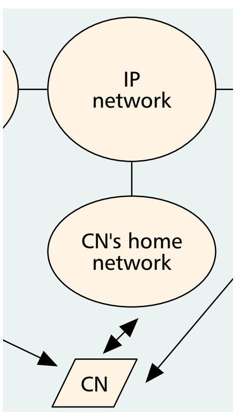
MOBILITY MANAGEMENT FOR VOIP SERVICE: MOBILE IP VS. SIP

TED TAEKYOUNG KWON AND MARIO GERLA, UCLA
SAJAL DAS, UNIVERSITY OF TEXAS AT ARLINGTON
SUBIR DAS, TELCORDIA TECHNOLOGIES, INC.



Wireless Internet access has gained significant attention as wireless/mobile communications and networking are becoming widespread. The voice over IP (VoIP) service is likely to play a key role in the convergence of the IP-based Internet and mobile cellular networks.

ABSTRACT

Wireless Internet access has recently gained significant attention as wireless/mobile communications and networking become widespread. Voice over IP service is likely to play a key role in the convergence of IP-based Internet and mobile cellular networks. In this article we explore different mobility management schemes from the perspective of VoIP services, with a focus on Mobile IP and Session Initiation Protocol. After illustrating the signaling message flows in these two protocols for diverse cases of mobility management, we propose a Shadow Registration concept to reduce the interdomain handoff (macro-mobility) delay in VoIP service in mobile environments. We also analytically compute and compare the delay and disruption time for exchanging signaling messages associated with the Mobile IP and SIP-based solutions.

INTRODUCTION

Recently, mobility support for Internet access has created significant interest among researchers as wireless/mobile communications and networking proliferate, especially boosted by the widespread use of laptops and handheld devices (e.g., PDAs and handsets). Considering the various wireless access technologies — 802.11, Bluetooth, second/2.5/third generation 2G/2.5G/3G cellular, and so on — and their complementary features, we expect that these pocket-sized mobile handheld nodes are going to be equipped with multiple wireless communication interfaces. Under this configuration and environment, the mobile node would be able to choose the most suitable interface for specific applications. This phenomenon is often called *wireless technologies convergence*. In such an environment, one of the crucial issues is how to support seamless mobility to mobile nodes.

Another important trend over the past few years is the emergence of voice over IP (VoIP) service and its rapid growth. Even though the original VoIP protocols and applications did not consider the mobility of the end nodes, there have been ongoing research efforts to support mobility in the current VoIP protocols. In providing the VoIP service in wireless technologies convergence, the most viable concern is the amount of disruption time to process the handoff of an ongoing VoIP call (or session).

The mobility itself can be largely divided into three types: roaming, macromobility, and micromobility. *Roaming* is the movement of the user in absence of the Internet connectivity. This roaming is usually triggered when a mobile node initiates the Internet connectivity. *Macromobility* and *micromobility* are the change of point of attachment with ongoing Internet connections and thus normally accompany the handoff. The macromobility is related to the movement of the user from one administrative domain to another. In such a case, the relevant domains must collaborate to ensure seamless connectivity to the moving user. Obviously, in wireless technologies convergence, macromobility will be invoked frequently since the different wireless networks are likely to be different administrative domains. Micromobility concerns the user's movement inside a given domain, which involves intradomain (subnet-level) handoff. A well defined *mobility management framework* or scheme should deal with all three types of mobility, especially seeking to reduce disruption in handoff.

Currently, there are two basic approaches to support mobility in VoIP services. The first one seeks to solve mobility in the network layer by using Mobile IP and related proposals. Although Mobile IP is not directly related to VoIP applications, mobility support for VoIP service can be realized via Mobile IP. The other approach is to solve the mobility problem in the application layer by augmenting existing VoIP protocols such as H.323 and Session Initiation Protocol (SIP). In our opinion, telecom-based H.323 is too complicated to evolve in practice. Therefore, we take into consideration only SIP in this article. Our main theme here is to compare the IP layer solution (Mobile IP) with the application layer solution (SIP) to support mobility in VoIP services.

Another aspect that accompanies macromobility is authentication, authorization, and accounting (AAA), which applies to both of the above solution approaches. A user of a mobile

 1070-9916/02/$17.00 © 2002 IEEE

node must identify (and authenticate) him/herself and interact with the AAA server of his/her home network. This AAA resolution should be performed not only when the user moves into the visited network but also when the user initiates Internet connectivity in the home network. As a number of diverse wireless access technologies and networks will be deployed in the near future, it is likely that a mobile node will frequently hand off between wireless networks of different service providers (i.e., different administrative domains). The problem is that the mobile node should resolve the AAA issue whenever it hands off (and changes the point of attachment) between different administration domains. Note that in the early stage of generic packet radio service (GPRS)/Universal Mobile Telecommunications System (UMTS) deployment, the handoff between GPRS and UMTS networks may not involve a change of the mobile node's IP address [1]; however, we believe that this is a temporary phenomenon.

To provide seamless VoIP service in such a challenging heterogeneous wireless/mobile communication environment, delay or disruption in dealing with macromobility and micromobility must be minimized because noticeable disruption during a voice conversation will make VoIP service users unhappy. After discussing the Mobile IP and SIP solutions for mobility management (note that Mobile IP is not designed only for VoIP), we will propose a *Shadow Registration* method to reduce the time to process interdomain handoff in both approaches. The key idea in *Shadow Registration* is to establish a registration status in the neighboring administrative domains a priori anticipating possible handoffs when the user registers in the given wireless/mobile network. We analytically derive various kinds of delay involved in both approaches and finally compare them.

The rest of this article is organized as follows. The Mobile-IP-based solution is discussed as well as the SIP-based approach. The Shadow Registration concept is introduced, and signaling message flows are illustrated. The analytical comparison of delay/disruption with a simplified network model is made, and concluding remarks are offered in the last section.

Network Layer Solution: Mobile IP

While there is some consensus that Mobile IP [2] will be used to manage roaming and macromobility in wireless/mobile access to the Internet, there have been a number of proposals for the micromobility issue, such as Regional Registration [3] and Cellular IP [4]. Since it takes considerable time to exchange a registration message between different mobility agents, most of these proposals have considered a special agent node in each administrative domain, which accommodates local handoff within the administrative domain without contacting the home agent (HA) of the mobile node. Here, we adopt the Regional Registration mechanism because it has a similar concept of operation as Mobile IP. (However, other micromobility proposals can also be adopted.) This section briefly summarizes Mobile IP and Regional Registration, and then discusses how to handle AAA resolution. We adopt Diameter for the AAA protocol since the current Internet Engineering Task Force (IETF) standardization efforts promote its use for Mobile IP authentication (e.g., [5]). We also illustrate the signaling flows in Mobile IP.

Mobile IPv4 Overview

Mobile IPv4 seeks to solve the mobility problem by two addresses: home address and care-of address (CoA). When a mobile node (MN) stays connected in its home network, it is reachable by its invariant home address. Each time the MN connects to a foreign network, it obtains a temporary address, the CoA, which is only valid for the time the MN will stay connected to this foreign network. The MN will then be reachable via both its home address and the CoA. There are two mobility agents that accommodate the MN: the foreign agent (FA) in the visited network and the HA in the home network. Whenever the MN obtains the CoA from the FA, it must inform its HA of the obtained CoA; this is the registration process. After this registration, the HA can forward the packets (originally sent to the MN's home address) to the FA by tunneling.

The basic working of Mobile IP leads to asymmetric routing; the packets from the correspondent node (CN) to the MN are first captured by the HA and tunneled to the MN, while the MN sends packets to the CN directly. To improve the efficiency of routing, Mobile IP defines the concept of mobility binding, which allows the CN to encapsulate packets directly to the current CoA of the MN. To implement mobility binding, the CN maintains a binding cache to store the mobility bindings for one or more MNs. The *Binding Update* message is used for the HA to inform the CN that the MN has changed its CoA [6].

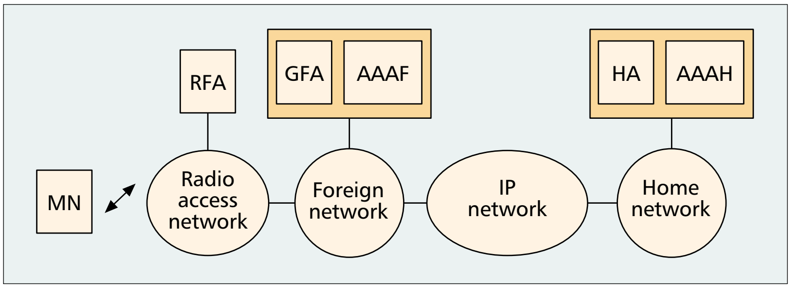
When an FA receives a tunneled packet for an MN that is not in its visitor list, it may deduce that the tunneling node has an out-of-date binding cache entry. If the FA has a mobility binding for the MN in its own binding cache, it should send a Binding Warning message to the HA of the MN and retunnel the packet to the CoA in the cache entry. On the other hand, if the FA has no binding cache entry for that MN, it sends the packet to the home address of the MN. The packet will be trapped by the HA which should encapsulate it to the current CoA of the MN.

Additionally, we assume that *Smooth Handoff* [7] is performed; that is, the old FA and the new FA can exchange the *Binding Update/Acknowledgment* message when the MN obtains a new CoA due to handoff. The new FA sends the *Binding Update* message to the old FA to inform the new CoA of the MN. When the old FA receives the *Binding Update* message, it updates the binding cache entry of the MN and then replies with the *Binding Acknowledgment* message to the new FA, if requested. We do not take into consideration buffering packets for the MN in the old FA.

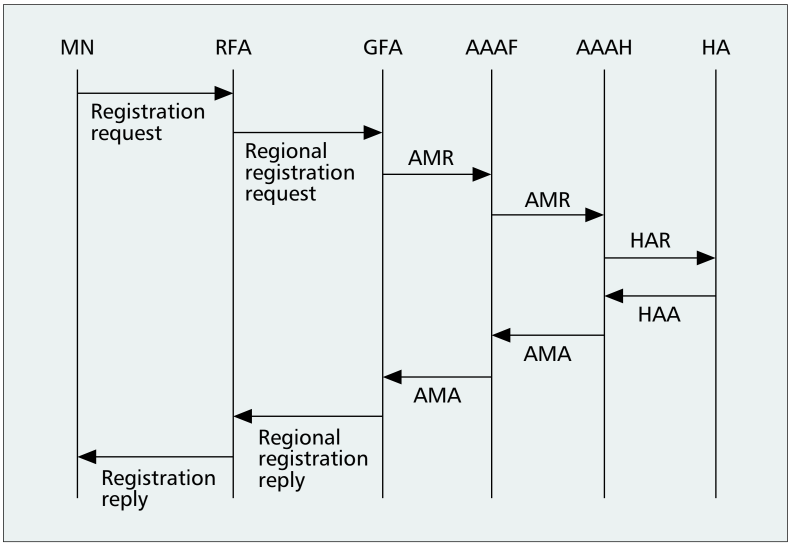
Regional Registration

Using Mobile IP, an MN registers with its HA each time it changes its CoA. If the distance between the visited network and the home net-

When a mobile node stays connected in its home network, it is reachable by its invariant home address. Each time the MN connects to a foreign network, it obtains a temporary address, the CoA, which is only valid for the time the MN will stay connected to this foreign network.



■ Figure 1. *Mobile IP architecture.*



■ Figure 2. *Mobile IP registration.*

work of the MN is large, the signaling delay for these registrations may be long. Gustafsson *et al.* [3] proposed a solution for performing registrations locally if the MN changes its CoA within the visited domain. This is called Mobile IP *Regional Registration*.

When an MN first arrives at a visited domain, it performs a registration with its HA. At this registration, we assume that the home network of the MN generates a registration key [8] for the MN. This registration key is distributed to the MN and visited domain, and can be used for authentication of regional registrations.

If the visited domain supports Regional Registration, the CoA that is registered at the HA is the publicly routable address of a *gateway foreign agent* (GFA). This CoA will not change when the MN changes FA under the same GFA. When changing GFA, the MN must perform a normal registration to its home network. On the other hand, when changing FA under the same GFA, the MN performs a regional registration within the visited domain. There are two new message types for this regional registration: *Regional Registration Request and Regional Registration Reply*.

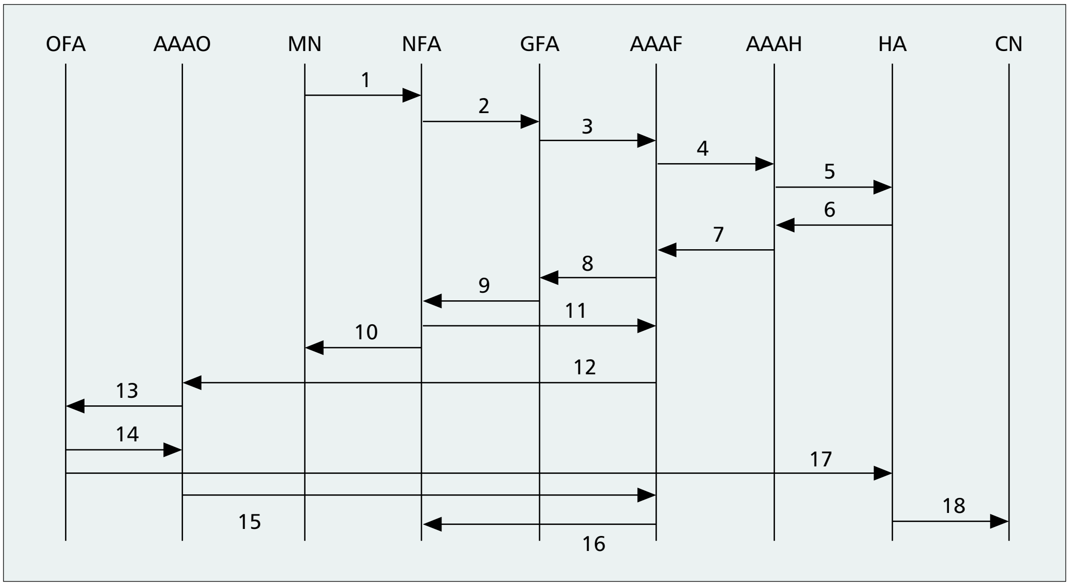
Mobile IP Message Flow

As stated earlier, mobility management should handle the AAA issue in regard to mobility in Internet service. Currently an AAA protocol such as RADIUS is used within the Internet to provide authentication services for dialup computers. However, the current IETF promotes the use of the Diameter protocol for authenticating mobile nodes during Mobile IP registration, which is adopted in this article. Mobile IP requires strong authentication services between the MN and its HA. Once the MN shares a security association (SA) with its home AAA server (AAAH), it is also possible to use that SA to create derivative SA between the MN and its HA, and again between the MN and the FA currently offering connectivity to the MN. The establishment of this SA lengthens the registration time in Mobile IP because security associations must be made among all entities (FA, HA, MN) involved in the process of registration.

The entities in the above-mentioned Mobile-IP-based approach are depicted in Fig. 1. In the foreign network, there is the regional foreign agent (RFA), which is the local FA that accommodates the MN in the subnet. The AAA server in the foreign network is denoted AAAF, while the AAA server in the home network is denoted AAAH. Since we choose regional registration, FAs are organized as a two-level hierarchy: RFA for each subnet and GFA for each foreign network. We assume that each radio access network (RAN) is an IP subnet, which consists of one or more base stations (or access points). Also, each foreign network is an administrative domain, and we assume there is only one GFA per administrative domain.

Figure 2 shows the message flow for initial registration at a foreign network. The MN sends the *Registration Request* message to the RFA. Then the RFA sends the *Regional Registration Request* message to GFA. The GFA then modifies that message into the *AA-Mobile-Node-Request* (AMR) message and sends it to the AAAF. The AAAF possibly adds or modifies some optional attribute value pairs (AVPs) and forwards this message to the AAAH of the MN. The AAAH generates a Home Agent Request (HAR) message and sends it to the HA. The HA processes this registration message and then responds with a Home Agent Answer (HAA) message. After receiving the HAA message, the AAAH generates and sends an AA-Mobile-Node-Answer (AMA) message to the AAAF. This AMA message is possibly modified and forwarded to the GFA. (The messages AMR, HAR, HAA, and AMA are Diameter-compliant and detailed in [5, 9].) Then the GFA sends the *Regional Registration Reply* message to RFA. Finally, RFA returns the *Registration Reply* message to the MN.

In the case of intradomain handoff, when the MN changes the point of attachment between FAs, it sends the *Registration Request* message to the new RFA (NFA). When the NFA receives this message, it modifies the message into the *Regional Registration Request* message as described above. In addition, the NFA also sends the *Binding Update* message to old the RFA (OFA) to inform the OFA of the new CoA of the MN. If requested, the OFA replies with *Binding Acknowledgment* message to confirm the update of binding cache entry on the MN. We assume that there is already a security association between the RFAs (NFA and OFA in this figure) in the same administrative domain, so the *Binding Update/Acknowledgment* message exchange is possible without additional authenti-



■ **Figure 3.** *Interdomain handoff in Mobile IP.*

cation in this scenario. Also, in this case, the *Binding Update* message to the CN is not necessary because the address of GFA (which is unchanged) is registered in the HA of the MN.

Figure 3 shows the signaling message flow for interdomain handoff. Messages 1–10 are exactly the same as in Fig. 2. However, in this case the *Binding Update* and *Binding Acknowledgment* messages should be authenticated since this message exchange is performed in different domains. Note that after the MN is authenticated (message 9), the NFA starts signaling for the *Binding Update*. Thus, messages 11, 12, 15, and 16 are the Diameter-compliant messages (AMR, AMA) that contains *Binding Update/Acknowledgment* information [10], and messages 13 and 14 are normal Binding Update/Acknowledgment messages. Here, AAAO is the AAA server of the old foreign network to which the OFA belongs, while the AAAF is the AAA server of the new foreign network to which the NFA belongs. Message 17 is the *Binding Warning* message, and message 18 is the *Binding Update* message.

Application Layer Approach: SIP

We first give an overview of the SIP architecture and then discuss how to augment mobility to SIP. Signaling message flows for SIP registration are also illustrated. We consider the configuration where a combination of SIP, DHCP, and Diameter (as an AAA protocol) is used to support mobility for SIP users.

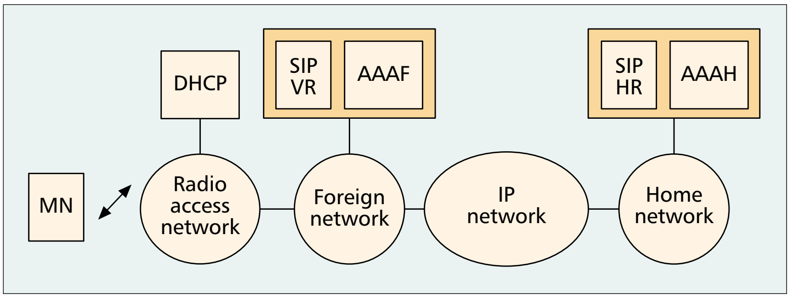
SIP Overview

SIP [11] is an application layer protocol used for establishing and tearing down multimedia sessions, both unicast and multicast. It has been standardized within the IETF for the invitation to multicast conferences and VoIP services.

The SIP user agent has two basic functions:

- Listening to the incoming SIP messages
- Sending SIP messages upon user actions or incoming messages

The SIP proxy server relays SIP messages so that

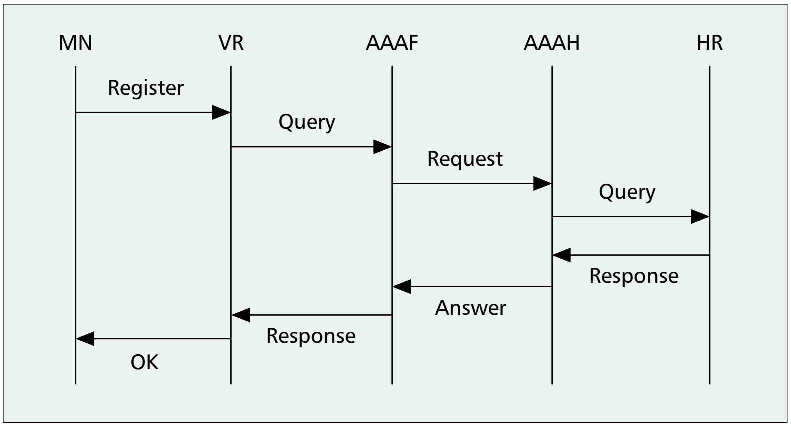


■ **Figure 4.** *SIP architecture.*

it is possible to use a domain name to find a user, rather than knowing the IP address or name of the host. A SIP proxy can thereby also be used to hide the location of the user. On the other hand, the SIP redirect server returns the location of the host rather than relaying the SIP messages. This makes it possible to build highly scalable servers, since it only has to send back a response with the correct location. The SIP redirect server has properties resembling those of the HA in Mobile IP with route optimization, in that it tells the caller where to send the invitation.

Although the load on a redirect server can be expected to be lower, we will discuss only proxy server from now on. The reason is that the message exchange delay is shorter in the case of SIP proxy server. Furthermore, the SIP proxy server can handle the firewall and the network address translation (NAT) problem. Figure 4 shows the SIP architecture. Here, the visited registrar (VR) is assumed to be a combination of the outgoing SIP proxy server, the location server, and the user agent server. Likewise, the home registrar (HR) is assumed to be a combination of the incoming SIP proxy server, the location server, and the user agent server. The MN will be a user agent client.

As stated in the previous section, SIP supports personal mobility; that is, a user can be found independent of location and network device (PC, laptop, IP phone, etc.). Originally



■ Figure 5. *SIP registration.*

the SIP was designed only for roaming. However, more recently, there have been efforts on how to maintain online connectivity during the SIP session in spite of handoff. The most promising approach is to reinvite the correspondent host by sending an INVITE message.

In SIP, faster handoff can be achieved by using an RTP translator [12]. With the RTP translator, the proxy server can rewrite the media destination in the outgoing INVITE message as the proxy server or the affiliated RTP translator, so the MN hands off in the same domain (more precisely, under the same RTP translator) without reestablishing the channel with the correspondent host. This mechanism is similar to the micromobility solution in the previous section. We assume that the outgoing proxy server provides this functionality.

SIP Message Flow

We assume that the MN and foreign network use Dynamic Host Configuration Protocol (DHCP) or one of its variants to configure its subnetwork. The MN broadcasts DHCP_DISCOVER message to the DHCP servers. Several servers may offer a new address to the MN via DHCP_OFFER that contains IP address, address of default gateway, subnet mask, and so on. (There is a proposal that DHCP_OFFER can also include SIP information [13], which is assumed in this article.) The MN then selects one DHCP server (and an IP address) and sends DHCP_REQUEST to the selected server. The DHCP server sends DHCP_ACK to confirm the assignment of the address to the MN.

After the MN is assigned an IP address from the DHCP server, the MN will initiate the signaling flow for SIP complete registration in a visited network, as depicted in Fig. 5 [14]. (DHCP message exchange is not shown here.) First, the MN sends a SIP *REGISTER* message with its new (temporary) IP and MN's profile to the VR. Note that the MN has obtained the address of the local SIP proxy server from DHCP messages upon its configuration (or reconfiguration) in the visited network. The VR queries the AAA entity of the visited network to verify the MN's credentials and rights by sending Diameter-compliant message (QUERY in Fig. 5). The AAA entity (AAAF) of the visited network sends a request (Diameter-compliant message) to the AAA entity (AAAH) of the home network to verify the MN's credentials and rights. The AAAH queries the HR and gets a reply from the HR, and then sends the appropriate answer to the AAAF. The AAAF sends an appropriate response to the VR. The VR sends either an SIP 200 OK response to the MN upon success, or a 401 unauthorized response upon failure of the registration. Note that the messages to/from AAA servers are Diameter-compliant.

After this registration, the MN can initiate the SIP session by sending the INVITE message to the callee. (Suppose the MN is the caller and a correspondent node, CN, is the callee.) Then the callee responds with a SIP OK message. (These messages are not shown in Fig. 5.) Here, we assume that the CN is located in its home network. For the detailed description of the signaling messages in SIP; please refer to [15].

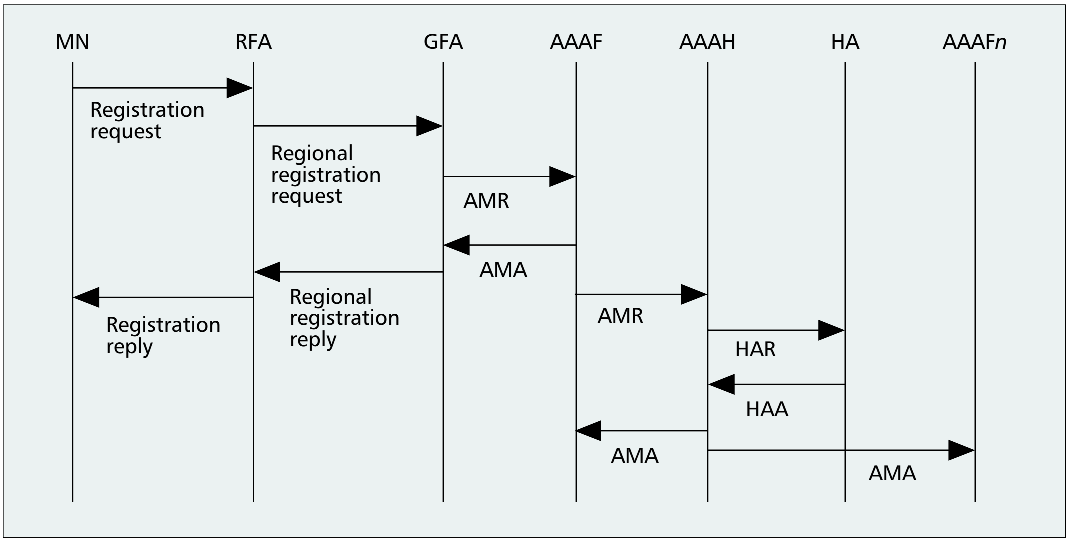
In the case of micromobility, there is no need to verify the user's credentials via the AAA server. The MN (SIP client) sends a SIP REGISTER message with the new MN's address. Then the VR verifies the user's credentials and registers the user of the MN in its contact database, and updates its contact list, which is called *expedited registration*. And then the VR replies with a SIP OK message. In the case of macromobility, the signaling message flow is the same as the SIP registration (Fig. 5).

Shadow Registration

In the previous two sections, we have illustrated how signaling messages are exchanged between entities in the Mobile IP and SIP approaches. In both the approaches, the signaling for the interdomain handoff takes much longer time and larger traffic than the intradomain handoff, which is likely to result in noticeable disruption in VoIP sessions.

In this section we introduce a *Shadow Registration* concept that can be applied to both approaches in order to reduce disruption time in the interdomain handoff (macromobility). The key idea is that the security association (SA) between the MN and the AAA server in neighboring domains is established a priori before the actual handoff occurs. Thus, when an MN hands off to a neighboring domain the registration request is processed locally within that domain without going all the way to the MN's AAAH.

This preestablishment of the SA can be performed in two fashions. The first one is a distributed fashion where the given AAA server directly contacts the neighboring AAA servers. The other approach is that the given AAA server informs the AAAH of the MN of the neighboring AAA server and let the AAAH contact them. We believe that the neighboring AAA servers are not necessarily cooperative among each other. On the contrary, the AAAH of the MN is expected to be able to accommodate the MN's SA establishment in the neighboring domains if Internet roaming is supported by the home network. Furthermore, the AAAF of the given domain is unlikely to know which neighboring domains are available to the MN. (That is, the given AAAF server know only the information such as domain name of the neighboring domains; it cannot know which



■ Figure 6. *Mobile IP interdomain handoff with Shadow Registration.*

neighboring AAA servers should provide the Internet connectivity to the MN.) Therefore, we assume that the AAAH will send messages for *Shadow Registration* only to the relevant neighboring AAA servers.

Mobile IP Case

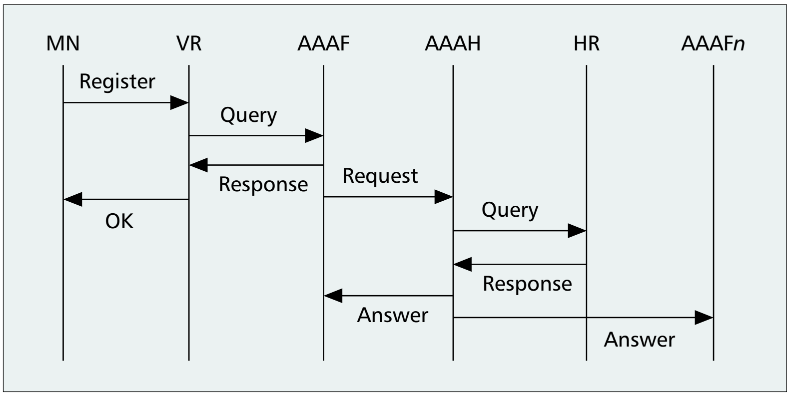
As mentioned above, when an MN triggers its registration at a given foreign network (administrative domain), the AAAF of the given network will send the AMR message to the AAAH (Fig. 2). At this time, we propose that the AAAF appends the information about all of its neighboring AAA servers (or neighboring administrative domains) to the AMR message. When the AAAH receives this message, it keeps this information. When the HA replies to the AAAH with positive certification of the MN, the AAAH checks out which neighboring AAA servers are available to the MN and sends the AMA message to those AAA servers for Shadow Registration.

The signaling message flow when an MN registers in the presence of Shadow Registration is as follows. All the messages in Fig. 2 are included in the same order; however, the contents of the message may be different. For example, the AMR message contains information about the neighboring AAA servers (AAA servers in the neighboring foreign networks of the given foreign network). The only message that is to be added is the AMA message for Shadow Registration from the AAAH to the relevant AAAFn (where the MN can connect to the Internet). There can be as many AMA messages as the number of relevant neighboring AAA servers.

Figure 6 shows the signaling message flows for the interdomain handoff in the presence of the Shadow Registration. Note that the AAAF responds to the MN's registration message without contacting the AAAH server. However, there is still message exchange for Shadow Registration since the neighboring AAA servers of the new AAAF are changed.

SIP Case

In SIP, the basic signaling mechanism for Shadow Registration is almost the same as for Mobile IP. The signaling message flow for SIP registration with Shadow Registration is almost the same as Fig. 5. However, one more message should be added: the ANSWER message from AAAH to AAAFn. The signaling flow for SIP call establishment (e.g., INVITE, OK) is not shown. A possible signaling flow in SIP for interdomain handoff with Shadow Registration is shown in Fig. 7. Note that the last ANSWER message from the AAAH to AAAFn is sent for Shadow Registration.

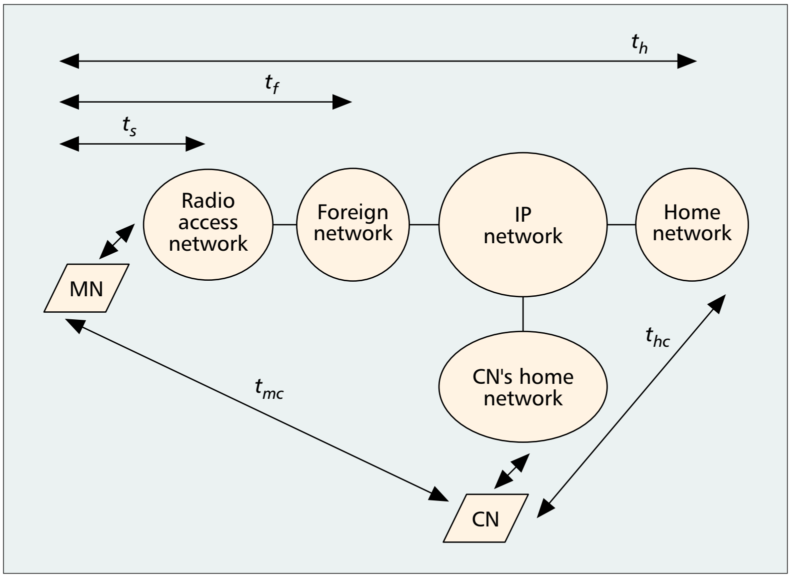


■ Figure 7. *SIP interdomain handoff with Shadow Registration.*

Delay/Disruption Analysis

In this section we make an analytic comparison between Mobile IP and SIP in terms of delay at initial registration, and disruption in intradomain and interdomain handoff, respectively. Handoff delay broadly consists of two components: link layer establishment delay and signaling delay. Link layer establishment is assumed to be negligible compared to signaling delay, so we focus on signaling delay. In addition, we disregard the quality of service (QoS) issue in signaling.

For simplicity, we assume the delay between the MN and the RFA (or DHCP server) is t_s, which is the time to send a message over the



■ Figure 8. *A simple model for analysis.*

subnet via wireless link. Also, the delay between the MN and the AAAF server (or VR) is assumed to be t_f, which is the time to send a message over the foreign network. The delay between the MN and the entities in its home network (HR, AAAH, or HA) is assumed to be t_h, which is the time to send a message to the home network. We can assume $t_s < t_f < t_h$ in general. Also, the delay between the MN and the CN is t_{mc}, and the delay between the MN's home network and the CN is t_{hc}. We only consider the scenario where the CN is in its home network in this article. The overall analytic model is depicted in Fig. 8.

Also, to make use of Mobile IP's mobility management, we consider a simple VoIP application (SVA) that is unaware of mobility. In other words, the SVA operates on top of Mobile IP. We assume the SVA has similar signaling messages as in SIP. We also assume that the home address of the callee (CN) is cached in the caller's (MN's) SVA. (That is why the SVA is mobility-unaware.) In the following, we derive some analytical results and compare the SVA while using Mobile IP and SIP as mobility protocols.

Initial Registration and Session Setup

Here we consider a scenario where an Internet connection is initiated when an MN triggers the VoIP session. That is, there is no Internet connectivity when we start a VoIP application (either the SVA or the SIP application). Thus, the initial delay will be the sum of the registration delay for Internet connectivity and VoIP signaling delay (only the round-trip time for the initial message exchange).

In the Mobile IP approach (Fig. 2), we assume that the MN will send the *Router Solicitation* message immediately when the user initiates the Internet connection. Thus, the *Router Solicitation* and *Router Advertisement* messages will take a round-trip time of ($2t_s$) in the subnet. Also, the round-trip registration message to the home network will take $2t_h$ time. After Mobile IP's registration, the SVA will initiate a VoIP session by sending the INVITE message with the CN's home address, and the CN will reply with the 200 OK (or 100 Trying) message. This will take $2t_{mc}$. To sum up, the total time to initiate a VoIP session with Mobile IP, T_{mip_init}, is given by

$$T_{mip\_init} = 2t_s + 2t_h + 2t_{mc}. \quad (1)$$

In the SIP approach (Fig. 5), there will be two round-trip delays for DHCP message interactions, which takes $4t_s$. During the DHCP message exchange, the client performs an address resolution protocol (ARP) to detect the duplicate address in the subnet, the time of which is denoted t_{arp}. Then a SIP REGISTER message will round-trip the MN's home network, which takes $2t_h$ time. Here we assume that the MN can initiate the signaling for SIP call establishment only after SIP registration. That is, during the process of SIP registration, the foreign network confirms the MN's certification and provides the Internet connectivity for SIP signaling to start a VoIP session. SIP call establishment will take $2t_{mc}$ time. Therefore, the total time to initiate the SIP session, T_{sip_init}, is given by

$$T_{sip\_init} = 4t_s + t_{arp} + 2t_h + 2t_{mc}. \quad (2)$$

Intradomain Handoff

In Mobile IP, the MN first detects a new base station or an access point (and the new IP subnet), then sends the *Router Solicitation* message to the RFA, which then replies with the *Router Advertisement* message. This will take $2t_s$ time. In intradomain handoff, the whole registration takes $2t_f$ time since intradomain handoff does not involve AAA resolution via the MN's home network. The total disruption time for intradomain handoff in Mobile IP, T_{mip_intra}, is given by

$$T_{mip\_intra} = 2t_s + 2t_f. \quad (3)$$

In SIP, the MN first detects a new wireless IP subnet and will initiate DHCP interactions as detailed in the previous section. This will take $4t_s$. Also, the ARP operation will take t_{arp}. After that, the MN will resend the REGISTER message to the VR; then the VR will reply with an OK message, which will take $2t_f$ time. The total disruption time for intradomain handoff in SIP, T_{sip_intra} is given by Eq. 4. In this case, the MN need not reinvite the CN since the VR will handle intradomain mobility.

$$T_{sip\_intra} = 4t_s + t_{arp} + 2t_f. \quad (4)$$

Interdomain Handoff

In Mobile IP, interdomain handoff will be handled as follows. First, the MN will detect the new wireless IP subnet of the different domain. The MN selects the new wireless network and then initiates handoff. First of all, the MN and NFA will exchange *Router Solicitation* and *Router Advertisement* messages, which will take $2t_s$ time. Then the MN will send a Mobile IP registration message, which will round-trip to its HA ($2t_h$). While the NFA catches this message ($2t_h - t_s$) (almost parallel) two signaling flows occur:

- Smooth handoff
- Route optimization

Let us first discuss the signaling for smooth handoff. The NFA catches the registration reply message from the HA ($2t_h - t_s$), then sends the *Binding Update* message to the OFA. Let t_{no} denote the time to send a message between the NFA and OFA. When the OFA receives the *Binding Update* message (t_{no}), it will start forwarding the packet for the MN to the NFA, which will take $t_s + t_{no}$. To sum up, smooth handoff will take total $2t_s + (2t_h - t_s) + t_{no} + (t_s + t_{no})$ since the MN starts interdomain handoff.

In the above procedure, when the OFA receives the *Binding Update* message, it updates its binding cache for the MN with the new CoA and sends the *Binding Warning* message to the HA of the MN ($t_h - t_s$). Then the HA sends the *Binding Update* message to the CN (t_{hc}). Finally, the CN will send the packets for the MN to the NFA, and then the NFA will forward the packets to the MN (t_{mc}). This route optimization will take total $2t_s + (2t_h - t_s) + t_{no} + (t_h - t_s) + t_{hc} + t_{mc}$ from when the handoff is triggered.

Considering the above two signaling flows, we can notice two points:

- The instant the packets forwarded from the OFA arrive at the MN
- The instant the packets from the CN directly arrive at the MN

There will be a blackout period until the first instant ($2t_s + 2t_h + 2t_{no}$). After that, the VoIP session resumes with possibly some disruption until the second instant. Here we take a conservative standpoint and consider the second instant as the end of disruption:

$$T_{mip\_inter} = t_{no} + 3t_h + t_{hc} + t_{mc}. \quad (5)$$

In interdomain handoff in SIP, after DHCP and ARP resolution, the MN will send SIP REGISTER message to its HR ($2t_h$), thereby enabling Internet connectivity. Then the MN reinvites the CN by sending an INVITE message, which will take $2t_{mc}$ time. The total disruption time in SIP interdomain handoff, T_{sip_inter}, is given by

$$T_{sip\_inter} = 4t_s + \text{tarp} + 2t_h + 2t_{mc}. \quad (6)$$

Numerical Results

In this section, we plot some results based on the above analysis. In the first two plottings (Figs. 9 and 10), we assume t_s = 10 ms, considering relatively low bandwidth in the wireless link. On the other hand, the delay in the wired foreign network is relatively short due to high bandwidth; thus, t_f is assumed to be t_s + 2 ms [16]. We also assume that the CN is connected to the Internet via a wireless link as well. Moreover, t_{no} is assumed to be 5 ms since the message is delivered over the wired network. Furthermore, we assume that processing time in each entity is negligible since it normally takes less than 1 ms [7]. In SIP, ARP resolution (t_{arp}) needs time in current implementations, which can be up to 1~3 s. We disregard this t_{arp} since we believe that as DHCP evolves with proliferation of mobile/wireless networks, t_{arp} will become negligible. (For example, there is no t_{arp} in DRCP [17], which can be thought of as a more evolved variant of DHCP.)

We take into consideration three configurations. In the first one, the MN is located in its home network and connected via a wireless link, while the CN's distance from the MN varies. In the second one, the MN and CN are close to each other, while the distance between the MN and its home network varies. In the last configuration, we plot the results while we vary the wireless link delay.

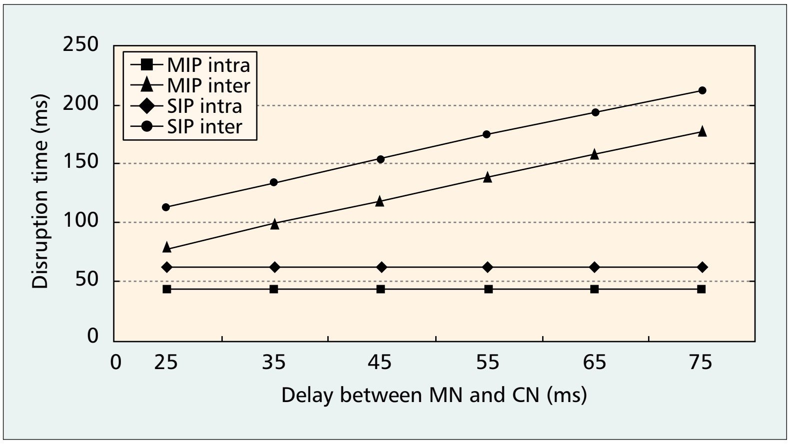


Figure 9. *Disruption time vs. delay between MN and CN.*

Figure 9 shows the disruption time as the delay from the MN to the CN, t_{mc}, increases. Since the MN and CN are connected to the network via wireless links, t_{mc} has fairly large values. Here we assume that the MN is located in its home network (t_h = 12 ms). Obviously, $t_h + t_{hc} = t_{mc}$ in this case. We plot the disruption time of intradomain and interdomain handoff in Mobile IP (denoted MIP in the legend) and SIP approaches. Overall, Mobile IP outperforms SIP. Recall that in the case of MIP interdomain handoff, the MN may start receiving VoIP data after $2t_s + 2t_h + 2\,t_{no}$, which is 54 ms in these experiments. That is, the SVA in the MN can play back the VoIP data during some portion of the interval between 54 ms and T_{mip_inter} due to smooth handoff. We believe that this forwarding of data between FAs can make the VoIP performance of Mobile IP superior to that of SIP in actual situations. Note that in SIP, the VoIP session is totally blacked out during the interval T_{sip_inter}.

Figure 10 shows the handoff disruption time as the delay from the MN to the MN's home network, t_h, increases. Here the MN and CN are assumed to be close: t_{mc} = 25 ms. (Since the wireless link delay is 10 ms, and both the MN and CN are connected via wireless links, we believe 25 ms is sufficiently small with this configuration.) Obviously, the disruption during interdomain handoff in SIP becomes shorter than that in Mobile IP as the distance between the MN and its home network increases, since SIP interdomain handoff mainly depends on t_{mc}.

The last experiments show the impact of the low-bit-rate wireless link on handoff disruption time. Figure 11 shows the disruption time as the message transmission delay over the wireless link increases. Note that this delay also applies to the wireless link to the CN. The basic configuration is the same as that of the first experiment (Fig. 9); that is, the MN is in its home network. Also, we assume that the MN and CN are at a moderate distance ($t_{mc} = 2t_s$ + 10 ms). As the wireless link delay increases, the overall signaling delay to handle handoff considerably increases. Espe-

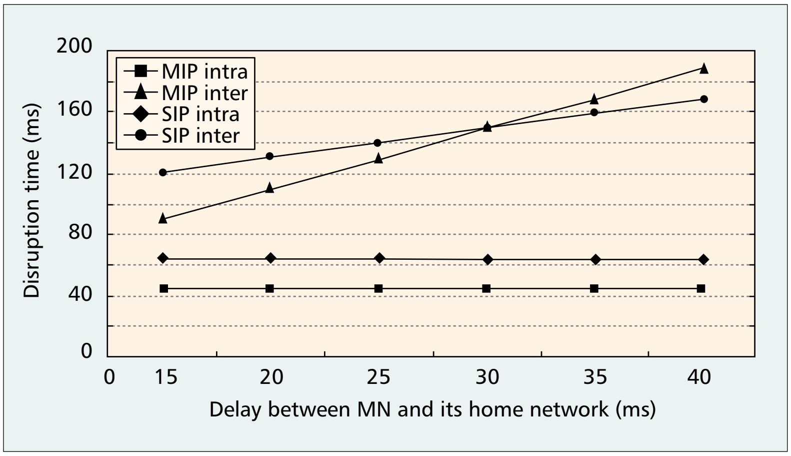


Figure 10. *Disruption time vs. delay between the MN and its home network.*

cially, the disruption time in SIP interdomain handoff increases to a large degree.

Disruption with Shadow Registration

With Shadow Registration, time to process interdomain handoff can be notably reduced since the AAA resolution for the MN can be performed in the local AAAF server. In the Mobile IP approach, the Router Solicitation/Advertisement message exchange takes $2t_s$ time. Then the MN's registration message will be handled in the current foreign network; therefore, the NFA will receive the registration reply message ($2t_f - t_s$). Then the same route optimization signaling flow will be done ($t_{no} + (t_h - t_s) + t_{hc} + t_{mc}$). Therefore, total disruption is given by

$$T_{mip\_inter\_shadow} = 2t_f + t_{no} + t_h + t_{hc} + t_{mc}. \quad (7)$$

Likewise, in SIP, the REGISTER message is handled in the local foreign network. Therefore, DHCP and ARP will take $4t_s + t_{arp}$. The REGISTER message is processed in the local AAAF and VR ($2t_f$). Then the MN reinvites the CN by sending a SIP INVITE message ($2t_{mc}$).

$$T_{sip\_inter\_shadow} = 4t_s + t_{arp} + 2t_f + 2t_{mc}. \quad (8)$$

Compared to the interdomain handoff analysis without Shadow Registration, we find that $2t_h$ is replaced with $2t_f$. Thus, Shadow Registration is useful when the MN (or user) is far from its home network.

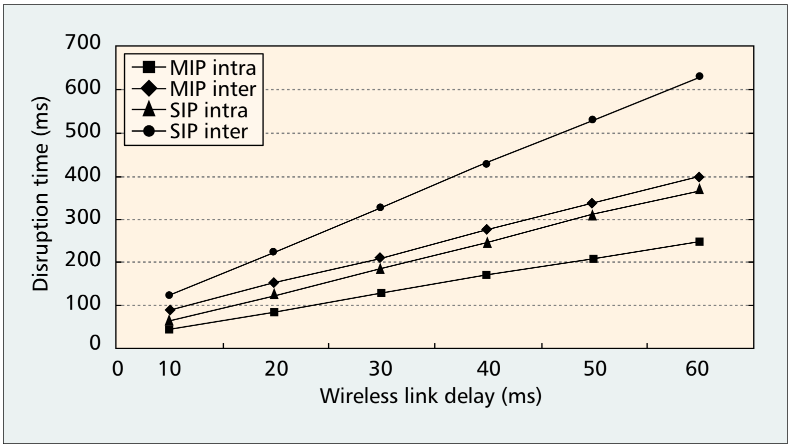


Figure 11. *Disruption time vs. wireless link delay.*

Conclusion

As wireless/mobile communications technologies become widespread, providing Internet access to mobile nodes (e.g., laptop, PDA) is of crucial importance. Also, the recent advent of VoIP services and their fast growth is likely to play a key role in successful deployment of IP-based convergence of mobile/wireless networks. In this article we focus on mobility management issues regarding VoIP services in wireless access technologies convergence. We first briefly describe Mobile IP (network layer solution) and SIP (application layer solution), and compare these two approaches in terms of mobility management. We also propose the Shadow Registration concept to reduce disruption time in interdomain handoff for VoIP sessions in mobile environments. Considering AAA functionality, we illustrated the signaling message flows of the two approaches in the presence/absence of Shadow Registration. Finally, we analyze and compare the initial delay and handoff disruption time. The disruption for handoff of the Mobile IP approach is smaller than that of the SIP approach in most situations; however, SIP shows shorter disruption when the MN and CN are close. Even though the smooth handoff scheme is not taken into consideration in the disruption analysis, we argue that smooth handoff will play an important role in reducing disruption in interdomain handoff in the Mobile IP approach.

References

[1] 3GPP, "Combined GSM and Mobile IP Mobility Handling in UMTS IP CN," TR 23.923, May 2000.
[2] C. Perkins, "IP Mobility Support," IETF RFC 2002, 1996.
[3] E. Gustafsson, A. Jonsson, and C. Perkins, "Mobile IP Regional Registration," Internet draft, draft-ietf-mobileip-reg-tunnel-05.txt, Sept. 2001, work in progress.
[4] A. Campbell *et al.*, "Cellular IP," Internet draft, draft-ietf-mobileip-cellularip-00.txt, Jan. 2000, work in progress.
[5] P. Calhoun and C. Perkins, "Diameter Mobile IPv4 Application," Internet draft, draft-ietf-aaa-diameter-mobileip-08.txt, Nov. 2001, work in progress.
[6] C. Perkins, "Route Optimization in Mobile IP," Internet Draft, draft-ietf-mobileip-optim-11.txt, Sept. 2001, work in progress.
[7] C. Perkins and Kuang-Yeh Wang, "Optimized Smooth Handoffs in Mobile IP," *Int'l. Symp. Comp. Commun.*, 1999, pp. 340–46.
[8] C. Perkins, "Mobile IP Joins Forces with AAA," *IEEE Pers. Commun.*, Aug. 2000, pp. 59–61.
[9] A. Hess and G. Shafer, "Performance Evaluation of AAA/Mobile IP authentication," Technical Report TKN-01-012, http://www-tkn.ee.tu-berlin.de/publications/papers/tkn01_012.pdf, Tech. Univ. Berlin, Aug. 2001.
[10] J. Song *et al.*, "MIPv6 User Authentication Support through AAA," Internet draft, draft-song-mobileip-mipv6-user-authentication-00.txt, Nov. 2001, work in progress.
[11] M. Handley *et al.*, "SIP: Session Initiation Protocol," IETF RFC 2543, Mar. 1999.
[12] H. Schulzrinne *et al.*, "RTP: A Transport Protocol for Real-Time Applications," IETF RFC 1889, Jan. 1996.
[13] H. Schulzrinne, "DHCP Option for SIP Servers," Internet Draft, draft-ietf-sip-dhcp-05.txt, Nov. 2001, work in progress.
[14] A. Vakil *et al.*, "Supporting Service Mobility with SIP," Internet draft, draft-itsumo-sip-mobility-service-00.txt, work in progress, Dec. 2000.
[15] H. Schulzrinne and E. Wedlund, "Application-Layer Mobility Using SIP," *Mobile Comp. and Commun. Rev.*, vol. 4, no. 3, July 2000.
[16] E. Hernandez and A. Helal, "Examining Mobile-IP Performance in Rapidly Mobile Environments: The Case of a Commuter Train," *26th Annual IEEE Conf. Local Comp. Net.*, Nov. 2001.
[17] A. McAuley *et al.*, "Dynamic Registration and Configuration Protocol (DRCP) for Mobile Hosts," Internet draft, draft-itsumo-drcp-01.txt, July 2000, work in progress.

Biographies

TED TAEKYOUNG KWON (tedkwon@cs.ucla.edu) is currently a post-doctoral researcher in the Computer Science Department at the University of California at Los Angeles (UCLA). He received his Ph.D., M.S., and B.S. degrees in computer engineering from Seoul National University in 2001, 1995, and 1993, respectively. He was a visiting student at IBM T. J. Watson Research Center in 1998 and a visiting scholar at the University of North Texas in 1999. His recent research areas include radio resource management, wireless technology convergence, mobility management, and adaptive multimedia networking. He has published over 20 technical papers on wireless/mobile communications and networking.

MARIO GERLA [SM] received his degree in electrical engineering from Politecnico di Milano, Italy, in 1966 and his M. S. and Ph. D. degrees in computer science from UCLA in 1970 and 1973, respectively. From 1973 to 1976 he was a manager at the Network Analysis Corporation, Glen Cove, New York, where he was involved in several computer network design projects. From 1976 to 1977 he was with Tran Telecommunication, Los Angeles, California, where he participated in the development of an integrated packet and circuit network. Since 1977 he has been on the Faculty of the Computer Science Department of UCLA. His current research projects cover the following areas: design and performance evaluation of protocols and control schemes for ad hoc wireless networks; routing, congestion control, and bandwidth allocation in wide area networks; and traffic measurements and characterization.

SAJAL K. DAS [SM] received a B.S. in 1983 from Calcutta University, an M.S. in 1984 from the Indian Institute of Science at Bangalore, and a Ph.D. in 1988 from the University of Central Florida at Orlando, all in computer science. Currently he is a full professor of computer science and engineering and also the founding director of the Center for Research in Wireless Mobility and Networking (CReWMaN) at the University of Texas at Arlington (UTA). His current research interests include resource and mobility management in wireless networks, mobile computing, QoS provisioning and wireless multimedia, mobile Internet, network architectures and protocols, distributed/parallel processing, performance modeling, and simulation. He has (co-)organized numerous IEEE and ACM conferences in the areas of parallel, distributed, wireless, and mobile computing, serving as technical program chair or general chair.

SUBIR DAS [M] (subir@research.telcordia.com) received his Ph.D. degree in 1996 from Indian Institute of Technology, Kharagpur. Since 1999 he has been at Telcordia Technologies Inc. and is currently a research scientist in the Wireless IP Research Laboratory. During 1997–1999 he was a faculty member in the Electronics and Electrical Engineering Department, Indian Institute of Technology, Kharagpur. He has designed several protocols (e.g., DRCP, IDMP, PANA) and architecture for next-generation wireless networks, particularly in the areas of auto-configuration, mobility management, and security. He has developed several mobile interworking prototypes for seamless connectivity over 802.11b and 3G cellular networks.

As wireless/mobile communications technologies become widespread, providing Internet access to mobile nodes is of crucial importance. Also, the recent advent of VoIP services and its fast growth is likely to play a key role in successful deployment of IP-based convergence of mobile/wireless networks.