MSIP: A protocol for efficient handoffs of real-time multimedia sessions in mobile wireless scenarios

Ranjeeth Kumar A Sridhar Iyer
[ranjeeth, sri]@it.iitb.ac.in
School of Information Technology
IIT Bombay

***Abstract*— The support for IP mobility has become very important with the increasing growth of mobile applications. One of the key challenges for the deployment of such wireless Internet infrastructure is to efficiently manage user mobility. Mobility is handled by Mobile IP (at the network layer) and Session Initiation Protocol (at the application layer). The main aim of these schemes is to provide seamless connectivity for ongoing communications, that is, to keep the handoff latency delay as less as possible. The delay constraint becomes even more crucial for real time applications. The Mobile IP scheme has drawbacks like triangular routing, longer delays and the need for tunneling management. SIP solves many of the problems in Mobile IP, but it involves the time consuming process of obtaining a new IP address from the DHCP server.**

In this paper we propose a hybrid scheme, MSIP, that is the integration of Mobile IP and SIP. MSIP reduces the handoff delay for realtime applications. It avoids the need for tunneling the packets throughout the handoff period and the communication need not be hung up till a new IP is given by the DHCP server.

I. INTRODUCTION

With the ever increasing use of mobile devices, such as laptops, the demand for host mobility and support for various mobile applications is also growing considerably. For any mobile node, if it wants to take full advantage of this mobility provided, the mobile node should be able to move to any point in the network and still retain all its ongoing communications. The cell handoff involves a delay called handoff delay, which is the period of time between the moment at which the mobile node detects the subnet change, and the time at which it receives the first packet of its ongoing communication in the new subnet. Existing mobility protocols have mainly been designed for network, and application layers, and the majority of studies refer to the inherent mobility support provided by the wireless network. The primary design goal of any scheme, that handles this mobility, is to keep the handoff delay as less as possible. If the applications are real time then this constraint on delay becomes even more crucial, as the real time applications are highly delay-sensitive. Schemes like Mobile IP and SIP exist that handle mobility at network and application layers respectively, but there are several issues that are yet to be resolved in these schemes. The problem with the Mobile IP scheme is the very famous triangular routing problem, and also that all the packets need to be tunneled through the home agent, then to the foriegn agent and then to the mobile node, leading to much longer delays [1] [2] [5]. The problem with SIP is that after the the mobile node moves to a new subnet, it should contact a DHCP server to get a new IP [6], and then inform the other communicating party about its new IP. The process of getting a new IP from the DHCP server is quite time consuming.
The solution proposed in this paper, **M**SIP, is a combination of network and application layer mobility management models reduces the signaling load and provides fast handoff for ongoing conversations. MSIP reduces the handoff delay, and gives better performance for realtime applications.

This report is organized as follows. The working of Mobile IP and its disadvantages have been discussed in section two. In the third section, the SIP working and its problems are discussed. The algorithm and working of MSIP scheme are discussed in the fourth section. Fifth section consists of Future work and conclusions.

II. MOBILE IP

Mobile IP is a transparent solution that handles mobility at the network layer. This scheme was designed to solve the mobility problem by allowing the mobile node to use two IP addresses: a fixed home address and a care-of address that changes at each new point of attachment. This solution is transparent because, when two nodes are communicating with each other and one of them moves to a new subnet, then the other node is completely unaware of this mobility, and it continues its communication as if nothing has happened. In Mobile IP, it is assumed that every mobile node (MN) has its home network, and a statically allocated IP address on its home network. In Mobile IP, the support for mobility is provided by adding IP tunneling to IP routing. The Mobile IP architecture mainly consists of

1) Mobile Nodes (MN)
2) Correspondent Nodes (CN) : The CN participates in communication with the MN. The CN can be either fixed or a mobile node.
3) Home Agents (HA) : The default router on the home network, HA should store the current locations of all the mobile nodes in its network. It should intercept the packets from CN and tunnel them to the current location of the mobile node.
4) Foriegn Agents (FA) : The default router on the foriegn network, FA should receive the tunneled packets from

HA and forward them to the MN in its network.

A. How Mobile IP Works

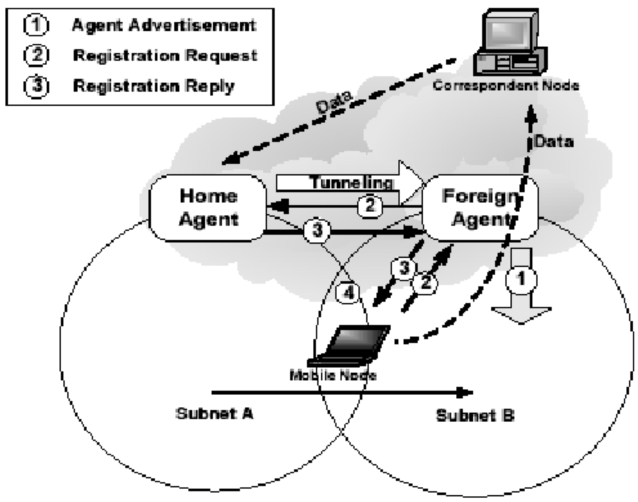

Fig. 1. Working of Mobile IP

Mobile IP consists of three main functions,

1) Mobile agent discovery
2) Registration with home agent
3) Packet delivery using IP tunneling

The Home agents and the Foriegn agents will broadcast agent advertisements in their subnets. Any mobile node detects that it has moved to a new subnet, when it hears the advertisement from an agent which is not its HA. Then it sends a registration request to its HA through FA with a care-of-address(CoA), which is generally the address of the foriegn agent itself. The HA then replies either accepting or denying that request. The CN will be completely unaware of this mobility and will transmit the packets to the mobile nodeś home address. The HA will intercept these packets and performs the IP-in-IP encapsulation and tunnels them to the CoA of the mobile node(i.e., FA). The FA then decapsulates the packets and forwards them to the mobile node.

B. Problems with Mobile IP

The famous problem with Mobile IP is *triangular routing*. All the packets from CN to MN, should take the longer path from CN to HA, from HA to FA, and then from FA to MN, but the packets from MN to CN, go directly to CN. One solution was proposed to solve this problem, *Route Optimization*.

Route Optimization : Route optimization solves the triangular routing problem by using binding updates to inform the correspondent host about the current IP address. However, route optimization has several drawbacks:

1) Route optimization requires changes in the IP stack of the correspondent host, since it must be able to encapsulate IP packets, and store care-of addresses of the foreign agent or mobile host.
2) Only the home agent may send binding updates to correspondent hosts. This means that there will be an extra delay before the correspondent host finds out where to send the packets, during which the old foreign agent must forward the packets to the correct location.
3) The mobile host needs to rely on the old foreign agent forwarding packets to its new foreign agent until the correspondent host has got the binding update. There is no requirement saying that the foreign agent must do so.

Because of the requirements that are put on the correspondent hosts, it cannot be expected that route optimization will be widely employed in a near future. Moreover, the home and foreign agent can become bottlenecks since they must handle the tunnels for a possibly large number of mobile hosts.

III. Session Initiation Protocol

The Session Initiation Protocol (SIP) is the application-layer solution to handle mobility. This protocol is used for establishing and tearing down multimedia sessions. It has been standardized by IETF for the invitation to multicast conferences and Internet telephone calls. A user is addressed using an email-like address üser@hosț, where user is a user name or phone number and host is a domain name or numerical address.

The session is active when the first INVITE request is issued and the call is established. The session terminates when a BYE request is issued. It also supports mobility using proxy or redirecting servers to reach the Callee. SIP is also independent of transport protocol. SIP is a client-server protocol where the client generates the SIP Request messages while the server receives the messages and send responses. There are three kinds of servers in use with SIP. They are namely the Proxy servers, Redirect Servers and the User Agent Servers (UAS). When a Proxy Server receives a Request message, it acts like a hop server or a router and forwards the Request message to the next Proxy Server or the User Agent Server at the receiving end-system. As for a Redirect Server, when it receives a Request message, it returns an IP address of the Callee's User Agent Sever so that the Caller, which is a client, would have to resend the request to the UAS.

The SIP Redirect server is preferred to the Proxy server, because the redirect server just relays the current location of the user, whereas the proxy server participates in the entire signaling session and thus increasing the load on the server. The working of SIP in presence of a redirect server is explained below.

A. How SIP Works

The above figure shows the working of SIP in the presence of a Redirect server. The steps followed are :

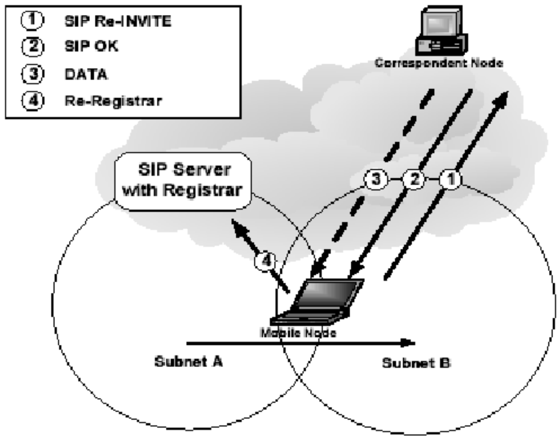

Fig. 2. SIP

1) Caller A, creates a session with an INVITE request for Callee B, sip:B@abc.com. This request is forwarded to the SIP Redirect Server of B.
2) The redirect server at Bś site looks for Callee B in the abc.com domain, with the help of a Location Server, usually a DNS.
3) The Location Server or DNS then replies that Callee B is logged in at BDomain.abc.com. The redirect server knows this through a static configuration, database entry or dynamic binding set up by Callee B with a SIP REGISTER message.
4) Caller A or UAC sends an ACK message to the Redirect Server that it received its response message.
5) Caller A now resends the INVITE request to Callee B directly at its current location, sip:B@BDomain.abc.com.
6) This time, Callee B or UAS, returns an ACK message to the UAC of Caller A to mean that Callee B acknowledges the last response from Caller A and the request is a success.
7) Finally, Caller A or UAC sends an ACK message to Callee B that it received its 200 response message for request success.

B. Problems with SIP

1) SIP is generally not preferred for TCP applications.
2) The process of getting a new IP from the DHCP server involves a significant amount of delay (typically 1-2 sec).
3) The communication resumes more quickly im Mobile IP than SIP. In Mobile IP, this communication starts as soon as the mobile node hears the Agent advertisement in the new subnet, which is not the case with SIP.

IV. Motivation For the New Scheme

The realtime applications are highly delay-sensitive. In order to support mobility for such applications, the main challenge would be to provide seamless connectivity for the ongoing communications with as lesser delay as possible. The motivation for MSIP is, the longer delay problems associated with the existing schemes, Mobile IP and SIP. The problems are

1) In Mobile IP all the packets for the entire session, the period in which the mobile node (MN) is in foriegn network, should travel through the longer path i.e., from correspondent node (CN) to home agent (HA) then to foriegn agent (FA) and then to MN, thus resulting in longer delays, not desirable for realtime applications.

Inspite of the above mentioned problems associated with Mobile IP and SIP, these schemes have certain advantages over each other. In Mobile IP, after the node moves to the new subnet, the communication resumes more quickly. This is because, the agents will broadcast the advertisements periodically with available CoA's. As soon as the mobile node listens this advertisement it sends a registration to its home agent and restarts the communication. In SIP, the packets need not be tunneled through the longer path after the handoff, but can be sent directly to the node's new location. The main aim of MSIP is to reduce the handoff delay for realtime applications by integrating these good features from both Mobile IP and SIP. In this new scheme we introduce a agent called **Mobility Agent** which is an integration of SIP Redirect Server and Mobile IP Home Agent.

V. MSIP - Proposed Solution

In this paper, we propose MSIP, an integrated scheme, combination of Mobile IP and SIP, that results in a very less handoff latency time. This combination of network and application layer mobility management models reduces the signaling load and provides fast handoff for ongoing conversations.

A. How MSIP Works

If the Correspondent Node(CN) wants to communicate with the Mobile node(MN), they start the communication using the SIP call establishment mechanism. If the MN moves to a new subnet, with the session being still active, then, the MN switches to Mobile IP scheme and receives the packets in the new subnet with tunneling being done by its *Home Mobility Agent*. The tunneling is required only until the MN receives a new IP address from the DHCP server in the new subnet. As soon as the MN gets the new IP, it sends the SIP REINVITE message, with the same Call-ID, to CN informing it about its new IP. From then on, tunneling is stopped, and the packets will be sent to the current location of the MN directly.

The detailed working of our new scheme is explained below

If Correspondent Node (CN) wants to communicate with Mobile Node (MN), then

1) CN sends an INVITE message to the Mobility Agent (MA) in MN's domain.

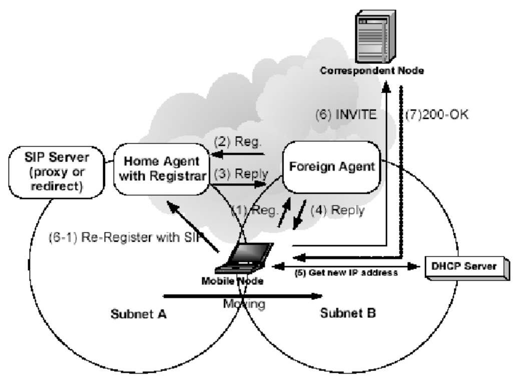

Fig. 3. Working of MSIP

2) The Mobility Agent performs a DNS lookup for the current location of the MN, and gives this information to the CN.
3) CN sends the INVITE message directly to the current location of MN, and the call is established.

   Meanwhile, if the MN moves to a new subnet during the ongoing communication with CN, results in a handoff. Then
4) MN sends a Registration request to its Home Mobility Agent (HMA) through Foriegn Mobility Agent (FMA).
5) MN gets the reply from its HMA accepting its registration.
6) Soon after this, the transmission resumes, as in normal Mobile IP scheme.
7) The packets will be tunneled to the MN through FMA by the HMA.
8) Mean while, the MN contacts DHCP server to get a new IP address.
9) After it gets a new IP, it sends the SIP RE-INVITE message to the Correspondent Node (CN), with its new location.
10) From then on, the packets will be sent directly to the MN s new current location without tunneling.
11) The MN sends a SIP RE-REGISTER message to its HMA, so that any new connections in future can be correctly redirected to the MN s current location.

The messages involved in the handoff process are given in Fig.4

B. Advantages

Our scheme, MSIP, overcomes the above mentioned problems associated with Mobile IP and SIP. The communication resumes as soon as the node hears the agent advertisement in the foriegn network as it happens in Mobile IP, without waiting until the node obtains the new IP address from the DHCP server asin SIP scheme. Also, the packets need not be tunneled through the longer path, for the entire session as in Mobile IP.

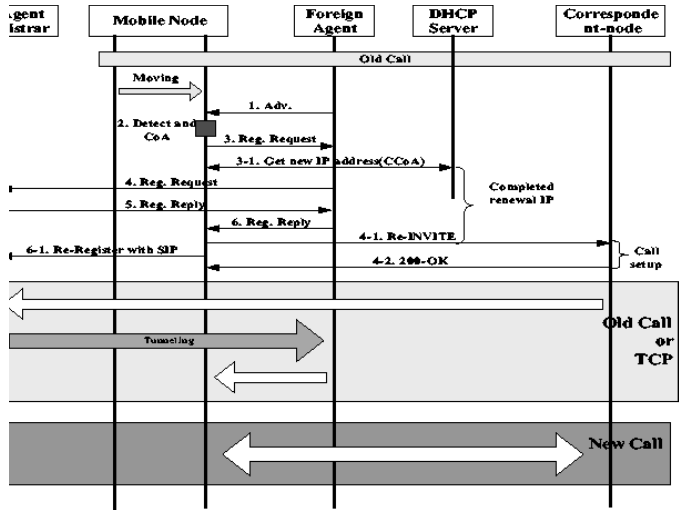

Fig. 4. Message flow during a Handoff

VI. Future Work and Conclusions

MSIP performs better than Mobile IP and SIP. The better performance here means a leser handoff delay. In our hybrid scheme it takes lesser time, for the first packet (after the handoff), of the ongoing communication, to reach the MN. There are some issues yet to be resolved in this scheme, like authentication. Experiments need to be done, to measure these authentication delays, required for exchange of extra control messages (Extra, because, in our scheme, control messages of both Mobile IP and SIP are involved).

The delay involved in obtaining a new IP address from a DHCP server mainly depends on its implementation. The duration of the hand-off period would be significantly reduced if the mobile host could immediately begin using the new IP address while verifying in the background that it is not in use by any other host. In addition, the accuracy and speed of determining when handoff is necessary could be increased if the 802.11 network interface card could notify the application when it associated with a new AP, to prevent continuous polling of the driver, and also if the access points could be queried for their subnet numbers.

References

[1] Charles Perkins, ”IP mobility support”, *Request for Comments (Proposed Standard) 2002, Internet Engineering Task Force*, 1996.

[2] D. B. Johnson and Charles Perkins, ”Route Optimization in Mobile IP”, *Internet Draft, Internet Engineering Task Force*, 1996.

[3] Charles Perkins, ”IP encapsulation within IP”, *Request for Comments (Proposed Standard) 2003, Internet Engineering Task Force* , 1996.

[4] Amir Mahajan Ben, ”Route Optimizations In Mobile IP”, citeseer.ist.psu.edu/570186.html

[5] Jon-Olov Vatn, ”Long Random Wait Times for Getting a Care-of Address are a Danger to Mobile Multimedia”, *International Workshop on Mobile Multimedia Communications, ”MoMuC 99”*, November 1999.

[6] Elin Wedlund and Henning Schulzrinne, ”Mobility Support Using SIP”, *”WOWMOM”*, 1999.

[7] M. Handley and H. Schulzrinne and E. Schooler, ”SIP: Session initiation protocol”, *Internet Draft, Internet Engineering Task Force, July 1997. Work in progress.*, 1997.

[8] G. Ayorkor Mills-Tettey and David Kotz, ”Mobile Voice over IP (MVOIP): An Application-Level Protocol”, citeseer.ist.psu.edu/571001.html

[9] Xinhua Zhao and Claude Castelluccia and Mary Baker, ”Flexible Network Support for Mobility”, *Mobile Computing and Networking*, 1998, citeseer.ist.psu.edu/zhao98flexible.html

[10] S. Giordano, M. Mancino, A. Martucci and S. Niccolini, ”Managing Multimedia Traffic in IP Integrated over Differentiated Services: SIP dynamic signaling inter-working”, *36th Annual Hawaii International Conference on System Sciences*, 2003.

[11] A. Dutta, F. Vakil, J.-C. Chen, M. Tauil, S. Baba, N. Nakajima, H. Schulzrinne, ”Application Layer Mobility Management Scheme for Wireless Internet”, *3G Wireless 2001*, May 2001, pp. 7.

[12] Nilanjan Banerjee, Wei Wu, Sajal K. Das, Spencer Dawkins and Jogen Pathak, ”Mobility Support in Wireless Internet”, *IEEE Wireless Communications, 2003*, October 2003.