a foreign network, it sends its new location, with its new care-of-address (see the part on Mobile IP), to its home agent. This scheme is efficient for small networks but will cause a lot of traffic overhead for the mobile registration/location update when the size of the network increases. It could nevertheless be used to offer location management without adding anything to SIP. It simply requires SIP to run on top of Mobile IP and this is the simplest solution to solve registration issues described earlier.

B. SIP location management server

In this scheme, the people mobility supporting features of SIP are used to support terminal mobility. In SIP, location servers keep track of the location of each user; i.e., on which terminal the user is working on (terminal identifier). These location servers could be used to keep track of the terminal mobility in addition to the people mobility. We can picture it as adding a third column to the location table, as shown in Table 1.

TABLE 1 LOCATION SERVER

User Name	Terminal Identifier	Terminal location
Mary@eniac	122.3.34.56	145.56.7.8
Gus@sjsu.edu	122.3.34.56	121.32.42.56
Pat@illy	125.34.343.23	165.38.66.12
Lisa@scrp.edu	143.32.34.57	122.3.34.56

This scheme would require additional software in the SIP location servers but allow a more scalable location management service, as it is external. To reduce traffic overhead, the location servers should administrate zones comprising a few cells. The size of this cluster of cells depends on the size of the cell. A hierarchy of location servers should be established, with the 2nd-level servers keeping track of which server is administrating each mobile and forwarding the requests for the mobiles location to the proper server. The number of levels of this hierarchy would depend on the size of the network. The choice of the number of levels in the hierarchy and the number of cells, and of servers for the upper levels, to be administered by each server, should be made with the preoccupation of minimizing the location update traffic.

Before establishing a communication with the mobile node, the caller will have to ask for its location. If the nodes are in the same zone, a simple request to the local location server will provide this information. Otherwise, the request will go through the hierarchy of location servers until the information is found. This structure is especially efficient for connection-oriented communications, as in voice communications. For voice communications, there must be a call establishment phase before the communication can take place. Finding the location of the mobile could be done during this phase. The loss of time resulting from the search of this location would be unimportant for such connections but would seriously add weight on connectionless transfer times. Once the connection is established, the routing can be done efficiently as triangular routing can be avoided.

This approach is definitely better to keep the amount of location updating reasonable while increasing network size. However, there will be a lot of overhead because of the two location management schemes working over each other (terminal and people mobility). In addition, there might be some redundancy in the updating traffic. For example, consider several users, referenced in different location servers, currently stored as users of the same machine. If this mobile machines moves to a new cell, its new temporary address will have to be updated in all location servers referencing the different users. The solution to this problem might be to update only the temporary address for the user currently using the mobile (we assume that there can not be more than one communication going on for one machine at the same time). The update of the other location tables would be delayed to a time when there is less traffic, so that it doesn't impair the on-going communications.

C. External Location Management Services

A third approach would be to design an independent location management service for terminal mobility. There would be two twin services, one for terminal mobility and one for people mobility. Both of them would be similar to the one described above. This would reduce useless updating traffic but would require frequent communications between the two systems. All in all, this system is probably less efficient than the previous ones, more complex to manage, and more costly.

3.3 Handoff

When a user initiates a PSTN phone call to an IP telephone on the Internet, the IP host may be currently on the move from networks to networks. Handoff occurs when a SIP mobile terminal is crossing a subnet boundary that causes the IP address of the mobile phone to change. It must be performed in such a way that the communication continues as if the mobile had stayed in the previous subnet. The goal is to achieve transparent handoff. The connections must be handed over fast enough to prevent loss of information. A new connection must be established before the old one is released.

As discussed in Section 2, SIP supports MCU services readily. As will be clear shortly, using MCU for handoff management provides a natural extension to SIP for mobility support. The proposed procedure is described as follows.

Initially, mobile terminal A wants to talk to B. A sends an INVITE request to B and indicates the address of the MCU in the Also header:

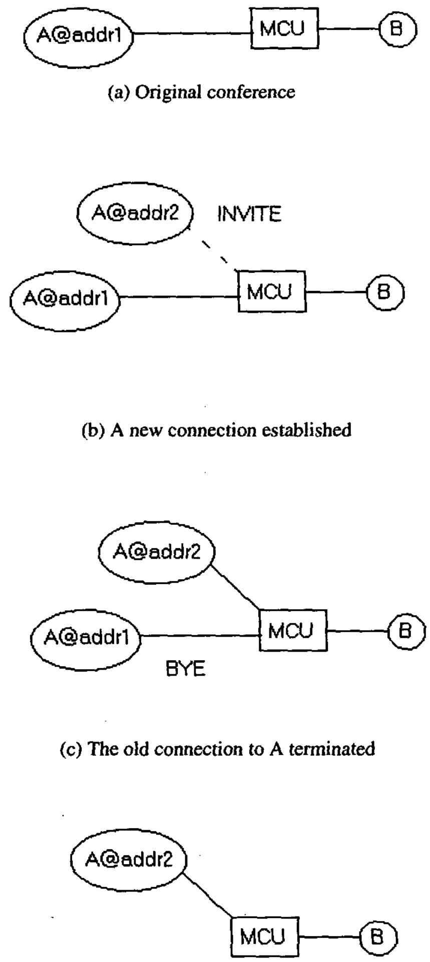
A → B: INVITE B SIP /2.0
Call-ID: 19990524C1234567.22@A
From: A @ addr1
Also: conference100 @ M

B → M: INVITE conference100 @ M
Call-ID: 19990524C1234567.22@A
Requested-By : A @ addr1

A and B then establish a connection in an MCU-based conference. Later, when A is roaming across the boundary of the subnet, he gets a new care-of address. Then he can send an INVITE request to MCU indicating his new address (represented by A1), with a Replace Header containing the old address of A. Note that the Call-ID field should not be changed during the process.

A→ MCU: INVITE A @ addr2 SIP/2.0
Call-ID: 19990524C1234567.22@A
From A @ addr1
Also: conference100@M
Replace : A @ addr2

After a new connection between A and MCU established, MCU terminates the old connection to A by sending BYE request. Hand-off is thus performed seamlessly. The whole process is illustrated in Fig. 2.



(a) Original conference

(b) A new connection established

(c) The old connection to A terminated

(d) The resulting conference

Fig. 2 Handoff procedures

Technically, there will then be two connections established at the same instant when a handoff occurs. The old connection is maintained until the new one is fully established. Having simultaneously two connections established guarantees that there will be no drop in the data rate when the first connection is released. There is no rerouting to be performed. However, this technique wastes some resources as two connections, with the QoS requirements of voice communications, are maintained for only one communication. To reduce this waste, the handoff has to be as fast as possible. To reduce handoff latency, the MCU has to be located as close as possible, in the hierarchical structure of the switches, from the BS managing the cells.

4. CONCLUSION

We have described mechanisms of providing mobility support to SIP, addressing the issues of registration in roaming, location management, and handoff. This shows that SIP may be readily extended to run on top of Mobile IP to support terminal mobility. The major weakness of Mobile IP for supporting mobile telephony is slow handoff management. We have shown that with its multipoint conference system, SIP provides a good solution. Similarly, the location management scheme implemented in SIP for people mobility may be extended to support terminal mobility in a more efficient way than in Mobile IP. It is clear, therefore, that SIP is able to compete with H.323 in the support of mobile telephony over the Internet.

The marriage of mobile networking technology with telephony could make the access points to the wired network very complicated, expensive devices. Indeed, these access points would have to incorporate the functions of the SIP servers (for fast handoff support) and the base stations. Their function as BS implies that they would incorporate both wired and wireless protocol stacks, with management functions, and the hardware enabling them to send electro-magnetic waves. Removing location management functions from these access points avoids adding another layer of complexity on these devices. Other management functions could be put on these servers, or on similar servers, to further decrease the price of these devices.

A point that has not been discussed here is the protocol choices on the wireless part of the network. Obviously, these protocols should take into account the different classes of service, as voice communications have to be supported. Also, they should be different from the standard protocols running over the wired network. Such protocols should apply different policies of services to voice communications than to data communications, insisting on a low-delay jigger and in-sequence delivery, and giving priority to voice communications. The research conducted in this field for Wireless ATM gives a good example of what should be done (readers may refer to [2], [8], [9] and [13] for more information).

REFERENCES

[1] U. Black, *Mobile and Wireless Networks*, Prentice Hall, 1996.

[2] J-C Chen, K. M. Sivalingam and R. Acharya , "Comparative Analysis of Wireless ATM Channel Access Protocols Supporting Multimedia Traffic" , *ACM/Baltzer Journal on Special Topics in Mobile Networking and Applications (MONET)*, Vol. 3, pp. 293-306, 1998.

[3] R. Droms, "Dynamic Host Configuration Protocol," IETF RFC 2131, March 1997.

[4] M. Handley and H.Schulzrinne, et. al. RFC 2543, SIP : Session Initiation Protocol.

[5] J. Lennox, H. Schulzrinne and T. Porta, "Implementing Intelligent Network Services with the Session Initiation Protocol ," http://www.cs.columbia .edu/~hgs/sip/papers.html

[6] W. Liao, "Mobile Internet telephony: Mobile Extensions to H.323," *Proceedings of IEEE INFOCOM' 99*, New York, March 1999.

[7] W. Liao, "Mobile Internet Telephony Protocol: An Application Layer Protocol for Mobile Internet Telephony Service," Proceedings of IEEE International Conference on Communications (ICC), Vancouver, Canada, June 1999.

[8] W. M. Moh and Y. Chen, "Multicasting Flow Control over Hybrid Wired/Wireless ATM," accepted by *Performance and Evaluation Journal.*

[9] W. M. Moh and H. Mei, ``Multicast GFR for Supporting TCP/IP over Hybrid Wired/Wireless ATM," *Proceedings of IEEE International Workshop on Wireless Mobile ATM, (wmATM'99)*, San Francisco, CA., June 1999; an extended version has been accepted by *ACM/Baltzer International Journal on Wireless Netwroks.*

[10] W. M. Moh, D. Yao, and K. Makki ``Wireless LAN: Study of Hidden-Terminal Effect and Multimedia Support," *the Proceedings of International Conference on Computer Communication Networks (IC3N'98)*, pp. 422-431, Lafayette, LA, Oct. 1998.

[11] C. Perkins, "IP Mobility Support," RFC 2002, October 1996.

[12] C. E. Perkins, " Mobile Networking Through Mobile IP," http://www.computer.org/internet/v2nl/perkins.htm

[13] D. Raychaudhuri, L.J. French, R.J. Siracusa, S.K. Biswas, R. Yuan, P. Narasimhan and C. Johnston, "WATMnet: A Prototype Wireless ATM System for Multimedia Personal Communication," *IEEE JSAC special Issue on Wireless ATM*, April 1997.

[14] H. Schulzrinne and J. Rosenberg, "SIP Call Control Services," Internet Draft, Internet Engineering Task Force, March 1998. Work in progress.

[15] H. Schulzrinne and J. Rosenberg, " A Comparison of SIP and H.323 for Internet Telephony," http://www.cs.columbia .edu/~hgs/sip/papers.html

[16] H. Schulzrinne and J. Rosenberg, "Signaling for Internet Telephony", http://www.cs.columbia .edu/~hgs/sip/papers.html.

[17] M. Umehira, A. Hashimoto and H. Matsue, "An ATM Wireless Access System for Tetherless Multimedia Services," *Proceedings of IEEE International Conference on Universal Personal Communications (ICUPC)*, Tokyo, Japan, Nov. 1995.

[18] ITU Recommendation H.323, "*Packet-Based Multimedia Communications Systems,*" Feb. 1998.