Multimedia SIP sessions in a Mobile Heterogeneous Access Environment

Ashutosh Dutta, Yibei Ling, Wai Chen, Jasmine Chennikara
Telcordia Technologies Inc., 445 South Street, Morristown, NJ 07960

Onur Altintas, Toyota InfoTechnology Center, 4009 Miranda Avenue, Palo Alto, CA 94304

Henning Schulzrinne
Computer Science Department, Columbia University, New York, NY 10027

Abstract—

Maintaining multimedia sessions by means of SIP signaling has often been termed as application layer mobility management scheme. There have been numerous proposals to support different types of mobility using SIP signaling which is generally used to set up and tear down the multimedia stream. Mobility management in heterogeneous networks plays a very important role since the user could be moving between multiple types of access networks involving many service providers during a multimedia session. These access networks could be 802.11b, CDPD, CDMA or GPRS based network supporting DHCP or PPP servers in the networks. The movement of the mobile host can be between the access networks where each access network may belong to the same subnet, same domain but different subnets or different different domain altogether. In most cases the end-client would have access to both the networks at the same time, but connectivity to the network would be determined by any local policy defined in the client itself such as signal strength or any other measurement based on QoS parameter of the traffic. This paper discusses various issues associated with SIP signaling for maintaining continuity of both signaling and media flow in a Local Area Network and Wide Area Network environment and presents the experimental results associated carried out in the testbed.

I. Introduction

SIP [3] is a signaling protocol to setup and tear down multimedia sessions between the end-clients. Such sessions can be set up using direct method between SIP user agents resident within the clients or via SIP proxy/redirect servers. Basic mobility can be defined when the client moves between two access points or RANs (Radio Access Networks) connected to different kinds of networks. Each type of RAN for different types of networks (e.g., 802.11b, CDPD, CDMA) may belong to the same subnet or different one within the same domain. It is important to reduce the transient data loss due to frequent hand-off associated with the client's mobility. There have been various proposals to take care of mobility in homogeneous networks using SIP signaling [1], [4], [2], [5]. However little has been studied about how SIP based mobility can take care of movement between heterogeneous access networks. When a client is multi-homed and is connected to a set of heterogeneous access networks, it may have multiple interfaces active at a particular time. Both pre-session and mid-session mobility of the SIP sessions would be affected due to heterogeneity of the networks during the client's movement.

A SIP session consists of both signaling and media delivery, where media delivery can be meant for TCP based non-real-time traffic or for RTP/UDP based real-time traffic. While switching between access networks, IP address acquisition method (e.g., DHCP, PPP) would be determined based on the kind of access network the client is connected to. For example, a client on a 802.11b network would depend upon DHCP server for IP address acquisition but on a CDPD or CDMA network, IP address would be acquired from a NAS (Network Access Server) via PPP. Few of the scenarios are described below with respect to SIP based mobility in heterogeneous access networks. Supporting mobility for heterogeneous access using SIP would involve various key issues such as network detection, active interface identification, registration, re-transmission of invites, redirection, session continuity, fast handoff and asymmetry of data delivery. Some of the issues associated with SIP sessions in heterogeneous networks are described later in section III. Although media delivery in a heterogeneous access network can be taken care of by using traditional Mobile IP [8] approach, where the mobile client registers with the home agent through a different active interface [10]; it is not the focus of discussion here.

This paper is organized is as follows. Section II describes several architecture alternatives and possible network access. Section III outlines several issues with respect to signaling and media delivery. Section IV concludes the paper.

II. Architecture Alternatives

This section discusses some of the architecture alternatives of SIP sessions in a heterogeneous access environment. It can mostly be divided into two main categories such as Local Area Network (LAN) and Wide Area Network (WAN). Types of networks are the factors that would determine the placement of SIP servers, ways of acquiring an IP address, security, and link layer access.

A. Local Area Access

Local Area access is limited to an autonomous system such as an enterprise, where moving between the subnets does not involve any sort of AAA (Accounting, Authorization, Authentication) mechanism [9]. Scenario 1 and scenario 2 provide two different kinds of LAN access based on the position and movement of the mobile host.

In scenario 1 which is mostly for enterprise based system, the end clients are equipped with SIP user agents. Each mobile client may be equipped with 802.11b and Bluetooth interface for communication. There are two cases within scenario 1, such as case I and case II based on the location of the CH. At any particular point of time, only one interface has access to the network. In this case, each interface card has a separate IP address but each belongs to the same subnet. SIP based signaling sets up the multimedia session between MH (Mobile Host) and CH (Correspondent Host). MH moves from one access point to another one where by one of the interface becomes active instead of the other one. In cases where both the interfaces have strong signal to noise ratio (SNR), then both the interfaces get assigned IP addresses. Depending upon the metric value within the OS, one of these is chosen as the active address while sending traffic out.

In scenario II, the mobile has connection to different access points which belong to two different subnets. They may share the same SIP server connected to a common subnet however. In this scenario the mobile's con-

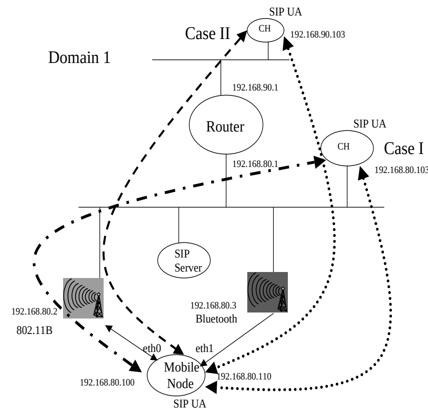

Fig. 1. Heterogeneous SIP in a LAN: Scenario 1

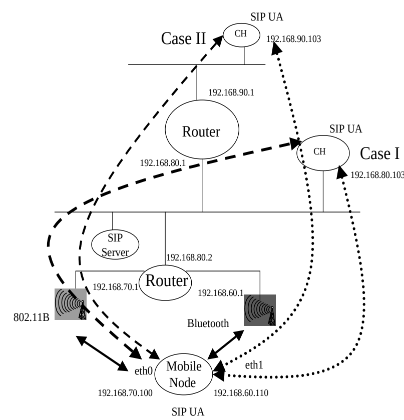

Fig. 2. Heterogeneous SIP in a LAN: Scenario 2

nection would move from one access point to another based on the signal quality or any other parameters. Thus an ongoing session would involve re-registration with a new SIP server, and multimedia session transfer. Figure 2 is a picture that depicts a similar scenario except that subnets associated with two interfaces are different.

Following is the description of a set of experiments to verify heterogeneous mobility functionality for multimedia SIP sessions. Mobile node has connection to two different IP addresses but on the same LAN thus same subnet. In this case the addresses assigned are 192.168.80.100 and 192.168.80.110. CH address is

192.168.80.103 or 192.168.90.103 depending on where it is connected. Experiments were tried for both SIP signaling and media delivery. Some of the results during this handoff are described later in the section. SIP server is connected in the network where registration takes place while the mobile moves between the private subnets. With the help of a public SIP server this experiment could be extended to a wide area network.

As part of experiment 1 both eth0 and eth1 are connected to two different access networks within an enterprise. CH is in the same subnet with the IP address (192.168.80.103). A video session was established between the mobile node and correspondent node. Initial communication was taking place via eth0. At this point, communication via eth0 was disrupted by either taking it down, unplugging the wire or moving away from the access point, and eth1 was made up, the video session still remains active. At this point CH still thinks it is sending traffic to 192.168.80.100 which is the IP address associated with eth0. Same sequence of operation was repeated when CH was connected to an interface which is not on the same LAN. In this case while bringing the interface up, we had to bring the default route up also. SIP based mobility (i.e Re-Invite feature) took care of the media redirection to the mobile host's new IP address. In both the cases video transmission breaks if the old interface card is taken out completely. As long as the interface card is installed inside, the there is an ARP entry in the cache and the other interface knows how to associate the old IP address with the new MAC address even if the client moves away from the access point. This routing table however gows away in the case of windows XP and windows 2000 when the client moves away from the access point.

B. Wide Area Access

Figure 3 depicts movement of the client with multiple access in a wide area environment involving 802.11b and CDPD or CDMA1xRTT networks. Handoff to CDPD and CDMA1xRTT networks was made possible using the connection provided by Verizon Wireless. This involves moving between a LAN within an enterprise and a public network such as a wireless ISP such as Verizon. In this case the client gets assigned the IP address and other parameters such as DNS, and SIP server on its 802.11b interface using a DHCP server, but gets an IP address and associated parameters via a wide area network using PPP connection over CDPD/CDMA.

As the client moves from one location to another, it connects to another subnet which may be part of a LAN or CDPD/CDMA network. It was observed that obtaining an IP address via CDPD/CDMA takes more time (few secs) than obtaining the address using DHCP without ARP option (1sec). According to the measurement it takes almost 30 sec. to establish the communication over the PPP link after getting initialized and authenticated properly with the NAS. In Windows platform it was observed that the clients get an temporary address called auto IP (169.x.x.x) before it gets the real address from the DHCP server because of the delay associated with the address acquisition. During the experiment it was found that average throughput over CDPD network is about 8 kbps and that over CDMA network it is about 60 kbps.

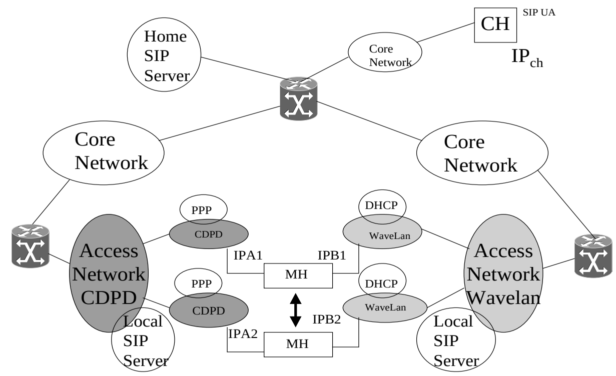

Fig. 3. Wide Area Access Network- Scenario A

Figures 3 and 4 show architecture alternatives for supporting SIP sessions in heterogeneous access networks where the client keeps on moving between different networks while still being connected to one or both of the networks. In some cases however, even if the MH is not moving (changing its location) it may alternate from one access interface to another access interface triggered by some local event thus creating a scenario for mobility. In the experiment we performed, Toshiba satellite has a switch to activate/de-activate the inbuilt 802.11b interface. In the following section we consider the scenario when the MH is moving from one place to another. In each location (before and after moving) mobile host has access to both the networks (CDPD, 802.11B through its respective interface). At any particular point of time a mobile host can have both the interfaces up and connected to a different network type. Each interface would be assigned a different IP address from the associated DHCP/PPP server in that access domain. It is to be de-

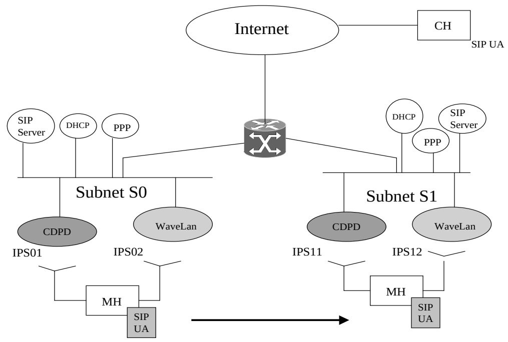

Fig. 4. Wide Area Access Network - Scenario B

termined by local decision as to how the mobile host would choose one of the interfaces active at any particular time in order to communicate with the CH. This can be done by assigning priority based scheme such as in Windows environment, routing metrics (gated approach) or allocating shadow address (i.e, choose one of the MAC address as shadow address and associate that with the IP addresses) scheme [6]. As the MH moves to a new location both of its interfaces would connect to two different access networks which are on different subnets than the previous ones, and the IP addresses would be configured accordingly by the DHCP servers. As the IP address gets configured from the DHCP server, it also gets configured with the SIP server for that access network. In most cases there is one SIP server for each access domain. In this particular case each access network has a local SIP server. It can so happen that when Mobile host moves from one place to another one it may try to communicate via 802.11b (location 1) or via CDPD (location 2), or via 802.11b (location3) thus making an alternate choice of access network for connection each time it moves from its present location. One needs to be careful about the assymeteric path taken when the mobile host moves. Re-Invite should take into account the IP address of the interface which is chosen to communicate through.

Figure 4 is another illustration of heterogeneous access network, where both the interfaces of the mobile host belong to the same subnet at any particular time, but each interface gets a different IP address on the same subnet. So as the client moves to another location, both of its interfaces get connected to another set of access points both belonging to subnet S1. It is important to figure out which interface is active (host is communicating

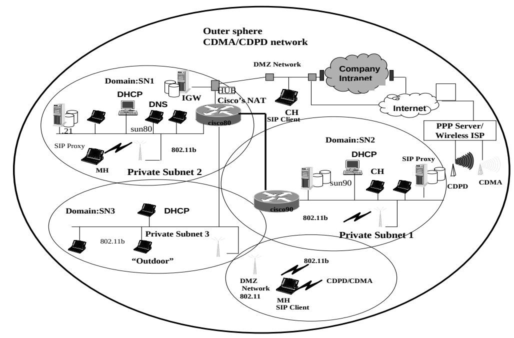

Fig. 5. Heterogeneous Access Testbed

through) at any particular point of time. SIP User Agent should take notice of the interface which is active at any particular point of time and accordingly send that as the contact address as part of its Re-INVITE. Choosing the right interface by the SIPUA is a method of implementation.

Figure 5 shows the testbed architecture where the SIP heterogeneity experiment is being carried out. It involves private subnets behind NAT on 802.11b network, CDPD, CDMA networks from wireless ISP connected to the Internet, and DMZ network associated with an enterprise with the wireless access. Several experiments were conducted where the mobile host moved between different kinds of networks, including the firewall traversal.

III. Issues

This section discusses issues related to SIP multimedia session in Heterogeneous Network. Each of the issues and solution there of has been discussed in the context of SIP sessions between the end clients which mostly comprises different signaling scheme such as Invite, Registration, and handoffs related to media delivery.

A. SIP Signaling setup

In case of the end client being multi-homed, SIP Invite messages can be either direct or proxy based. When the Invite is proxy based, the proxy server takes care of the redirection, but in case of direct signaling, it is important to take care of both signaling and media delivery on the UA side. Since completion of SIP signaling consists of several intermediate steps (Invite, Provisional responses (180, 183 etc.), Ack, OK), it may so happen that MH would change its access network even before the call setup is complete. In this scenario dynamic DNS can be taken advantage of. In most cases there is a re-transmission of Invites, in case the CH cannot reach mobile host or mobile host changes its address or moves to a different network even before Invite is received by the mobile host. CH's SIP user agent would try to re-transmit the Invites for up to seven times, until it times out. If the SIP UA is not using the canonical address scheme for host name resolution, subsequent request has to go through a SIP server which knows the most recent active interface of the Mobile host. Subsequent responses (e.g., 180 ringing, Ack, OK) can be sent direct to the correspondent host.

B. SIP sessions with NAT

As one of the alternate architectures, correspondent host can be within a LAN behind a NAT (Network Address Translator) box, and can be wired or wireless. Where as the mobile host can be moving around between two different wide area networks (e.g. CDMA, CDPD or external 802.11 provided by an ISP) which are outside the NAT box. In this case if a correspondent host makes a call to the mobile host, SIP Invite request will reach the mobile host, but response to this Invite request (e.g., 180, 200), should be sent to the NAT gateway's address instead which should forward this on. But if this is taken care of in a standard way, UAS on the callee side would try to respond to the NAT address which is shown in the Via header. In this case we may need to have some sort of application level gateway functionality within the NAT box. If the call that is being established uses a SIP proxy server, then the call leg is completed using "received" and "rport" parameter [11] in the via header which is added by the proxy before forwarding the Invite request further. However the NAT implementation in our testbed does not change the port associated with the source address. Reference [11] also takes care of establishing the SIP calls when either the caller or callee is behind the NAT box. In case we are not using any proxy between the caller and callee, rather the session is established using direct signaling, then we may have to revert to the case where a SIP based application level gateway (ALG) takes care of the SIP calls or some other approach such as STUN [12] can also be used in conjunction with sipc.

C. Multiple IP registrations

A SIP user is usually identified with a unique URI scheme such as sip:bob@abc.com. A SIP user agent on a single terminal can be associated with multiple IP addresses, if there are multiple IP addresses active on the same terminal. At any particular instance a mobile terminal may have multiple IP addresses assigned to each of its physical interfaces. Both the addresses could be active in some situation unless the host decides a policy table to make one of the interfaces active such as in the case of 802.11b vs. CDPD. If dynamic DNS is being used, any new address would update the associated DNS table to reflect the host's IP address change in the DNS database. In such cases we need to have a way to register the active interface with the SIP server. But in some instances both the interfaces could be active, and depending upon the destination route metrics or local policy decision one particular interface is chosen for transmission or reception of the traffic. Just like DNS resolves to multiple IP addresses, unique URI scheme would also resolve to multiple IP addresses and one of these interfaces would be picked up.

D. Registration of the active IP address

Usually the IP address is sent in the Contact header of the registration message. If both the interfaces are up, contact header can use two IP addresses instead of one in a single registration message. Otherwise, if the end host has the ability to choose the active interface, it can choose the right interface and register that interface. Thus SIP user agent would need to be able to send both the IP addresses in its contact header instead of one, which is usually done in a SIP ua case. In some cases registration can also be done using the source IP address of the SIP UA, rather than relying on the contact header. This is often the case when the SIP UA is behind the NAT.

E. De-registration

When the client moves from one access network to other one and obtains a new IP address, it would need to de-register the IP address assigned to the interface which may not be active currently, but this address can be in IP routing table itself. In this scenario a local SIP server in the separate access network is chosen to take care of third party de-registration, since it may not have connectivity to the old access network. In some cases even if there is dead registration because of inactive interface, it can be taken care of by a forking proxy. However there may be a potential conflict, if the previous IP address which got de-registered gets re-used by another potential client in between. However this kind of problem is easily avoidable in an environment where IP address is distributed using DHCP. If both the networks share a common SIP server, then it is possible to send a list of old IP addresses for de-registration, but in this case SIP UA would need to keep a listing of all the old IP addresses in its database.

F. User Agent with multiple IP addresses

When both the interfaces are active, the interface that gets picked up is determined by the SIP user agent's priority for a particular interface. Thus associating a SIP user agent with a specific interface is implementation specific. In some cases even if both the interfaces are up, SIP user agent should be able to pick up the interface based on kind of traffic it is expecting on an interface.

G. Session Handoff

In case of heterogeneous access networks, several methods can be used so as to trigger a SIP based session handoff. Triggering mechanism can be based on channel change, QoS of the received traffic, server based advertisement, or layer 3 router solicitation.

However it is an important criterion to initiate the handoff based on layer 2 signal strength in case of homogeneous access networks. In case of heterogeneous access, other methods need to be devised to make sure that the client is connected to the right interface. In case of availability of access to both the kinds of networks, a policy needs to be determined on the client itself as to which network to connect to at any paricular moment. In a typical handoff scenario for homogeneous access such as described in [1], a simple Re-Invite is sent from the mobile host to the correspondent host and the new IP address is given as the contact addr. In the case of heterogeneous access the Re-Invite, source address of the Invite request may be used by the user agent of the corresponding host (CH), so as to avoid the complicacy associated with dual contact address in a Re-Invite message.

IV. CONCLUSION AND OPEN ISSUES

This paper illustrates few scenarios for supporting SIP based sessions in local area and wide area networks where the end-client has heterogeneous access capability. While discussing several architecture alternatives for local Area and wide Area, it has also pointed out several issues and the solutions thereof which should be taken into account to support mobility. It also presented some of the experiments carried out while moving in a heterogeneous environment.

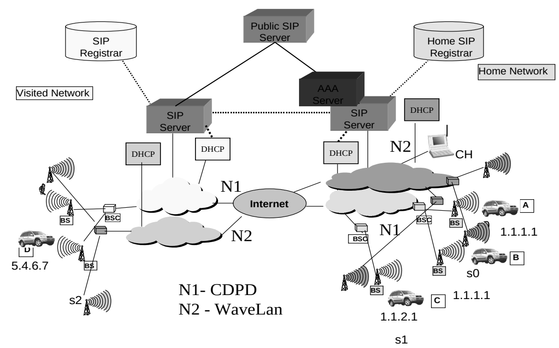

Fig. 6. Mobile Internet Heterogeneous Access Network

V. ACKNOWLEDGEMENT

Authors would like to acknowledge Ted Lu, Sunil Madhani, Val Glinskiy, Chiao-Wei, Yosihiro Suwa for their contribution to the project and Xiaotao Wu for discussion towards sipc implementation.

REFERENCES

[1] Elin Wedlund, H. Schulzrinne , " Mobility Suport using SIP" in *IEEE/ACM Multimedia conference WOWMOM 1999*

[2] H. Schulzrinne, Elin Wedlund, "Application-Layer Mobility using SIP", Mobile Computing and Communications Review, vol 1

[3] Mark Handley, Eve Schooler, H. Schulzrinne, Jonathan Rosenberg, "Session Initiation Protocol", RFC 2543

[4] F. Vakil et. al, "Host Mobility Management Protocol for 3G network ". Internet Draft Work in progress

[5] A. Dutta, F. Vakil, S. Baba, H. Schulzrinne et al, "Application Layer Mobility Management Scheme for Wireless Internet", 3G Wireless 2001, San Francisco.

[6] Tao Zhang, Jyh-Cheng Chen, Prathima Agrawal, "Distributed Soft Handoff in All-IP Wireless Networks" 3G Wireless 2001, San Francisco

[7] Henning Schulzrinne, "SIP registration", draft-schulzrinne-sip-register-00.txt, IETF Draft Work in Progress

[8] C. Perkins, "IP mobility support", RFC 2002, IETF October 1996"

[9] Pat Calhoun et al, "AAA requirements for IP Telephony/Multimedia" SIP WG, IETF, Work in Progress

[10] Caroline Lebre et al, "Handover between heterogeneous IP networks for Mobile Multimedia Services" BT Technical Report

[11] J. Rosenberg et al, "SIP Expensions for NAT traversal" draft-ietf-sip-nat-01.txt, IETF work in progress.

[12] J. Rosenberg et al., "Simple Traversal of UDP through NATs"draft-rosenberg-midcom-stun-00.txt, Nov. 2001, work in Progress IETF.

[13] S. Denie, "NAT friendly application design guidelines, Internet Draft, Mar 2001, work in progress.