Fast Handoff for Hierarchical Mobile SIP Networks

F. Chahbour, N. Nouali, K. Zeraoulia

***Abstract*—**Although it was originally developed as an application layer signalling protocol, SIP offers mobility management capability. However SIP Terminal Mobility suffers from considerable handoff latency which is unsuitable for real time applications and frequent local movements (intra domain handoffs). Hierarchical Mobile SIP (HMSIP) was then proposed as a solution to reduce the SIP handoff delay and to provide a complete micro-mobility solution for SIP environments. Unfortunately HMSIP handoff delays remain too high for real time multimedia applications which require handoff delay between 50ms and 200ms to guarantee the Quality of Service. In this paper we propose to reinforce the HMSIP by integrating the Predictive Address Reservation (PAR) mechanism to get better handoff performance. A simple integration of PAR-SIP with HMSIP allows not only reducing handoff delay but also signaling overhead.

***Keywords*—**Fast handoff, Multimedia applications QoS, SIP Mobility management, Wireless communication**.**

I. Introduction

THE tremendous efforts seen in the last years towards the development of wireless mobile networks as well as the increase of mobile devices performance are making viable the mobile Internet infrastructure for multi-media services and the deployment of real time multimedia applications. These applications concern all modern life aspects: peer-to-peer communications, video conferences, remote learning, etc. Several standard bodies such as the IETF [1] and the 3GPP [2] are working on the specifications of an all IP wireless network that allows roaming users to access integrated data, voice, and multimedia services. A wide variety of mobility management schemes have been proposed working at different layers of the protocol stack. However, these schemes are not suitable for all types of applications. Thus, dedicated schemes capable of acting more semantically must be developed; e.g., in the case of HTTP or FTP applications, handoff latency is not of vital importance (waiting one or two seconds extra when downloading a web page is not critical). But for real-time media, latency and packet losses are extremely important and even a small disturbance can make a media stream unintelligible.

The mobility requirements can be satisfied by a simple application layer mobility management protocol. Although the Session Initiation Protocol (SIP), originally developed as an application layer signalling protocol, offers mobility management capability, however, SIP is not suitable for supporting streaming media with stringent delay requirements [3]. Therefore the SIP handoff delay needs careful investigation to offer reasonable delays for real time applications.

Manuscript received March 30, 2005.

This work is supported by the Ministère de l'Enseignement Supérieur et de la Recherche Scientifique.

F. Chahbour is with CERIST, Mobile Computing Department, Rue des 3 Frères Aissou, Ben Aknoun, 16030 Algiers, Algeria (Tel: 213 21 20 25, Fax: 213 21 91 21 26, e-mail: cfairouz@yahoofr).

N. Nouali, also is with CERIST (e-mail: nanouali@yahoo.com).

K. Zeraoulia , is with CERIST (e-mail: Zekhaled@yahoo.fr).

II. Mobility Management using SIP

A. Review of the session initiation protocol

The Session Initiation Protocol (SIP) [4] is a simple scalable, text-based application layer signalling protocol. It was originally specified by the MMUSIC (Multiparty Multimedia Session Control) working group of the IETF in the RFC 2543 in March 1999 and is getting widely accepted for Internet multimedia and telephony services. Communications in SIP can be either unicast or multicast and participants are end-users, media servers (audio, video…), SIP signalling servers, or gateways towards other networks. Various methods are defined in SIP, viz. INVITE, ACK, BYE, OPTIONS, CANCEL, and REGISTER. INVITE is used to invite a user to a session. The message body contains the session description with the address at which the host wants to receive the media stream. ACK is for acknowledging the INVITE request and BYE is for releasing a call. OPTIONS provide the feature for querying the server about the capabilities. CANCEL is for cancelling a pending request and REGISTER is for registering a host with a SIP server. SIP end users are addressed using SIP URLs. The format of a SIP URL is similar to that of an e-mail address, user-info@domain. Session establishment between two end users is shown in the Fig. 1. SIP is built on Client/Server architecture and uses a number of logical entities, namely User Agents, redirect servers, proxy servers and registrars.

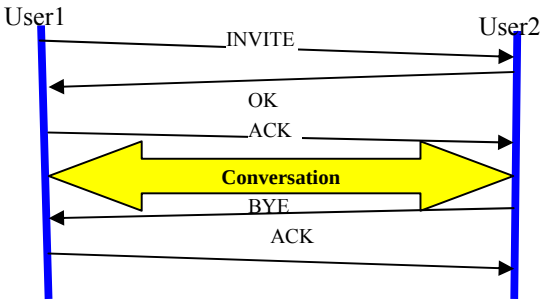

Fig. 1 SIP session establishment between two users.

B. SIP Mobility support

SIP was initially designed to be a call control protocol; therefore it does not cater for Terminal Mobility. SIP can handle User Mobility in an intrinsic way independently of the device being currently used as it uses logical addressing to identify SIP users. To be able to manage Terminal Mobility, SIP needs to have some minor extensions. Terminal Mobility impacts SIP at three stages, pre-Call, mid-call and to recover from network partitions [5]. In this paper we focus only on pre-call and mid-call mobility as they are the main issues in mobility management namely location and handoff management respectively. The pre-call Mobility represents the case when the Mobile Host (MH) changes its attachment point prior to receiving or making a call. In this case MH acquires a new IP address before it decides to initiate or to accept new SIP calls. MH must update its location information; this is ensured simply by the re-registration of the MH with its home registrar each time it obtains a new IP address (Fig. 2).

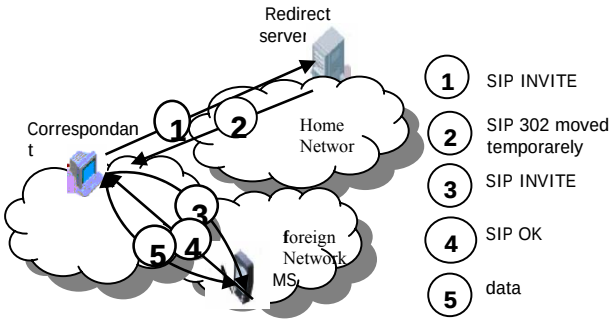

Fig. 2 pre-call mobility.

If the Mobile Host changes its location (IP address) during a traffic flow, it must inform its correspondent (CH) about its current location by sending to him another INVITE request, as shown in Fig. 3. The INVITE request contains an updated session description with the new IP address [6]. These extensions allows SIP to manage Terminal Mobility but they are not sufficient in high mobility environments as they require end-to-end signaling exchange to re-establish a communication after a handoff (the mobile host should inform its new IP address to both home registrar and the correspondent during handoff). This session re-establishment results in long update delays and high signaling overhead in the backbone network and consequently, considerable handoff latency which is unsuitable for real time applications and frequent local movements (intra domain mobility). A hierarchical based registration scheme is proposed in [5] to accelerate the registration process and to reduce backbone signaling. This idea is then adopted by Hierarchical Mobile SIP (HMSIP) [7] aiming to offer an efficient micro-mobility solution for sip networks. Although HMSIP offers low handoff delay, it remains too high for real time applications. In the following sections we give an overview of the HMSIP scheme and we try to explain why the HMSIP handoff delay is not suitable for the real time applications and finally we propose to reinforce the HMSIP scheme by the Predictive Address Reservation mechanism [8] to get better handoff performance.

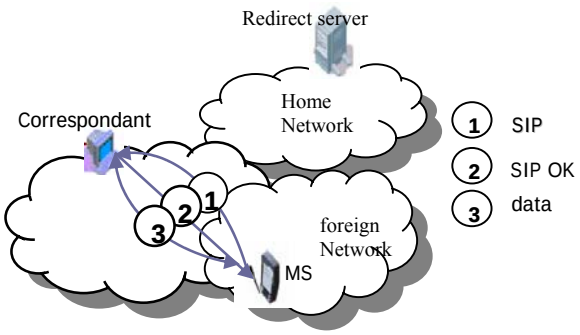

Fig. 3 SIP handoff (Mid-call Mobility).

III. Overview of HMSIP

HMSIP aims to reduce signaling overhead and handoff delay inside the same domain in a SIP environment. It follows the general regional registration approach found in various intra-domain mobility schemes[1] (MIP-RR, HMIPv6) and builds on SIP hierarchical registration proposed in [5]. HMSIP defines a new entity responsible for locally handling intra-domain mobility called SIP Mobility Agent (SIP MA). SIP MA is a domain border router enhanced with the functionality of SIP Proxy and SIP Registrar. HMSIP allocates two IP addresses to the Mobile Host (MH), a local address (LA) reflecting the MH current point of attachment allocated by the serving access router and a global domain address (DA) used to uniquely identify the MH during its roaming inside the same domain, and is allocated by the SIP MA. The SIP MA maintains and manages a data base of soft state mappings between the SIP URI, the DA and the LA for each mobile that roams inside the domain. HMSIP caters for intra-domain mobility and relies on Mobile SIP[2] for handling inter-domain mobility.

A. Pre-call Mobility (registration procedure)

After powering on inside a visiting domain, the MH is allocated a new LA from the serving access router and a new DA from the SIP MA Fig. 4. The MH registers its new

[1] We prefer to use intra-domain mobility and inter-domain mobility terms instead of macro and micro-mobility to avoid any conflict in the terminology concerning the granularity of mobility.

[2] Mobile SIP refers to the extensions proposed to SIP to support Terminal Mobility.

location with the visited network by sending a SIP REGISTER message with the contact field at LA value towards the serving SIP MA. The SIP registrar entity of the SIP MA associates LA to the SIP URI. This completes the regional registration procedure. In case of MH is newly arrived to the domain, a new DA is allocated to it by the domain SIP MA. The SIP Registrar entity store a mapping between SIP URI, LA and the DA allocated and then instructs the SIP proxy to initiate a SIP REGISTER message, containing the DA information, towards the MH SIP Home Registrar. The latter creates a mapping between the SIP URI and the DA in its database. Incoming SIP calls towards the

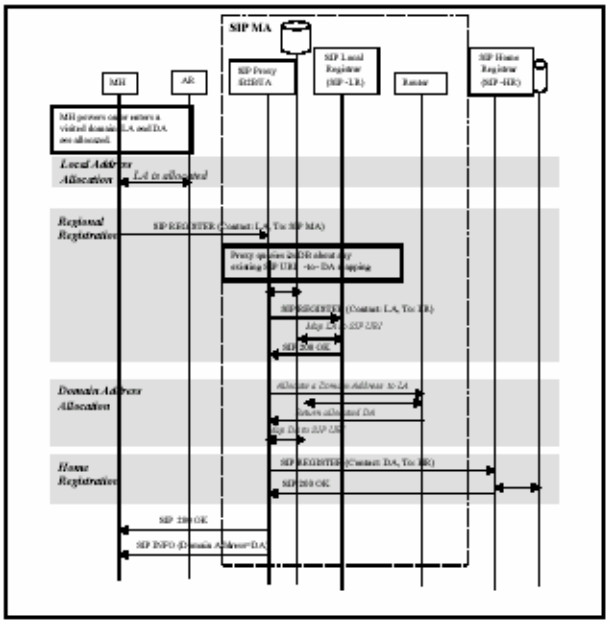

Fig. 4 Regional and home registration.

MH are directed to the DA which will be the static address of MH inside the roaming domain and used as the source/destination address outside the domain. Encapsulation of SIP calls and data packets takes place inside the domain between the SIP MA and the current location of the MH (LA).

B. Mid-call mobility (handoff)

When the MH hands off to a new access router belonging to the same domain, it obtains a new LA while the DA remains the same. The MH performs only regional registration to reflect its new location. The mobility related signaling cost inside the domain boundary is optimized as the home registration does not take place (DA hasn't changed); resulting in low handoff latency. If the MH changes the domain, it performs both regional and home registration procedures using the new allocated LA and DA. After this the session re-establishment for on-going sessions takes place as proposed by the Mobile SIP approaches (fig. 4).

IV. Fast Hierarchical Mobile SIP

The regional registration procedure proposed by HMSIP offers low handoff delay, but it remains too high for real time applications which require handoff delay between 50ms and 200ms to avoid service degradation. HMSIP can't offer such delays even by eliminating home registration procedure because in addition to time necessary for *Link layer handoff,* the MH must detect movements in the network layer (it must discover that it is on a new sub-network using Router Advertisement (RA), which a router periodically broadcasts), obtain a new IP Address, re-configure its own network interface and some network parameters to communicate again, and finally update its location information by sending the *SIP REGISTER* message to the local registrar. Among the above steps, address allocation takes the most part of the handoff delay in the all approaches proposed to support mobility using SIP including the HMSIP scheme. When DHCP is used, address allocation takes more than a second. Even though DRCP [9] reduces the address allocation time, a handoff still requires a few hundred milliseconds [8]. Address acquisition becomes then the main cause in degrading the service quality in real-time applications. For this reason and in order to reduce handoff delay at a value suitable for real time applications, we propose to reinforce HMSIP with a Predictive Reservation Address mechanism. A simple integration of the PAR

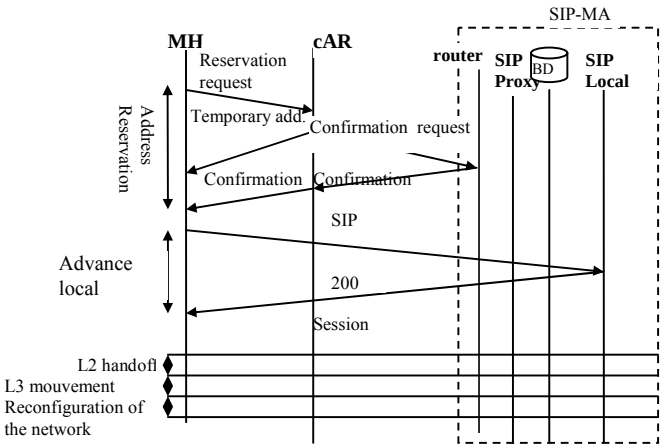

Fig. 5 FHMSIP handoff flow.

(Predictive Address Reservation) mechanism with HMSIP allows not only reducing handoff delay but also signaling overhead. Fast Hierarchical Mobile SIP (F_HMSIP) proposes to perform address allocation and registration procedures before a link layer handoff to reduce HMSIP handoff latency. This can be achieved by employing the movement detection scheme using link layer information. Fig. 5 presents the order of an intra-domain handoff in F_HMSIP. The base idea is to allocate a new IP address to the MH and allow him to re-REGISTER with its serving SIP-MA (regional registration) before the link layer handoff triggering. In fact, address acquisition and SIP registration procedures are executed in parallel within an existing SIP session.

V. Editorial Policy

A. Network layer movement detection

To obtain a new IP address in advance, MH must forecast its network layer movements before receiving RA (Router Advertisement) from a router. Network layer movements detection is done as follows : When the Signal to Noise Ratio value (SNR) of the current access point (AP) falls bellow the Cell Search Threshold, the MH starts to search another reachable AP with higher SNR than the current one (active scan). After the selection of such an AP and using an AP list, MH can verify whether the predictive AP belongs to the same access router or to a neighbor one. If the predictive AP belongs to the same access router, a link layer handoff is performed using its address MAC, else MH starts the Address Reservation Process (detailed in the section below). To the SIP-MA entity is added a data base to manage the domain access routers information. Each AR contains a neighbor AR information table, updated periodically by the SIP MA entity, to allocate a temporary address and to create an AP list for the MH.

B. Address Reservation

The address reservation process is carried out in the same manner as in [8] with some modifications adapting it to the hierarchical environment. When a MH detects network layer handoff, it sends a reservation request to the current AR, cAR, as shown in the fig.6. The cAR sends to MH a temporary local address, topologically correct for the new sub-network, and asks the SIP-MA router to validate this address. The SIP-MA router verifies that the allocated address is not used and replies with a validation message. The cAR then authorizes MH to use the allocated address. After this, MH sends a SIP REGISTER message to the SIP registrar of the serving SIP-MA and receives the OK message as a reply before the triggering of the link layer handoff.

All mobility support using SIP proposals, privilege the re-establishment of the ongoing sessions (the SIP re-INVITE message) against the location information update after a handoff. But, in HMSIP environment there is no need to re-establish the ongoing SIP sessions after an intra-domain handoff, because the SIP-MA is informed about the new LA of the MH via the regional registration procedure and is able to redirect the data path to the new location. Therefore there is also no need to re-establish in advance an ongoing session in FHMSIP. This allows reducing the number of procedures to be performed before the link layer handoff triggering and therefore the increase of the probability so that the tasks to be performed before the link layer handoff terminate before the triggering of the link layer handoff (Fig. 6).

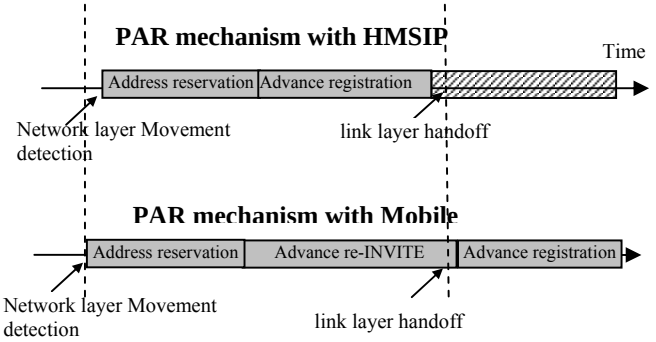

Fig. 6 An example of link layer handoff triggering in both HMSIP and Mobile SIP.

VI. Conclusion

Both Hierarchical Mobile SIP and Predictive Address Reservation schemes have been proposed to reduce the handoff latency in their own ways. The Hierarchical Mobile SIP allows to reduce handoff latency by reducing the delay concerned with the location update in which the Mobile Node registers locally with its serving SIP_MA registrar rather than the home SIP registrar and the correspondent that are typically far away. On the other hand, the Predictive Address Reservation uses link layer information for earlier movement detection and address configuration for the new point of attachment so as to minimize the disruption of the services during the handoff. The integration of the PAR mechanism with the Mobile SIP networks [8] reduces significantly the handoff latency but its integration with the HMSIP environment provides better handoff performance and minimizes the signalling overhead by eliminating the ongoing SIP session re-establishment phase (the send of the re-INVITE message).

References

[1] The Internet Engineering Task Force, http://www.itef.org.
[2] The 3rd Generation Partnership Project (3GPP), http://www.3gpp.org.
[3] N. Banerjee, k. Basu and S.K. Das, "Hand-off Delay Analysis in SIP-based Mobility Management in Wireless Networks," in *Proc. Of the IEEE Int. Parallel and Distributed Processing Symposium, IPDPS'03.*
[4] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg, SIP: Session Initiation Protocol, *RFC 2543, Internet Engineering Task Force*, March 1999.
[5] H. Schulzrinne. E. Wedlund, "Application Layer Mobility Using SIP," *Mobile Computing and Communication Review, (MC2R)*, vol 4, n°3, 2000.
[6] E. Wedlund. H. Schulzrinne, "Mobility support using SIP' WOWMOM'99," *in Proc. of Second ACM International Workshop on Wireless Mobile Multimedia*, Seattle, Washington, USA, 1999.
[7] D. Vali, S; Paskalis, A. Kaloxylos and L. Merakos, "An Efficient Micro-mobility Solution for SIP Networks," *IEEE globecom 2003*, San Francisco, CA, USA, 2003.
[8] W. Kim, M. Kim, K. Lee, C. Yu and B. L. Link. "Layer Assisted Mobility Support Using SIP for Real-time Multimedia Communications," ACM MobiWac, 2004.
[9] A. Mcauley Subir Das, Sunil Madhani, Shinichi Baba, Yasuro Shobatake, *Dynamic Registration and Configuration Protocol, draft-itsumo-drcp-01.txt*, IETF July 2000.