WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. 2005; **5**:825–845

Signaling cost and performance of SIGMA: A seamless handover scheme for data networks

Shaojian Fu[1*,†], Mohammed Atiquzzaman[1], Liran Ma[1] and Yong-Jin Lee[2]

[1] *Telecommunications and Networks Research Laboratory, School of Computer Science, University of Oklahoma, Norman, OK 73019-6151, U.S.A.*
[2] *Department of Technology Education, Korea National University of Education, San 7-1, Darakri, Chongwon-Gun, Chungbuk, 363-791, Korea*

Summary

Mobile IP has been developed to handle mobility of Internet hosts at the network layer. Mobile IP (MIP), however, suffers from a number of drawbacks such as requirement of infrastructure change, high handover latency, high packet loss rate, and conflict with network security solutions. In this paper, we describe and evaluate the performance of SIGMA, a Seamless IP diversity-based Generalized Mobility Architecture. SIGMA utilizes multihoming to achieve seamless handover of mobile hosts, and is designed to solve many of the drawbacks of MIP, including requirement for changes in infrastructure. We first evaluate the signaling cost of SIGMA and compare with that of hierarchical Mobile IPv6 (an enhancement of Mobile IP) by analytical modeling, followed by comparison of handover performance of SIGMA and Mobile IPv6 enhancements. Criteria for performance evaluation include handover latency, packet loss, throughput, and network friendliness. Our results indicate that in most cases SIGMA has a lower signaling cost than Hierarchical Mobile IPv6. Moreover, for a typical network configuration, SIGMA has a higher handover performance over Mobile IP. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: mobile handover; SIGMA; mobile IP; IP diversity; signaling cost

---

1. Introduction

Mobile IP (MIP) [1] has been designed to handle mobility of Internet hosts at the network layer to manage mobile data communication. It allows a TCP connection to remain alive when a mobile host (MH) moves from one point of attachment to another. Several drawbacks exist when using MIP in a mobile computing environment, the most important ones identified to date are high handover latency, high packet loss rate [2], and requirement for change in infrastructure. MIP is based on the concept of home agent (HA) and foreign agent (FA) (which requires modification to existing routers in Internet) for routing packets from previous point of attachment to the new one. An MH needs to complete the following four steps before it can receive forwarded data from the previous point of attachment: (i) perform layer 2 (L2) handover, (ii) discover the new care of address (CoA), (iii) registering the new CoA with the HA, and (iv)

---

*Correspondence to: Shaojian Fu, Telecommunications and Networks Research Laboratory, School of Computer Science, University of Oklahoma, Norman, OK 73019-6151, U.S.A.
†E-mail: sfu@ou.edu

Contract/grant sponsor: NASA; contract/grant number: NAG3-2922.

forwarding packets from the HA to the current CoA. During this period, the MH is unable to send or receive packets through its previous or new point of attachment [3], giving rise to a large handover latency and high packet loss rate.

1.1. Recent Research on Improving MIP

Many improvements to MIP have been proposed to reduce handover latency and packet loss. IP micromobility protocols like hierarchical IP [4], HAWAII [5], and cellular IP [6] use hierarchical foreign agents to reduce the frequency and latency of location updates by handling most of the handovers locally. Low latency Handoffs in Mobile IPv4 [2] uses preregistrations and post-registrations, which are based on utilizing link layer event triggers to reduce handover latency.

Optimized smooth handoff [7] not only uses a hierarchical FA structure, but also makes the previously visited FA buffer and forward packets to MH's new location. To facilitate packet rerouting after handover and reduce packet losses, Jung *et al.* [8] introduces a location database that maintains the time delay between the MH and the crossover node. Mobile routing table (MRT) has been introduced at HA and FA in Reference [9], and a packet forwarding scheme similar to Reference [7] is also used between FAs to reduce packet losses during handover. A reliable mobile multicast protocol (RMMP), proposed in Reference [10], uses multicast to route the data packets to adjacent subnets to ensure low packet loss rate during MH roaming.

Mobile IPv6 [11] removes the concept of FA to reduce the requirement on infrastructure support (only HA required). Route optimization is built in as an integral part of Mobile IPv6 to reduce triangular routing encountered in MIPv4 [11]. Fast Handovers for Mobile IPv6 (FMIPv6) [3], aims to reduce the handover latency by configuring a new IP address before entering the new subnet. This results in a reduction in the time required to prepare for new data transmission; packet loss rate is thus expected to decrease. Like the hierarchical IP in MIPv4, hierarchical MIPv6 mobility management (HMIPv6) [12] also introduces a hierarchy of mobile agents to reduce the registration latency and the possibility of an outdated collocated CoA (CCOA) address. FMIPv6 and HMIPv6 can be used together, as suggested in Reference [12], to improve the performance further (in this paper, we refer to this combination as FHMIPv6). Even with the above enhancements, Mobile IP still can not completely remove the latency resulted from the four handover steps mentioned earlier, still resulting in a high packet loss rate [13].

1.2. Motivation of SIGMA

As the amount of real-time traffic over wireless networks keeps growing, the deficiencies of the network layer based MIP, in terms of high latency and packet loss, becomes more obvious. The question that naturally arises is: Can we find an alternative approach to network layer based solution for mobility support? Since most of the applications in the Internet are end-to-end, a transport layer mobility solution would be a natural candidate for an alternative approach. A number of transport layer mobility protocols have been proposed in the context of TCP, for example, MSOCKS [14] and connection migration solution [15]. These protocols implement mobility as an end-to-end service without the requirement to change the network layer infrastructures; they, however, do not aim to reduce the high latency and packet loss resulted from handovers. Both these protocols follow a make-after-break approach, and they disable MH to communicate with CN until the new network path is ready and signaling process is finished. As a result, the handover latency for these schemes is in the scale of seconds.

The objective of this paper is to describe the architecture of a new scheme for supporting low latency, low packet loss mobility architecture called Seamless IP diversity-based Generalized Mobility Architecture (SIGMA), and evaluate its signaling cost and performance compared with MIPv6 enhancements. Similar in principle to a number of recent transport layer handover schemes [16–18], the basic idea of SIGMA is to exploit multihoming to keep the old path alive during the process of setting up the new path to achieve a seamless handover.

Traditionally, various diversity techniques have been used extensively in wireless communications to combat channel fadings by finding independent communication paths at physical layer. Common diversity techniques include: space (or antenna) diversity, polarization diversity, frequency diversity, time diversity, and code diversity [19,20]. Recently, increasing number of mobile nodes are equipped with multiple interfaces to take advantage of overlay networks (such as WLAN and GPRS) [21]. The development of Software Radio technology [22] also enables integration of multiple interfaces into a single network interface card. With the support of multiple IP addresses in one mobile host, a new form of diversity: IP diversity can be achieved. On the other hand, a new transport protocol proposed by IETF, called stream control transmission protocol (SCTP), has recently received much attention from the research community [23]. In the field of mobile and wireless communications, the performance of SCTP over wireless links [24], satellite networks [25,26], and mobile ad-hoc networks [27] is being studied. Multihoming is a built-in feature of SCTP, which can be very useful in supporting IP diversity in mobile computing environments. Mobility protocols should be able to utilize these new hardware/software advances to improve handover performance. Although we illustrate SIGMA using SCTP, it is important to note that SIGMA can be used with other transport layer protocols that support multihoming. It can also cooperate with normal IPv4 or IPv6 infrastructure without any support from Mobile IP.

1.3. Contributions of Current Research

The contributions of our paper can be outlined as follows:

- Propose and develop SIGMA. Here ‘Seamless’ means low latency and low packet loss.
- Evaluate and compare the signaling cost of SIGMA and HMIPv6 using analytical models.
- Compare the handover performance of SIGMA with various MIPv6 enhancements including FMIPv6, HMIPv6, and FHMIPv6, taking into accounts handover latency, throughput, and packet loss rate as the performance measures.

The authors are not aware of any previous studies comparing these MIPv6 enhancements with transport layer mobility solutions.

1.4. Paper Structure

The rest of this paper is structured as follows: First, Section 2 provides an overview of MIPv6 enhancements that currently developed by IETF. Section 3 describes the handover signaling procedures, timing diagram, and location management method of SIGMA. We develop analytical models to evaluate and compare the signaling costs of SIGMA and HMIPv6 in Section 4. We then compare the handover performance of SIGMA with MIPv6 enhancements by simulation in Section 5. Finally, concluding remarks are presented in Section 6.

2. Overview of MIPv6 Enhancements

One of the objectives of this paper is to compare the performance of SIGMA with a number of MIPv6 enhancements. We, therefore, briefly describe the protocols of the MIPv6 enhancements in this section.

2.1. Hierarchical Mobile IPv6

The objective of HMIPv6 is to reduce the frequency and delay of location updates caused by MH’s mobility. In HMIPv6, operation of the correspondent node and HA are the same as MIPv6. A new network element, called the mobility anchor point (MAP), is used to introduce hierarchy in mobility management. A MAP covers several subnets under its domain, called a region in this paper. A MAP is essentially a local HA. The introduction of MAP can limit the amount of MIPv6 signaling cost outside its region as follows:

- When an MH roams between the subnets within a region (covered by a MAP), it only sends location updates to the local MAP rather than the HA (that is typically further away and has a higher load).
- The HA is updated only when the MH moves out of the region.

HMIPv6 operates as follows. An MH entering a MAP domain receives router advertisements containing information on one or more local MAPs. The MH updates the HA with an address assigned by the MAP, called regional COA (RCoA), as its current location. The MAP intercepts all packets sent to the MH, encapsulates, and forwards them to the MH’s current address. If the MH changes its point of attachment within a MAP domain, it gets a new local CoA (LCoA) from the AR serving it; the MH only needs to register the LCoA with the MAP. MH’s mobility (change of the LCoA) is transparent to the HA, and the RCoA remains unchanged (thus no need to update HA) as long as the MH stays within a MAP’s region.

2.2. Fast Handovers for Mobile IPv6

The objective of FMIPv6 is to reduce the handover latency and packet loss experienced by an MH during handover. FMIPv6 achieves this goal by two mechanisms: (i) resolve the new CoA address to be used before the MH enters into the coverage of the new AR; (ii) setup a temporary tunnel between previous access router (PAR) and new access router (NAR) to forward packets to the new location.

The protocol operates as follows: when an MH senses a link-specific event (e.g., layer 2 'trigger,' such as higher signal strength from a new access point), it sends a router solicitation for proxy (RtSolPr) message to its PAR to resolve information about the anticipated new subnet. In response to RtSolPr, PAR sends a proxy router advertisement (PrRtAdv) message which contains the binding between adjacent APs and ARs. From the information provided in the PrRtAdv message, the MH formulates a prospective new CoA (NCoA) that will be used in the new subnet, and sends a fast binding update (FBU) message to the PAR. The purpose of FBU is to inform the PAR to bind previous CoA (PCoA) to NCoA, so that arriving packets can be tunneled to the new location of the MH. If the NAR considers the NCoA formulated by MH as acceptable, NAR will acknowledge MH by sending a fast binding acknowledgment (FBack).

To reduce the packet loss during a handover, a tunnel is established between PAR and NAR. Upon receiving FBU from the MH, PAR sends a handover initiate (HI) message to NAR, in response to which a handover acknowledge (HAck) message is sent by NAR to setup the tunnel with NCoA as the exit point. Once MH attached successfully with NAR, it sends a Fast Neighbor Advertisement (FNA) to NAR, and all packets cached at NAR are delivered to the MH.

The initiation of the fast handover procedure depends on the wireless link layer triggers which inform the mobile node of an imminent handoff between the wireless access points attached to PAR and NAR. This operation requires a cross-layer communication, and the performance of FMIPv6 greatly relies on the accuracy and timing of the link layer trigger. Packet loss occurs if the FBU is sent too late or too early with respect to the time that the MH detaches from the PAR and attaches to the NAR.

2.3. Hierarchical Mobile IPv6 with Fast Handover (FHMIPv6)

HMIPv6 and FMIPv6 can be used together to further reduce signaling overhead and packet loss. A natural way to integrate HMIPv6 and FMIPv6 is to place the MAP at an aggregation point above the NAR and PAR. In this case, the forwarding of packets between PAR and NAR would be inefficient, since these data packets will traverse the MAP–PAR link twice before arriving at the NAR. A mechanism has been proposed [12] to move the FMIPv6's HI/HAck message exchange from between PAR and NAR to MAP and NAR to establish a temporary tunnel. Also, FBU and FBack will be exchanged between MH and MAP instead of PAR. After receiving FBU and HAck, MAP will begin tunneling packets to NAR.

The combination of fast handover and HMIPv6 allows performance improvement by taking advantage of both hierarchial structure and link layer triggers. However, like FMIPv6, FHMIPv6 also relies heavily on accurate link layer information. MH's high movement speed or irregular movement pattern may reduce the performance gains of these protocols.

3. Architecture of SIGMA

In this section, we outline the SIGMA's signaling procedure involved in the mobile handover process. The whole procedure can be divided into five parts, which will be described below. The main idea of SIGMA is trying to keep the old data path alive until the new data path is ready to take over the data transfer by exploiting the IP diversity at MH, thus achieve a low latency, low loss handover between adjacent subnets.

In this paper, we illustrate SIGMA using SCTP. SCTP's multihoming allows an association between two end points to span multiple IP addresses or network interface cards. An example of SCTP multihoming is shown in Figure 1, where both endpoints A and B have two interfaces bound to an SCTP association. The two end points are connected through two types of links: satellite at the top and ATM at the

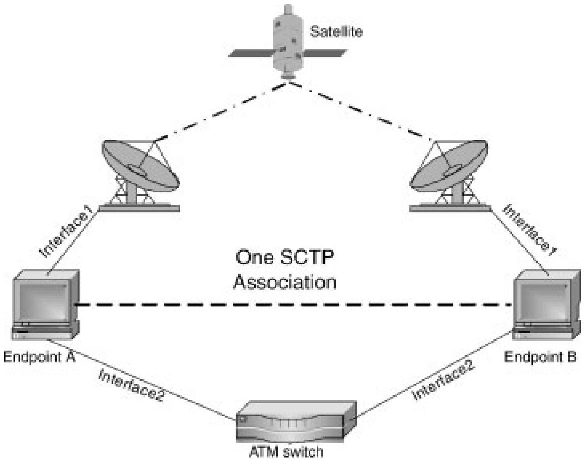


Fig. 1. Stream Control Transmission Protocol (SCTP) multihoming.

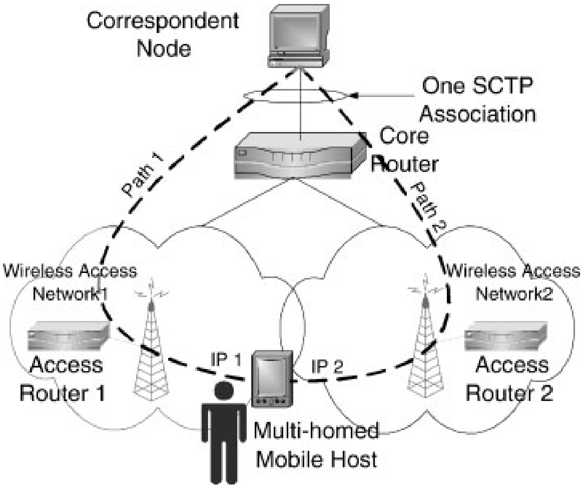


Fig. 2. An SCTP association with multihomed mobile host.

bottom. One of the addresses is designated as the primary while the other can be used as a backup in the case of failure of the primary address, or when the upper layer application explicitly requests the use of the backup.

A typical mobile handover in SIGMA using SCTP as an illustration is shown in Figure 2, where MH is a multihomed node connected to two wireless access networks. Correspondent node (CN) is a node sending traffic to MH, corresponding to the services like file download or web browsing by mobile users.

3.1. Handover Process

The handover process of SIGMA can be described by the following five steps.

Step 1: Layer 2 Handover and Obtain New IP Address: Refer to Figure 2 as an example, the handover preparation procedure begins when MH moves into the overlapping radio coverage area of two adjacent subnets. In the state of the art mobile system technologies, when a mobile host changes its point of attachment to the network, it needs to perform a layer 2 (data link layer) handover to cutoff the association with the old access point and re-associate with a new one. For example, in IEEE 802.11 WLAN infrastructure mode, this layer 2 handover will require several steps: detection, probe, and authentication and reassociation with new AP. Only after these procedures have been finished, higher layer protocols can proceed with their signaling procedure, such as layer 3 router advertisements. Once the MH finishes layer 2 handover and receives the router advertisement from the new access router (AR2), it should begin to obtain a new IP address (IP2 in Figure 2). This can be accomplished through several methods: DHCP, DHCPv6, or IPv6 stateless address auto-configuration (SAA) [28]. We call the time required for MH to acquire the new IP address as address resolution time.

Step 2: Add IP Addresses Into the Association: Initially, when the SCTP association is setup, only CN's IP address and MH's first IP address (IP1) are exchanged between CN and MH. After the MH obtained the IP address IP2 in STEP 1, MH should bind IP2 also into the association (in addition to IP1) and notify CN about the availability of the new IP address through SCTP address dynamic reconfiguration option [29]. This option defines two new chunk types (ASCONF and ASCONF-ACK) and several parameter types (Add IP Address, Delete IP address, and Set Primary Address, etc.).

Step 3: Redirect Data Packets to new IP Address: When MH moves further into the coverage area of wireless access network2, CN can redirect data traffic to new IP address IP2 to increase the possibility that data can be delivered successfully to the MH. This task can be accomplished by sending an ASCONF from MH to CN, through which CN set its primary destination address to MH's IP2. At the same time, MH needs to modify its local routing table to make sure the future outgoing packets to CN using new path through AR2.

If MH can utilize the information from layer 2, such as radio link Signal/Noise Ratio (SNR), Bit Error Rate (BER), or available bandwidth, MH has much more accurate information about when the primary data path should be switched over to the new path. One disadvantage of this method is that it requires cross-layer communication in the protocol stack, which may result in difficulties in protocol deployment.

Step 4: Update Location Manager (LM): SIGMA supports location management by employing a location manager which maintains a database recording the correspondence between MH's identity and MH's current primary IP address. MH can use any unique information as its identity, such as home address (like MIP), or domain name, or a public key defined in public key infrastructure (PKI).

Following our example, once MH decides to handover, it should update the LM's relevant entry with the new IP address, IP2. The purpose of this procedure is to ensure that after MH moves from wireless access network1 into network2, subsequent new association setup requests can be routed to MH's new IP address

(IP2). Note that this update has no impact on the existing active associations.

We can observe an important difference between SIGMA and MIP: the location management and data traffic forwarding functions are coupled together in MIP, while in SIGMA they are decoupled to speedup handover and make the deployment more flexible.

Step 5: Delete or Deactivate Obsolete IP Address: When MH moves out of the coverage of wireless access network1, no new or retransmitted data should be directed to address IP1. In SIGMA, MH notifies CN that IP1 is out of service for data transmission by sending an ASCONF chunk to CN to delete IP1 from CN's available destination IP list.

A less aggressive way to prevent CN from sending data to IP1 is to let MH advertise a zero receiver window (corresponding to IP1) to CN. This will give CN an impression that the interface (on which IP1 is bound) buffer is full and cannot receive data any more. By deactivating, instead of deleting, the IP address, SIGMA can adapt more gracefully to MH's zigzag movement patterns and reuse the previous obtained IP address (IP1) as long as the IP1's lifetime is not expired. This will reduce the latency and signaling traffic caused by obtaining a new IP address.

3.2. Timing Diagram of SIGMA

Figure 3 summarizes the signaling sequences involved in SIGMA, the numbers before the events correspond to the step numbers in Subsection 3.1. The meaning of each signaling message has been explained in Subsection 3.1, so we will not go through Figure 3 step by step. Note that here we assume IPv6 SAA is used for MH to get new IP address, the timing diagrams for using other methods can be drawn similarly. It should also be noted that until the old IP is deleted at CN (including the time for discovering new IP address), MH can always receive data packets (not shown in the figure) from old IP in parallel with the exchange of signaling packets.

3.3. Location Management

As mentioned in Step 4 of Subsection 3.1, SIGMA needs to setup a location manager for maintaining a database of the correspondence between MH's identity and its current primary IP address. Unlike MIP, the location manager in SIGMA is not restricted to the same subnet as MH's home network (in fact, SIGMA has no concept of home or foreign network). Subsections 5.3.2 and 5.4.2 will show through simulation that the location of the LM does not have impact

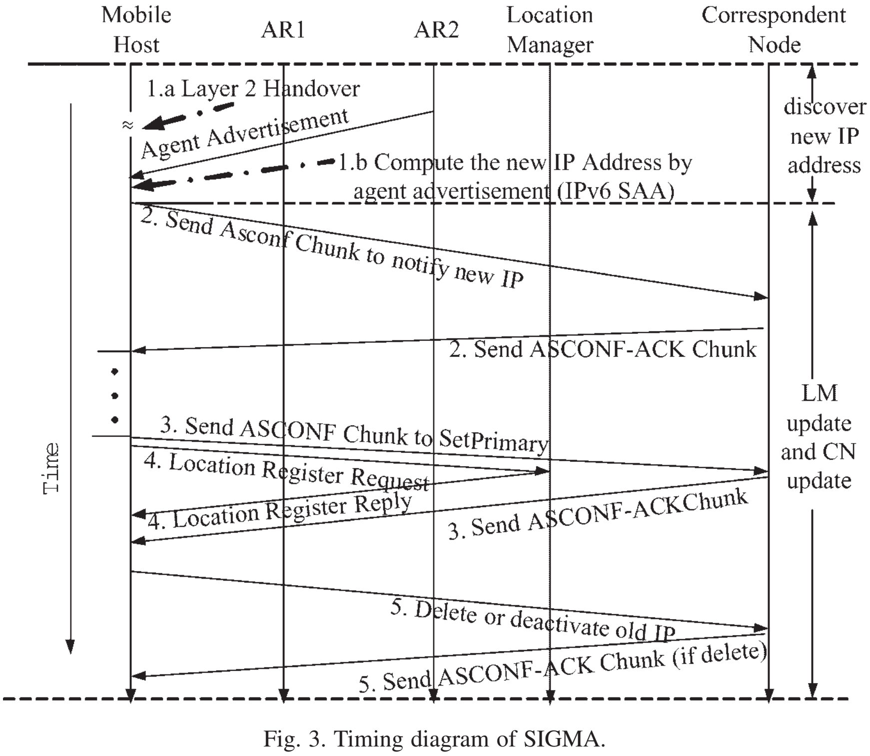


Fig. 3. Timing diagram of SIGMA.

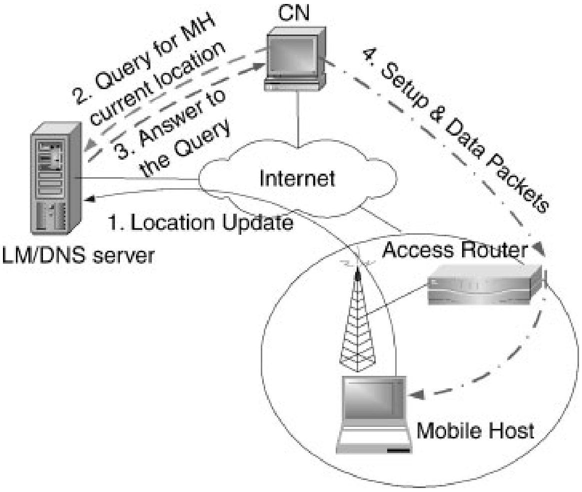


Fig. 4. Location management in SIGMA.

on the handover performance of SIGMA. This will make the deployment of SIGMA much more flexible than MIP.

The location management can be done in the following sequence as shown in Figure 4:

(1) MH updates the location manager with the current primary IP address.
(2) When CN wants to setup a new association with MH, CN sends a query to the location manager with MH's identity (home address, domain name, or public key, etc.)
(3) Location manager replies to CN with the current primary IP address of MH.
(4) CN sends an SCTP INIT chunk to MH's new primary IP address to setup the association.

If we use the domain name as MH's identity, we can merge the location manager into a DNS server. The idea of using a DNS server to locate mobile users can be traced back to [30]. In SIGMA, CN only needs to know MH's domain name, and it just needs to perform normal DNS queries to find the current IP address of MH. Note that only one round of interaction between CN and DNS server is shown in Figure 4, although this process could involve several rounds of referrals between DNS servers in practice [31]. The advantage of this approach is its transparency to existing network applications that use domain name to IP address mapping. An Internet administrative domain can allocate one or more DNS servers for its registered mobile users. Compared to MIP's requirement that each subnet must have a location management entity (HA), SIGMA can reduce system complexity and operating cost significantly by not having such a requirement.

4. Signaling Cost Analysis of SIGMA and HMIPv6

In this section, we develop an analytical model for comparing the signaling cost of SIGMA and HMIPv6. We choose HMIPv6 as the benchmark protocol for signaling cost comparison because HMIPv6 is designed to reduce the signaling cost of base MIPv6, and it has the lowest signaling cost in all versions of MIPv6 enhancements. First, the network structure being considered and the notations to be used in the model are described in Subsections 4.1 and 4.2, respectively. We analyze and develop models for the signaling cost for SIGMA and HMIPv6, in Subsections 4.3 and 4.4, respectively. The numerical results are presented in Subsection 4.5.

4.1. Network Structure

In this section, we describe the network structure that will be used in our analytical model. Figure 5 shows a

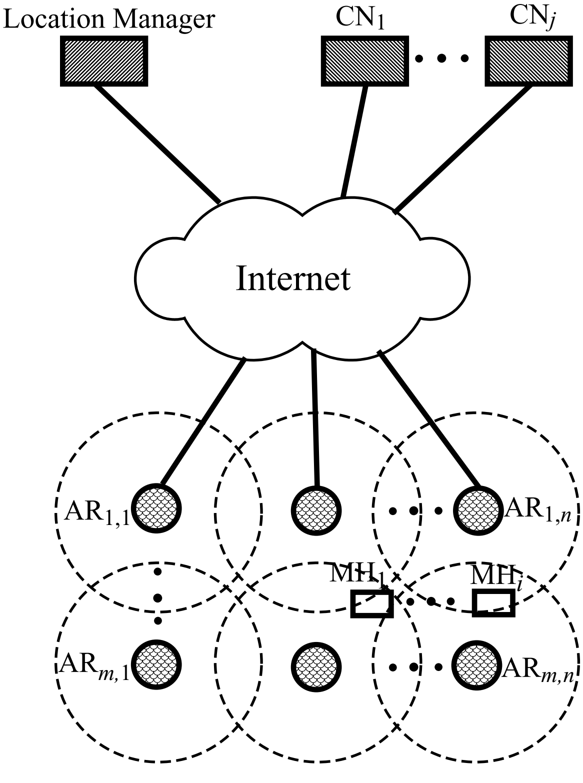


Fig. 5. Network structure considered.

two dimensional subnet arrangement for modeling MH movement, where $\mathrm{AR}_{1,1}, \cdots \mathrm{AR}_{m,n}$ represent access routers. There is one location manager (same as HA in the case of HMIPv6) and a number of CNs connected to the Internet. The MHs are roaming in the subnets covered by $\mathrm{AR}_{1,1}, \cdots \mathrm{AR}_{m,n}$, and each MH communicates with one or more of the CNs. Between a pair of MH and CN, intermittent file transfers occur caused by mobile users requesting information from CNs using protocols like HTTP. We call each active transfer period during the whole MH-CN interactivity as a session.

4.2. Notations

The notations to be used for developing the analytical models of SIGMA and HMIPv6 are given below. They are divided into three categories depending on whether they are required for SIGMA, HMIPv6, or both. For the sake of consistency, the notations for HMIPv6 modeling are similar to those used in Reference [32].

4.2.1. Notations that apply to both SIGMA and HMIPv6 signaling cost modeling

N_{mh}	total number of MHs
N_{cn}	average number of CNs with which a MH is communicating
T_{r}	MH residence time in a subnet
S	number of sessions during an MH-CN transport layer association (connection) time
λ_{sa}	average session arrival rate
λ_{pa}	average packet arrival rate
ϕ	session-mobility ratio defined as $\lambda_{\mathrm{sa}} \times T_{\mathrm{r}}$

4.2.2. Notations that apply only to SIGMA signaling cost modeling

l_{ml}	average distance between MH and location manager in hops
l_{mc}	average distance between MH and CN in hops
LU_{ml}	transmission cost of one location update from MH to location manager
γ_{l}	processing cost at location manager for each location update
υ_{l}	location database lookup cost per second for each transport layer association at LM
Ψ^{S}_{LU}	SIGMA location update cost per second for the whole system, including transmission cost and processing cost incurred by location update of all MHs
BU_{mc}	transmission cost of one binding update between MH and CN
Ψ^{S}_{BU}	SIGMA binding update cost per second between MHs and CNs for the whole system
Ψ^{S}_{PD}	SIGMA packet delivery cost per second from CNs to MHs for the whole system
Ψ^{S}_{TOT}	SIGMA total signaling cost per second for the whole system including location update cost, binding update cost and packet delivery cost, $\Psi^{S}_{\mathrm{TOT}} = \Psi^{S}_{\mathrm{LU}} + \Psi^{S}_{\mathrm{BU}} + \Psi^{S}_{\mathrm{PD}}$

4.2.3. Notations that apply only to HMIPv6 signaling cost modeling

l_{mh}	average distance between MAP and HA in hops
l_{mm}	average distance between MH and MAP in hops
LU_{mh}	transmission cost of one location update from MH to HA
LU_{mm}	transmission cost of one location update from MH to MAP
γ_{h}, γ_{m}	processing cost for each location update at HA and MAP, respectively
υ_{h}, υ_{m}	processing cost for each data packet at HA and MAP, respectively
C_{mh}	registration cost of one location update from MH to HA, including transmission cost and processing cost
C_{mm}	registration cost of one location update from MH to MAP, including transmission cost and processing cost
R	number of subnets under a MAP
M	average number of subnet crossings that will cause a HA registration in HMIPv6, i.e. MH moves out of region covered by a MAP
Ψ^{H}_{LU}	HMIPv6 location update cost per second for the whole system which includes transmission cost and processing cost incurred by location update of all MHs to their HA and/or MAP
Ψ^{H}_{PD}	HMIPv6 packet delivery cost per second for the whole system from CNs to MHs, including the encapsulation/decapsulation processing cost at mobile agents
Ψ^{H}_{TOT}	total HMIPv6 signaling cost per second for the whole system including location update cost, binding update cost and packet delivery cost, $\Psi^{H}_{\mathrm{TOT}} = \Psi^{H}_{\mathrm{LU}} + \Psi^{H}_{\mathrm{PD}}$

4.3. Signaling Cost Analysis of SIGMA

In this section, the signaling cost of SIGMA will be analyzed. Subsections 4.3.1–4.3.3 develop the cost for location update, binding update and packet delivery, respectively. Finally, Subsection 4.3.4 gives the total signaling cost of SIGMA.

4.3.1. Location update cost

In SIGMA, every subnet crossing (happens every T_r seconds) by an MH will trigger a location update, which incurs a transmission cost (LU_{ml}) and processing cost (γ) for the location update message. Since there is only one location update per subnet crossing, no matter how many CNs an MH is communicating with, the number of CNs does not have any impact on the location update cost. Therefore, the average location update cost per second in the whole system can be estimated as the number of MHs multiplied by the location update cost for each MH, divided by the average subnet residence time:

$$\Psi^S_{LU} = N_{mh}\frac{LU_{ml} + \gamma_l}{T_r} \tag{1}$$

Due to frame retransmissions and medium access contentions at the data link layer of wireless links, transmission cost of a wireless hop is higher than that of a wired hop; we denote this effect by a proportionality constant, ρ. Let the per-hop location update transmission cost be δ_U, for a round trip, LU_{ml} can be calculated as:

$$LU_{ml} = 2(l_{ml} - 1 + \rho)\delta_U \tag{2}$$

Where $(l_{ml} - 1)$ represents the number of wired hops. Therefore,

$$\Psi^S_{LU} = N_{mh}\frac{2(l_{ml} - 1 + \rho)\delta_U + \gamma_l}{T_r} \tag{3}$$

4.3.2. Binding update cost

In the analysis of binding update cost, processing costs at the endpoints (MH and CN) are not counted into the total signaling cost, since these costs stand for the load that can be scattered into user terminals and hence do not contribute to the network load. Because we are more concerned about the load on the network elements, this assumption enables us to concentrate on the impact of the handover protocol on network performance. This same assumption was also made by other previous works [32–34].

Similar to the analysis in Subsection 4.3.1, every subnet crossing will trigger a binding update to CN, which incurs a transmission cost (BU_{mc}) due to the binding update message. For each CN communicating with an MH, the MH need to send a binding update after each handover. Therefore, the average binding update cost can be estimated as:

$$\Psi^S_{BU} = N_{mh}N_{cn}\frac{BU_{mc}}{T_r} \tag{4}$$

Let the per-hop binding update transmission cost be δ_B. The BU_{mc} can be calculated as:

$$BU_{mc} = 2(l_{mc} - 1 + \rho)\delta_B \tag{5}$$

Therefore, the binding update cost per second in the whole system can be calculated by multiplying the number of MHs, the average number of communicating CNs, and the average cost per binding update:

$$\Psi^S_{BU} = N_{mh}N_{cn}\frac{2(l_{mc} - 1 + \rho)\delta_B}{T_r} \tag{6}$$

4.3.3. Packet delivery cost

Unlike the analysis of packet delivery cost in Reference [32], we do not consider the data packet transmission cost, IP routing table searching cost, and bandwidth allocation cost since these costs are incurred by standard IP switching, which are not particularly related to mobility protocols. Instead, we only consider the location database lookup cost at LM. Moreover, we take into account the processing cost caused by packet tunneling to better reflect the impact of mobility protocol on overall network load.

For SIGMA, a location database lookup at LM is required when an association is being setup between CN and MH. If each session duration time is independent from each other, the association setup event happens every S/λ_{sa} seconds. If we assume the database lookup cost has a linear relationship with N_{mh}, and φ_l and ψ be the per location database lookup cost and the linear coefficient at LM, then the per-second per-association lookup cost υ_l can be calculated as:

$$\upsilon_l = \frac{\varphi_l\lambda_{sa}}{S} = \frac{\psi N_{mh}\lambda_{sa}}{S} \tag{7}$$

Since SIGMA is free of packet encapsulation or decapsulation, there is no processing cost incurred at intermediate routers. So the packet delivery cost from

CN to MH can be calculated by only counting the location database lookup cost. This cost can be expressed as:

$$\Psi^{S}_{\mathrm{PD}} = N_{\mathrm{mh}} N_{\mathrm{cn}} \upsilon_{\mathrm{l}} = N^{2}_{\mathrm{mh}} N_{\mathrm{cn}} \frac{\psi \lambda_{\mathrm{sa}}}{S} \tag{8}$$

4.3.4. *Total signaling cost of SIGMA*

Based on above analysis on the location update cost, binding update cost, and packet delivery cost shown in Equations (3), (6), and (8), we can get the total signaling cost of SIGMA as:

$$\Psi^{S}_{\mathrm{TOT}} = \Psi^{S}_{\mathrm{LU}} + \Psi^{S}_{\mathrm{BU}} + \Psi^{S}_{\mathrm{PD}} \tag{9}$$

4.4. Signaling Cost Analysis of HMIPv6

The analysis in this section follow a logic which is similar to the previous work on HMIP signaling cost analysis [32]. However, our analysis differs from [32] in three ways: (i) we do not consider the packet delivery costs incurred by standard IP switching, since they are not particularly related to mobility protocols; (ii) the tunneling costs at HA and MAP are considered explicitly; (iii) we removed the processing costs at FAs to match the operation of HMIPv6. These modifications to the analysis of Reference [32] enables us to compare the signaling cost of SIGMA and HMIPv6 more consistently. If we consider HMIPv6 operating at the bidirectional tunneling mode [11], there is no binding update cost since the MH will not send a binding update to CN. Subsections 4.4.1 and 4.4.2 develop the cost for location update and packet delivery respectively, and Subsection 4.4.3 gives the total signaling cost of HMIPv6.

4.4.1. *Location update cost*

In HMIPv6, an MH does not need to register with the HA until the MH moves out of the region covered by a MAP, instead it only registers with the MAP. Therefore, every subnet crossing within a MAP (happens every T_{r} seconds) will trigger a registration to the MAP, which incurs a transmission cost to MAP (LU_{mm}) and processing cost at MAP (γ_{m}) of the location update message. Therefore, $C_{\mathrm{mm}} = LU_{\mathrm{mm}} + \gamma_{\mathrm{m}}$.

For every region crossing between MAPs (happens every $M \times T_{\mathrm{r}}$ seconds), MH needs to register with HA, which incurs a transmission cost to HA (LU_{mh}), processing cost at HA (γ_{h}), and processing cost at MAP ($2\gamma_{\mathrm{m}}$, since MAP needs to process both registration request and reply messages). Therefore, $C_{\mathrm{mh}} = LU_{\mathrm{mh}} + \gamma_{\mathrm{h}} + 2\gamma_{\mathrm{m}}$.

Similar to SIGMA, the number of CNs that an MH is communicating with have no impact on the location update. Therefore, the average location update cost per second in the whole system can be estimated as the number of MHs multiplied by the location update cost for each MH, then divided by the average subnet residence time:

$$\Psi^{H}_{\mathrm{LU}} = N_{\mathrm{mh}} \frac{MC_{\mathrm{mm}} + C_{\mathrm{mh}}}{MT_{\mathrm{r}}} \tag{10}$$

Similar to Equation (2), for a round trip, LU_{mh} and LU_{mm} can be calculated as:

$$LU_{\mathrm{mh}} = 2(l_{\mathrm{mm}} + l_{\mathrm{mh}} - 1 + \rho)\delta_{\mathrm{U}} \tag{11}$$

$$LU_{\mathrm{mm}} = 2(l_{\mathrm{mm}} - 1 + \rho)\delta_{\mathrm{U}} \tag{12}$$

Also, M can be calculated from the total number of subnets ($m \times n$) and the number of subnets beneath a MAP (R) [32]:

$$M = \frac{\mathrm{mn} - 1}{\mathrm{mn} - R} \tag{13}$$

Note that Equation (13) is slightly different from the result given in Reference [32], since the definition of M is different. In Reference [32], the definition of M assumes that after first movement, MH always stay within the same MAP. Compared with the definition in this paper, there will be a difference of one in Equation (13).

Therefore,

$$\Psi^{H}_{\mathrm{LU}} = N_{\mathrm{mh}} \left[\frac{2(l_{\mathrm{mm}} - 1 + \rho)\delta_{\mathrm{U}} + \gamma_{\mathrm{m}}}{T_{\mathrm{r}}} + \frac{2(l_{\mathrm{mm}} + l_{\mathrm{mh}} - 1 + \rho)\delta_{\mathrm{U}} + \gamma_{\mathrm{h}} + 2\gamma_{\mathrm{m}}}{T_{\mathrm{r}}} \times \frac{\mathrm{mn} - R}{\mathrm{mn} - 1} \right] \tag{14}$$

4.4.2. *Packet delivery cost*

Similar to the analysis of Subsection 4.3.3, for packet delivery cost analysis, we only consider the location database lookup cost and tunneling-related costs at

HA and MAP. For each packet sent from CN to MH, processing costs incurred in sequence are: one location database lookup and one encapsulation at HA; one location database lookup, one decapsulation, and one encapsulation at MAP.

Let φ_{h}, φ_{m} be the per location database lookup costs at HA, MAP, respectively; let τ be the per encapsulation/decapsulation cost at HA or MAP; and let ψ be the linear constant for location database lookup as defined in Equation (7); then we have:

$$\upsilon_{\mathrm{h}} = \varphi_{\mathrm{h}} + \tau = (\psi N_{\mathrm{mh}}) + \tau \tag{15}$$

$$\upsilon_{\mathrm{m}} = \varphi_{\mathrm{m}} + 2\tau = \left(\psi \frac{N_{\mathrm{mh}} R}{\mathrm{mn}}\right) + 2\tau \tag{16}$$

So the packet delivery cost from CN to MH can be calculated by summing up the processing cost due to database lookup and tunneling in the system, as shown in Equations (15) and (16). This cost can be expressed as:

$$\begin{aligned}\Psi_{\mathrm{PD}}^{H} &= N_{\mathrm{mh}} N_{\mathrm{cn}} \lambda_{\mathrm{pa}} (\upsilon_{\mathrm{h}} + \upsilon_{\mathrm{m}}) \\ &= N_{\mathrm{mh}} N_{\mathrm{cn}} \lambda_{\mathrm{pa}} \left(\psi N_{\mathrm{mh}} \frac{\mathrm{mn} + R}{\mathrm{mn}} + 3\tau\right)\end{aligned} \tag{17}$$

Where packet arrival rate (λ_{pa}) can be calculated from the session arrival rate and packet size. Here for HMIPv6, we need to compute packet arrival rate (λ_{pa}) instead of just using session arrival rate (λ_{sa}) as in the analysis of SIGMA. This is because in HMIPv6, we need to consider the tunneling cost for each packet sent from CN to MH. In comparison, SIGMA does not have a per-packet delivery cost component. Let F be the file size being transferred by the session, and *PMTU* be the path MTU between CN and MH, then the packet arrival rate can be calculated as:

$$\lambda_{\mathrm{pa}} = \lambda_{\mathrm{sa}} \frac{F}{PMTU} \tag{18}$$

4.4.3. *Total HMIPv6 signaling cost*

Based on above analysis of the location update cost and packet delivery cost shown in Equations (14) and (17), we can get the total signaling cost of HMIPv6 as:

$$\Psi_{\mathrm{TOT}}^{H} = \Psi_{\mathrm{LU}}^{H} + \Psi_{\mathrm{PD}}^{H} \tag{19}$$

4.5. Results and Signaling Cost Comparison of SIGMA and HMIPv6

In this Subsection, we present results showing the effect of various input parameters on SIGMA's total signaling cost. In all the numerical examples, using the following parameter values, which are obtained from previous work [32] and our calculation based on user traffic and mobility models [35,36]: $\gamma_{\mathrm{l}} = 30$, $\psi = 0.3$, $F = 10$ kbytes, $PMTU = 576$ bytes, $S = 10$, $\rho = 10$, $l_{\mathrm{ml}} = 35$, $l_{\mathrm{mc}} = 35$, $m = 10$, $n = 8$, $R = 10$, $\gamma_{\mathrm{h}} = 30$, $\gamma_{\mathrm{m}} = 20$, $\tau = 0.5$, $\lambda_{\mathrm{sa}} = 0.01$, $l_{\mathrm{mh}} = 25$, and $l_{\mathrm{mm}} = 10$.

4.5.1. *Impact of number of MHs for different subnet residence times*

The impact of number of MHs on total signaling cost of SIGMA and HMIPv6 for different subnet residence times is shown in Figure 6. Here, the values used for other parameters are: $N_{\mathrm{cn}} = 1$ and $\delta_{\mathrm{U}} = \delta_{\mathrm{B}} = 0.2$. From the figure, we can see that under different residence time, the signaling cost of both SIGMA and HMIPv6 increases with the increase of the number of MHs. When the moving speed is higher, the subnet residence time T_{r} decreases, resulting in a increase of the location update and binding update costs per second (see Equations (3), (6), and (14)). We can also observe that the total signaling cost of SIGMA is less than HMIPv6 in this scenario; this is because when δ_{U} and δ_{B} are small, the location update and binding update costs are not high, and the high

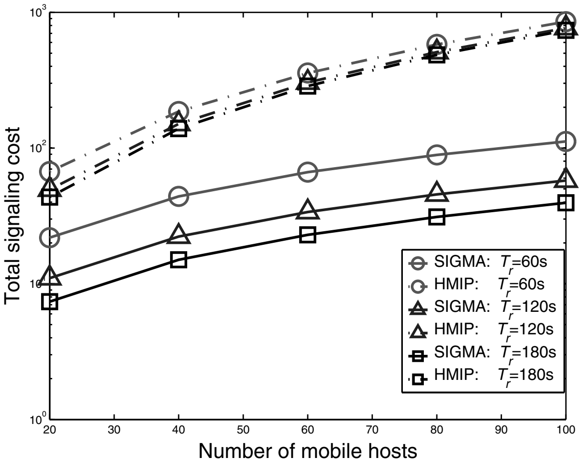


Fig. 6. Impact of number of mobile hosts (MHs) under different subnet residence times.

packet delivery cost will make the signaling cost of HMIPv6 much higher than that of SIGMA.

4.5.2. Impact of average number of communicating CN and location update transmission cost

Next, we set subnet residence time $T_r = 60$ s, and number of MHs $N_{mh} = 80$. The impact of the number of average CNs with which an MH communicates with for different per-hop transmission cost for location update cost (δ_U) is shown in Figure 7. It can be observed from this figure that when the average number of communicating CNs increases, the total signaling cost increases (see Equations (3), (6), (8), (14), and (17)). Also, when δ_U increases, the location update cost per second will increase as indicated by Equations (2), (11), and (12), which will result in the increase of the total signaling cost of both SIGMA and HMIPv6. However, we can see that the impact of δ_U is much smaller in HMIPv6; this is because HMIPv6's signaling cost is less sensitive to location update cost due to its hierarchical structure. In this scenario, signaling cost of HMIPv6 is higher than that of SIGMA when $\delta_U = 0.4$ or 1.6. However, when $\delta_U = 6.4$, SIGMA requires a higher signaling cost due to frequent location update for each subnet crossing (compared to HMIPv6's hierarchical mobility management policy).

4.5.3. Session to mobility ratio

Session to Mobility Ratio (*SMR*) is a mobile packet network's counterpart of call to mobility ratio (CMR) in PCS networks. We vary T_r from 75 to 375 s with λ_{sa} fixed to 0.01, which yields a SMR of 0.75–3.75. The impact of SMR on total signaling cost for different N_{mh} is shown in Figure 8. We can observe that a higher SMR results in lower signaling cost in both SIGMA and HMIPv6. This is mainly because high SMR means lower mobility, and thus lower signaling cost due to less location update and binding update. Also, we can see that the decrease of HMIPv6's signaling cost as a function of SMR is not as fast as that of SIGMA. This again is because HMIPv6's hierarchy structure reduces the impact of mobility on the signaling cost. The signaling cost, therefore, decreases slower than that of SIGMA when MH's mobility decreases.

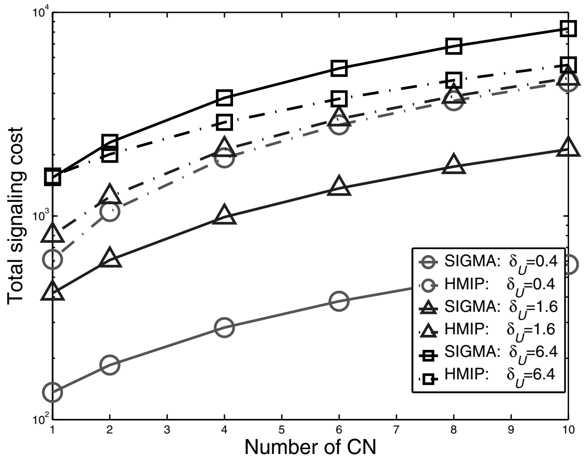


Fig. 7. Impact of number of CNs and per-hop binding update transmission cost.

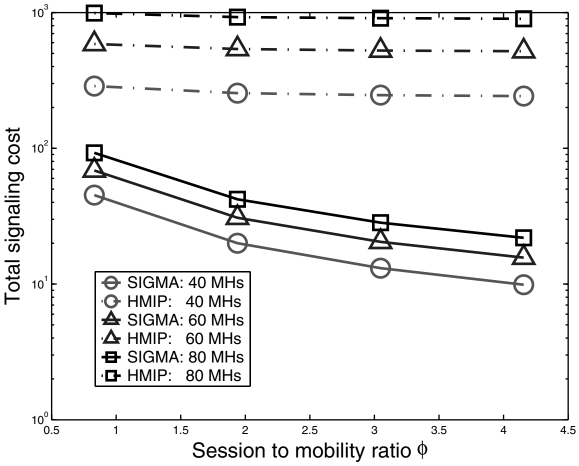


Fig. 8. Impact of SMR on total signaling cost for different N_{mh}.

4.5.4. Relative signaling cost of SIGMA to HMIPv6

Figure 9 shows the impact of (location update transmission cost)/(packet tunneling cost) ratio (δ_U/τ) on the relative signaling cost between SIGMA and HMIPv6. A higher δ_U/τ ratio means that the location update requires more cost while packet encapsulation/decapsulation costs less. This ratio depends on the implementation of the intermediate routers. We can see that as long as $\delta_U/\tau < 15$, the signaling cost of SIGMA is less than that of HMIPv6 due to the advantage of no tunneling required. After that equilibrium point, the cost of location update will take dominance, and the signaling cost of SIGMA will become higher than that of HMIPv6.

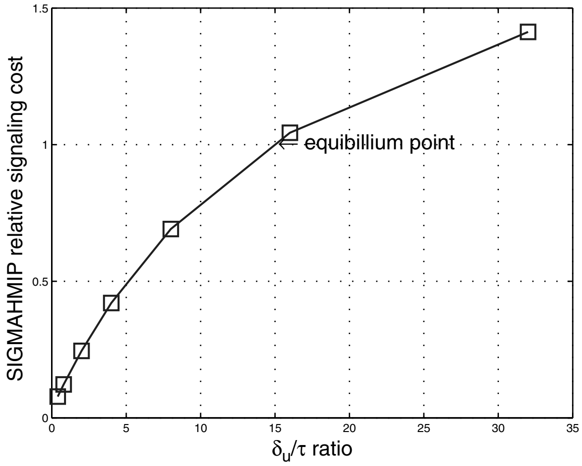


Fig. 9. Impact of δ_U/τ ratio an SIGMA to HMIPv6 relative signaling cost.

5. Performance Evaluation of Handover

In this section, we first describe the simulation topology and configurations that have been used to compare the performance of SIGMA and MIPv6 enhancements. Next, we will show two simulation packet traces of SIGMA to illustrate the seamless handovers of SIGMA, followed by comparison results between SIGMA and MIPv6 enhancements in terms of handover latency, throughput, packet loss rate, and network friendliness.

We have used *ns*-2 simulator that supports SCTP as the transport protocol, and incorporated FMIPv6, HMIPv6, FHMIPv6 implementations [37] and MIP route optimization [38]. We have also implemented SIGMA on *ns*-2.

Standard *ns*-2 simulator does not have direct support for layer 2 handover latency simulation; an MH can communicate with two APs simultaneously once the MH entering into the overlapping region of the two APs. In order to simulate mobile handovers between real-world infrastructure mode WLANs, we also implemented layer 2 handover latency in *ns*-2 IEEE 802.11 code by introducing layer 2 beacons and a set of timers.

5.1. Simulation Topology and Configurations

The network topology used in our simulations for FMIPv6, HMIPv6, FHMIPv6 and SIGMA is shown in Figure 10. This topology has been used extensively in earlier MIP performance studies [12,37]. In the simulation, MH initially has an IP address of 2.0.1 (IP1) when it is associated with AR1. After moving into the overlapping region, MH will get new IP address 3.0.1 (IP2) from AR2, which will make it have two IP (IP1 and IP2) available at the same time. Once MH moves out of the coverage of AR1, the IP1 is deleted and only IP2 is available. In Figure 10, MIPv6 enhancements use HA, while SIGMA uses it as location manager. Router2 in the topology acts as an MAP point in the case of HMIPv6 and FHMIPv6, while as a normal router in FMIPv6 and SIGMA. The link characteristics, namely the bandwidth (Megabits/s) and propagation delay (milliseconds), are shown on the links.

The following configurations are used in our simulations:

- A pair of FTP source and sink agents are attached to the CN and MH, respectively, to transfer bulk data from CN to MH. To stabilize the result, each simulation run lasted for 500 s of MH's linear back and forth movement between AR1 and AR2.
- IEEE 802.11 is used as the MAC layer, and each AR has a radio coverage area of approximately 40 m in radius. The overlapping region between two ARs is 10 m. The advertisement period of HA/AR1/AR2 is one second, although the advertisements are not synchronized.
- To allow a fair comparison between SIGMA and MIPv6 enhancements, we have used standard

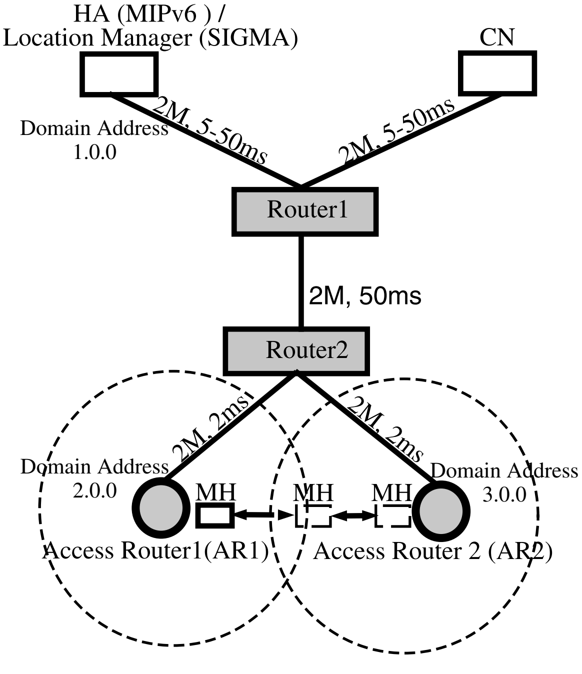


Fig. 10. Simulation topology.

SCTP (without mobility related modifications) as the transport layer protocol for the MIPv6 enhancements. This ensures that all the handover schemes considered in this paper use the same connection setup and congestion control mechanisms; the results are thus only affected by the handover schemes.

5.2. Packet Trace of SIGMA

In this section, we show packet traces and congestion window traces of SIGMA to illustrate the seamless handover of SIGMA. The traces can be classified into two categories: (1) no layer 2 handover latency, and (2) layer 2 handover latency of 200 ms. For both the categories, the IP address resolution latency is set to 500 ms.

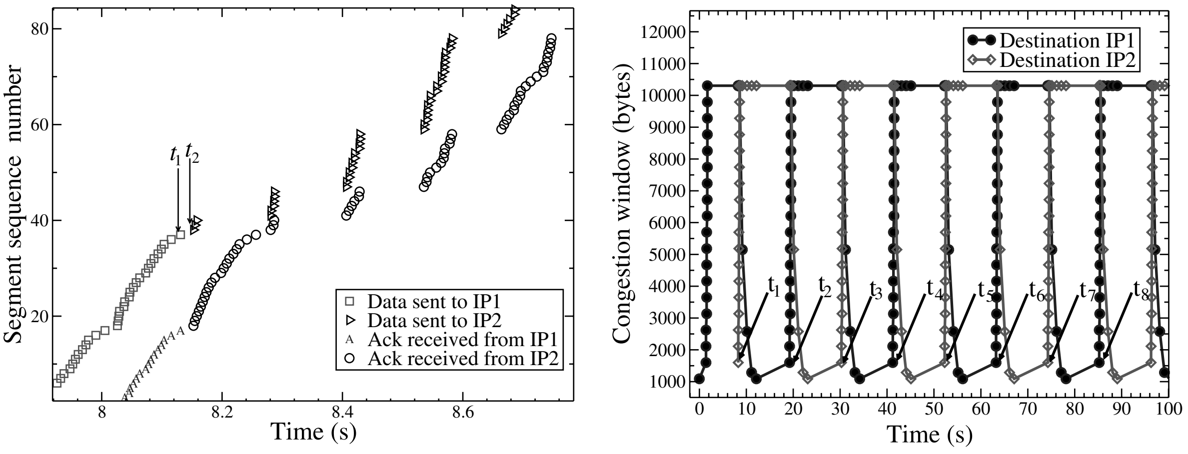
5.2.1. No layer 2 handover latency

Figure 11(a) shows packet trace observed at the CN during a typical SIGMA handover, with data sent from CN to MH. The segment sequence numbers are shown as MOD 100. For zero layer 2 handover latency, the SIGMA handover finishes immediately. From Figure 11(a), we can see that SCTP data segments are sent to MH's old IP address (IP1) until time 8.140 s (point t_1), and then to the new IP address (IP2) almost immediately (point t_2); all these packets are successfully delivered to MH. Since the change of routing table at MH occurs at the same time as the sending of SetPrimary chunk to CN (see STEP3 in Subsection 3.1), the ACKs sent to CN after time 8.134 s (time when handover decision is made) use the new path through AR2, which is not the same as the path receiving data packets before 8.140 s. Also note that, at t_2 a slow start begins at transport address IP2. The initial congestion window (*cwnd*) is three instead of two (as specified in RFC2960) because CN received an ACK from the new path resulting in the *cwnd* increasing by one segment size. At time 8.266 s, the next window of data is sent to IP2 using *cwnd* of six (according to the slow start algorithm).

Figure 11(b) shows CN's congestion window evolution when there is no layer 2 latency. The time instants labeled with odd subscripts (t_1, t_3, t_5, and t_7) represent handovers from AR1 to AR2, while the ones labeled with even subscripts (t_2, t_4, t_6, and t_8) represent handover from AR2 to AR1. The figure demonstrates that SIGMA achieve seamless handover, as evidenced by the fact that *cwnd* for the new path picks up before *cwnd* for the old path drops (due to no data being directed to the old path after the new path becomes the primary). Moreover, to probe the new network gradually after the handover, *cwnd* for the new path increases according to the slow start algorithm; this means SIGMA is network friendly by complying with the congestion control principles of the Internet.

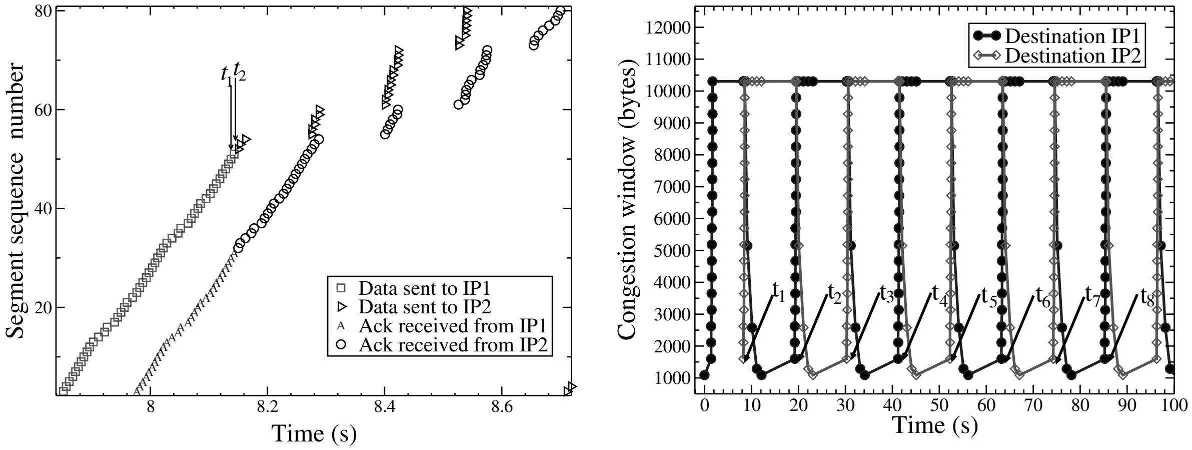
5.2.2. Layer 2 handover latency of 200 ms

To reflect a real-world scenario, we introduce a layer 2 handover latency of 200 ms. Figure 12(a) shows the corresponding packet trace. We can see that data segments are sent to MH's old IP address (IP1) until time 8.160 s (point t_1), and then almost immediately



(a) (b)

Fig. 11. SIGMA with no L2 handover latency. (a) Packet trace. (b) Congestion window.



(a) (b)

Fig. 12. SIGMA with 200 ms L2 handover latency. (a) Packet trace. (b) Congestion window.

afterwards to the new IP address (IP2) (point t_2). All packets are successfully delivered to MH without any loss. Because SIGMA prepared the new path in parallel with data forwarding over the old path, it experienced a seamless handover.

Figure 12(b) shows CN's congestion window evolution corresponding to the case of 200 ms layer 2 handover latency. This figure demonstrates that SIGMA can still achieve seamless handover in the presence of layer 2 latency. *cwnd* for the paths through IP1 and IP2 pick up and drop alternatively in a smooth manner. Similar to Figure 11(b), the time instants in Figure 12(b) labeled with odd and even subscripts represent handovers from AR1 to AR2 and AR2 to AR1, respectively.

We found that the only impact of layer 2 handover is to push the time instant of transport layer handover by 20 ms (8.140 s vs. 8.160 s comparing point t_1 of Figures 11(a) and 12(a)). This is the basic reason for SIGMA achieving a low handover latency, which eventually results in low packet loss rate and high throughput as will be shown in Subsections 5.3 and 5.4, respectively.

5.3. Handover Latency

We define handover latency as the time interval between the last data segment received through the old path and the first data segment received through the new path by the MH. In this section, we examine the impact of moving speed, link delay between HA (location manager in the case of SIGMA) and Router1, and the link delay between CN and Router1 on the overall handover latency of SIGMA and MIPv6 enhancements.

5.3.1. Impact of moving speed

First, we vary the speed of MH from 1.0 up to 15.0 m/s, while fixing link delays between HA(LM) and Router1, and CN and Router1 to 20, 5 ms, respectively. As MH moves faster, all MIPv6 enhancements and SIGMA experience a higher handover latency due to shorter time to prepare for the handover (see Figure 13(a)). However, since FMIPv6 relies on the assumption that the discovery of the NAR takes place well in advance of the actual handover, increase in speed has the most significant effect on FMIPv6. The assumption can easily break down at high moving speed. When MH moves fast, it is very difficult for FMIPv6 to complete the signaling process before MH moves to a new attachment point. Therefore, the handover latency will increase significantly.

Because HMIPv6 and SIGMA do not rely on the above assumption, the effect of moving speed on these two protocols is smaller. However, as the moving speed increases, the possibility of packets being forwarded to the old path and getting lost increases. Therefore, the time when MH can receive packets from the new path is postponed, resulting in increase in the handover latency.

5.3.2. Impact of link delay between HA(LM) and Router1

We vary the link delay between HA (LM) and Router1 (which affects the time taken by MH to update HA or LM) from 5 up to 200 ms, while keeping the MH's moving speed, and CN-Router1 link delay at 5 and 5 ms, respectively. The effect of the HA (LM)-Router

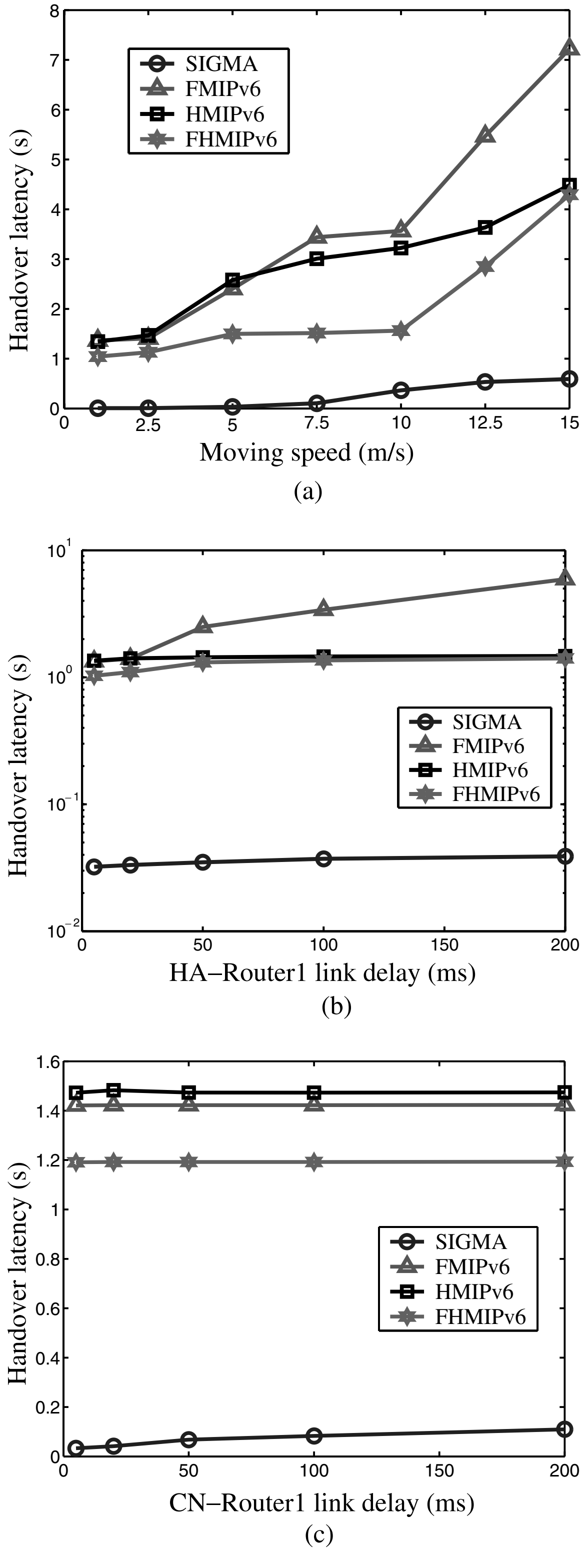


Fig. 13. Impact of different parameters on overall handover latency. (a) Impact of moving speed. (b) Impact of HA-Router1 link delay. (c) Impact of CN-Router1 delay.

link delay on the overall latency is shown in Figure 13(b). Since SIGMA decouples location management from the critical handover process (see step 4 of Figure 3), the HA (LM)-Router link delay does not have noticeable impact on the handover latency of SIGMA. This implies that the location manager of SIGMA can be placed anywhere in the Internet, without sacrificing handover performance.

For HMIPv6 and FHMIPv6, when MH moves between AR1 and AR2, it only needs to register with the MAP node (Router2). The HA-Router1 link delay, therefore, does not have much impact on these two MIPv6 enhancements. However, each location update in FMIPv6 goes through the HA-Router1 link; the overall latency, therefore, increases with an increase of the HA-Router1 link delay.

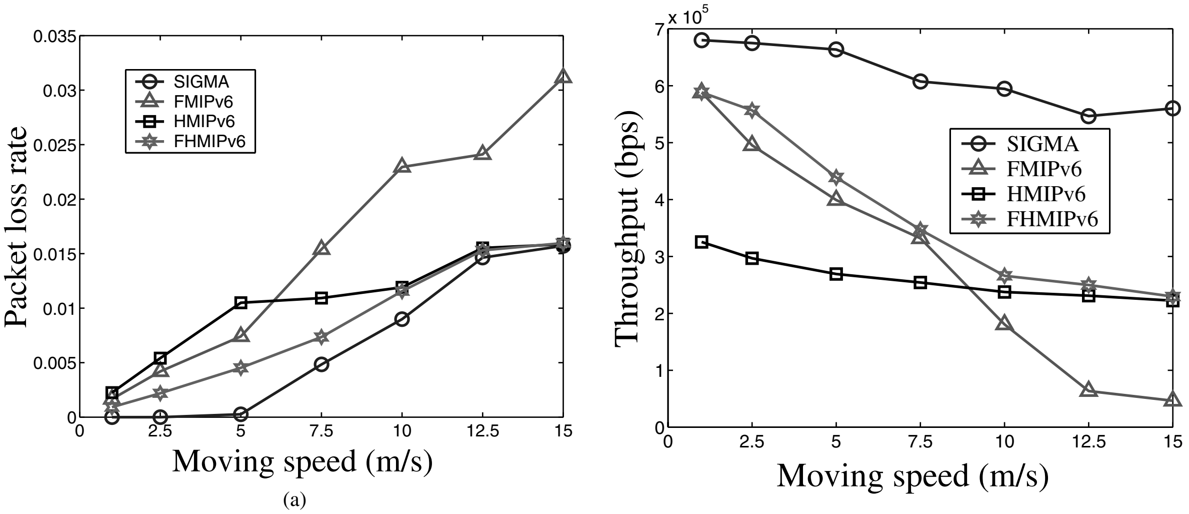
5.3.3. Impact of link delay between CN and Router1

We vary the CN-Router1 link delay from 5 up to 200 ms, while fixing MH's moving speed and HA (LM)-Router1 link delay at 5 and 20 ms, respectively. The CN-Router1 link delay decides the time taken by the MH to update the binding cache at CN (or CN's protocol control block in OS kernel, in the case of SIGMA).

The effect of the CN-Router1 link delay on the overall latency is shown in Figure 13(c). Since our definition of handover latency does not require route optimization in MIPv6 enhancements to finish, the delay required for route optimization signaling (such as return routability test [11] does not have impact on the handover latency. As long as the MH receives packets from the new path, either directly from CN or forwarded from HA, the handover is regarded as finished. Therefore, the CN-Router1 link delay does not have much impact on the handover latency of MIPv6 enhancements. In contrast, SIGMA always requires updating of CN before a packets can be received from the new path. Therefore, an increase of this link delay increases the handover latency (up to 109 ms in the case of 200 ms delay between CN-Router1).

5.4. Throughput and Packet Loss Rate

We define throughput as the total number of useful bits that can be delivered to MH's upper layer application divided by the simulation time. This provides an estimate of the average transmission speed that can be achieved by an SCTP association. Packet loss rate



(a)

(b)

Fig. 14. Impact of moving speed on packet loss rate (a) and throughput (b).

is defined as the number of packets lost due to handover divided by the total number of packets sent by the CN. In this Subsection, we will examine the impact of a number of parameters (same as those used earlier in Subsection 4.3) on the throughput and packet loss rate of SIGMA and MIPv6 enhancements.

5.4.1. Impact of moving speed

As the MH moves faster, all MIPv6 enhancements and SIGMA experience higher packet loss rates (Figure 14(a)) and reduced throughput (Figure 14(b)). This is because of the possibility of packets being forwarded to an outdated path with an increase in the speed. The packets are dropped by AR1/AR2, either because they are not aware of MH's current location or the buffer space is full. We also notice that increase in speed has the most significant effect on FMIPv6 since it relies on the assumption that detection of the new agent is well in advance of the actual handover, which may not hold when MH moves fast.

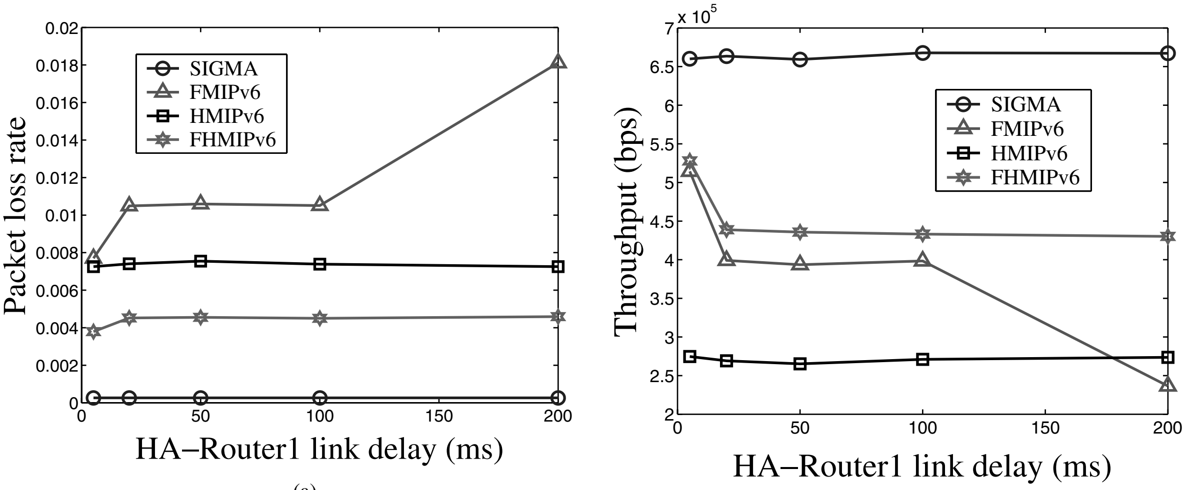
5.4.2. Impact of link delay between HA (LM) and Router1

As pointed out in Subsection 4.3.2, since SIGMA decouples location management from the critical handover process, the HA(LM)-Router1 link delay does not have impact on the packet loss and throughput of SIGMA (Figure 15(a) and (b)). For HMIPv6 and FHMIPv6, when MH moves between AR1 and AR2, it only needs to register with the MAP node (Router2). Therefore, the HA-Router1 link delay also does not have much impact on these two MIPv6 enhancements.

Each location update in FMIPv6 needs to go through the HA-Router1 link. A higher delay in this link, therefore, results in the packets forwarded by HA to have a higher possibility of being sent to an outdated location and dropped.

5.4.3. Impact of link delay between CN and Router1

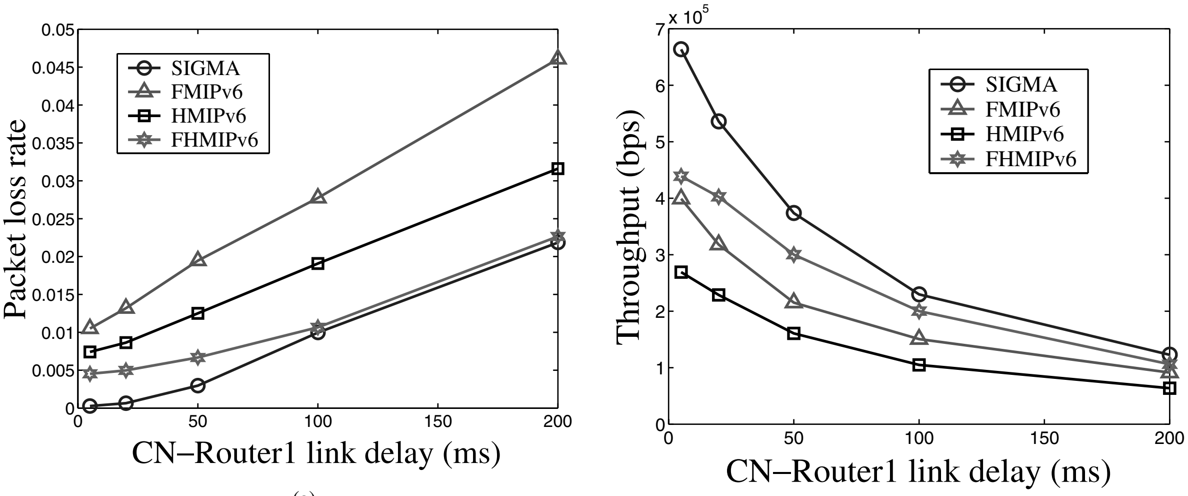
As shown in Subsection 5.3.3, CN-Router1 link delay does not have noticeable impact on the handover latency. As a result, the number of packets lost remains the same with an increase of this link delay. However, a higher value of this link delay increases the RTT. Since the throughput of an SCTP association decreases as RTT increases, the total number of packets sent to the MH will be reduced. When we compute packet loss rate by dividing the number of packets lost by the total number of packets sent by the CN, the resulting loss rate increases (Figure 16(a) and (b)). For SIGMA, as this link delay increases, it has a negative effect on both packet loss (due to non-timely CN update) and throughput (longer RTT). As a result, the packet loss rate increases relatively faster as compared to FHMIPv6 (Figure 16(a)).



(a)

(b)

Fig. 15. Impact of HA-Router1 delay on packet loss rate (a) and throughput (b).



(a)

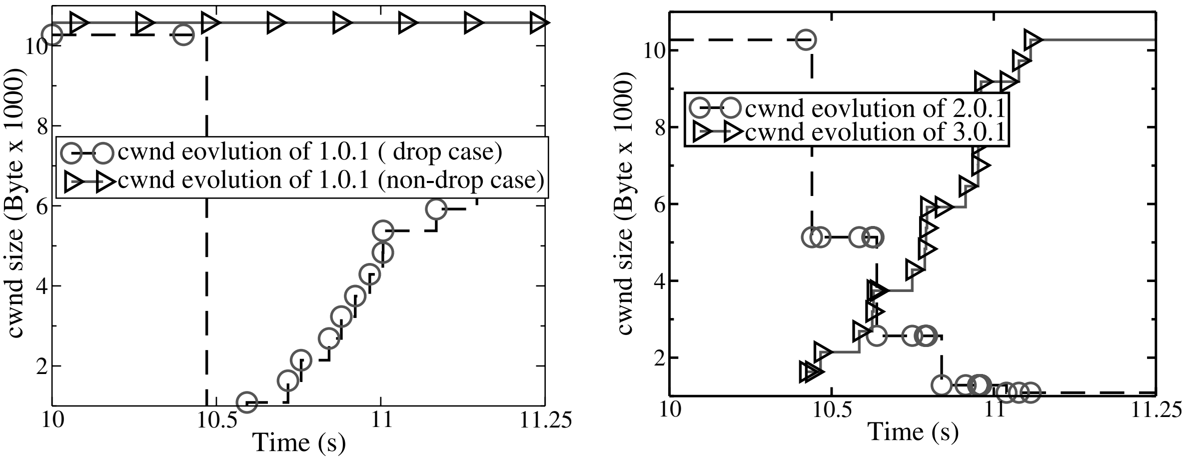
(b)

Fig. 16. Impact of CN-Router1 delay on packet loss rate (a) and throughput (b).

5.5. Network Friendliness

A network friendly mobility protocol requires that when MH enters a new domain, the CN should probe for the new domain's network condition. Unfortunately, in MIP, CN's transport protocol is not aware of the handover; it continues to use the old congestion window and slow start threshold that are only suitable for the old domain.

As shown in Figure 17(a), CN's *cwnd* for MIP remains constant after a handover takes place around time 10.5 s for the case where the handover latency is small, i.e., the CN does not encounter a timeout which could result in a drop of *cwnd*. This means that the CN implicitly assumes that the new network path has the same capacity as the old one. This apparently wrong assumption may cause network congestion in the new domain if the new path does not have enough capacity. Although this network unfriendliness can sometimes help MIP achieve better throughput, it is not preferable from the point of network performance. Note that an MIP sender may be forced to begin a slow start after a handover, due to packet losses during the handover, as shown in Figure 17(a) where the CN goes through a slow start starting at around time 10.6 s.



(a)

(b)

Fig. 17. *cwnd* evolution for MIP (a) and SIGMA (b) during handover.

In contrast to MIP, SIGMA exhibits better network friendliness. In SIGMA, the sender always probes the new network path after a handover, regardless of whether segment are dropped. As shown in Figure 17(b), the new network path is used starting at time 10.4 s, when the CN automatically begins slow start to avoid possible congestions at the new network path. This is because the CN in SIGMA switches over to a new transport address, after a handover, which automatically has a different set of congestion control parameters from the old transport address.

6. Conclusions

We have presented SIGMA, to manage handovers of mobile nodes. Using an analytical model, we have evaluated the signaling cost of SIGMA and compared with that of HMIPv6. Numerical results show that, in most practical scenarios, the signaling cost of SIGMA is lower than HMIPv6. However, there is a tradeoff between location update transmission cost (δ_{U}) and packet tunneling cost (τ); very high δ_{U}/τ ratio results in the signaling cost of SIGMA being higher than that of HMIPv6.

We also compared the handover performance of SIGMA with three different Mobile IPv6 enhancements including FMIPv6, HMIPv6, and FHMIPv6. Different performance measures, including handover latency, packet loss, and throughput, have been compared. Our results indicate that for typical network configuration and parameters, SIGMA has a lower handover latency, lower packet loss rate, and higher throughput than the three MIPv6 enhancements. SIGMA has also been shown to be network friendly due to probing of the new network at every handover.

Acknowledgment

We thank William Ivancic, Justin S. Jones, and Song Lu for the numerous discussion that greatly improved quality of this paper.

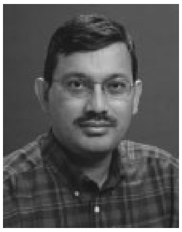
References

1. Perkins CE. Mobile networking through Mobile IP. *IEEE Internet Computing* 1998; **2**(1): 58–69.
2. Malki KE (editor). Low latency handoffs in Mobile IPv4. IETF DRAFT, draft-ietf-mobileip-lowlatency-handoffs-v4-07.txt, October 2003.
3. Koodli R (editor). Fast handovers for Mobile IPv6. IETF DRAFT, draft-ietf-mipshop-fast-mipv6-03.txt, October 2004.
4. Gustafsson E, Jonsson A, Perkins CE. Mobile IP regional registration. IETF DRAFT, draft-ietf-mobileip-reg-tunnel-04.txt, March 2001.
5. Ramjee R, Porta TL, Thuel S, *et al*. IP micro-mobility support using HAWAII. IETF DRAFT, draft-ietf-mobileip-hawaii-00.txt, June 1999.
6. Cambell AT, Kim S, Gomez J, *et al*. Cellular IP. IETF DRAFT, draft-ietf-mobileip-cellularip-00.txt, December 1999.
7. Perkins CE, Wang KY. Optimized smooth handoffs in mobile IP. In *IEEE International Symposium on Computers and Communications*, July 1999; pp. 340–346.
8. Jung MC, Park JS, Kim DM, Park HS, Lee JY. Optimized handoff management method considering micro mobility in wireless access network. In *5th IEEE International Conference on High Speed Networks and Multimedia Communications*, July 2002, pp. 182–186.
9. Wu IW, Chen WS, Liao HE, Young FF. A seamless handoff approach of Mobile IP protocol for mobile wireless data networks. *IEEE Transactions on Consumer Electronics* 2002; **48**(2): 335–344.
10. Liao W, Ke CA, Lai JR. Reliable multicast with host mobility. *IEEE Global Telecommunications Conference (GLOBECOM)*, November 2000; pp. 1692–1696.

11. Johnson D, Perkins CE, Arkko J. Mobility support in IPv6. IETF RFC 3775, June 2004.
12. Soliman H, Catelluccia C, Malki KE, *et al.* Hierarchical Mobile IPv6 mobility management (HMIPv6). IETF DRAFT, draft-ietf-mipshop-hmipv6-04.txt, December 2004.
13. Hsieh R, Seneviratne A. A comparison of mechanisms for improving Mobile IP handoff latency for end-to-end TCP. *ACM MobiCom*, San Diego, U.S.A., September 2003; pp. 29–41.
14. Maltz DA, Bhagwat P. MSOCKS: An architecture for transport layer mobility. *INFOCOM*, San Francisco, U.S.A., March 1998; pp. 1037–1045.
15. Snoeren AC, Balakrishnan H. An end-to-end approach to host mobility. *ACM MobiCom*, Boston, MA, August 2000; pp. 155–166.
16. Koh SJ, Lee MJ, Ma ML, Tuexen M. *Mobile SCTP for Transport Layer Mobility.* draft-sjkoh-sctp-mobility-03.txt, February 2004.
17. Xing W, Karl H, Wolisz A. M-SCTP: design and prototypical implementation of an end-to-end mobility concept. In *5th International Workshop on the Internet Challenge: Technology and Applications*, Berlin, Germany, October 2002.
18. Li L. PKI based end-to-end mobility using SCTP. *MobiCom 2002*, Atlanta, GA, U.S.A., September 2002.
19. Rappaport TS. *Wireless Communications Principles and Practice*. Upper Saddle River, NJ: Prentice Hall, 1996.
20. Caire G, Taricco G, Biglieri E. Bit-interleaved coded modulation. *IEEE Transactions on Information Theory* 1998; **44**(3): 927–946.
21. Holzbock M. IP based user mobility in heterogeneous wireless satellite-terrestrial networks. *Wireless Personal Communications* 2003; **24**(2): 219–232.
22. Glossner J, Iancu D, Lu J, Hokenek E, Moudgill M. A software-defined communications baseband design. *IEEE Communications Magazine* 2003; **41**(1): 120–128.
23. Fu S, Atiquzzaman M. SCTP: State of the art in research, products, and technical challenges. *IEEE Communications Magazine* 2004; **42**(4): 64–76.
24. Fu S, Atiquzzaman M, Ivancic W. Effect of delay spike on SCTP, TCP Reno, and Eifel in a wireless mobile environment. In *11th International Conference on Computer Communications and Networks*, Miami, FL, October 2002; pp. 575–578.
25. Fu S, Atiquzzaman M, Ivancic W. SCTP over satellite networks. In *IEEE 18th Annual Workshop on Computer Communications*, Dana Point, October 2003; pp. 112–116.
26. Atiquzzaman M, Ivancic W. Evaluation of SCTP multistreaming over wireless/satellite links. In *12th International Conference on Computer Communications and Networks*, Dallas, TX, October 2003; pp. 591–594.
27. Ye G, Saadawi T, Lee M. SCTP congestion control performance in wireless multi-hop networks. *MILCOM2002*, Anaheim, California, October 2002; pp. 934–939.
28. Thomson S, Narten T. IPv6 stateless address autoconfiguration. IETF RFC 2462, December 1998.
29. Stewart R, Ramalho M, Xie Q, *et al.* Stream control transmission protocol (SCTP) dynamic address reconfiguration. IETF DRAFT, draft-ietf-tsvwg-addip-sctp-09.txt, June 2004.
30. Awerbuch B, Peleg D. Concurrent online tracking of mobile users. *ACM SIGCOMM Symposium on Communications, Architectures and Protocols*, September 1991; pp. 221–233.
31. Stevens WR. TCP/IP Illustrated, Volume 1 (The Protocols). Addison Wesley: Reading, MA, November 1994.
32. Xie J, Akyildiz IF. An optimal location management scheme for minimizing signaling cost in Mobile IP. *IEEE International Conference on Communications (ICC)*, New York, April 2002; pp. 3313–3317.
33. Chung YW, Sun DK, Aghvami AH. Steady state analysis of P-MIP mobility management, *IEEE Communication Letters* 2003; **7**(6): 278–280.
34. Xie J, Akyildiz IF, A novel distributed dynamic location management scheme for minimizing signaling costs in Mobile IP. *IEEE Transactions on Mobile Computing* 2002; **1**(3): 163–175.
35. Bettstetter C, Hartenstein H, Perez-Costa X. Stochastic properties of the random waypoint mobility model: Epoch length, direction distribution, and cell change rate. In *5th ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems*, September 2002; pp. 7–14.
36. Crovella ME, Bestavros A. Self-similarity in world wide web traffic: evidence and possible causes. *IEEE/ACM Transactions on Networking* 1997; **5**(6): 835–846.
37. Hsieh R, Guang Z, ZheSeneviratne Aruna. S-MIP: A seamless handoff architecture for Mobile IP, *IEEE INFOCOM*, San Francisco, CA, April 2003; pp. 1774–1784.
38. Chen H, Trajkovic LJ. Route optimization in Mobile IP. *Workshop on Wireless Local Networks (WLN)*, Tampa, FL, November 2002; pp. 847–848.

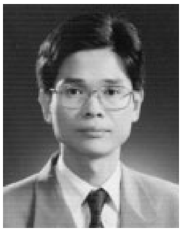
Authors' Biographies

Shaojian Fu received the B.E. degree in Transportation Engineering in 1997, and the M.E. degree in Systems Engineering in 2000, both from Northern Jiaotong University, Beijing, China. During 2000–2001, he worked with Bell Labs China, Lucent Technologies in the area of network surveillance and performance management. He is currently working toward his Ph.D. in the School of Computer Science, University of Oklahoma. His research interests are in the area of wireless communications, IP mobility, transport protocols, network simulation, and embedded reconfigurable computing.



Mohammed Atiquzzaman received the M.Sc. degree and his Ph.D. in Electrical Engineering from the University of Manchester, England. Currently, he is a professor in the School of Computer Science at the University of Oklahoma. He is co-editor-in-chief of *Computer Communications Journal*, and serves on the editorial boards of *IEEE Communications Magazine*, *Telecommunications Systems Journal, Wireless and Optical Networks Journal, and Real-Time Imaging Journal*. He has guest edited many special issues in various journals, and organized special sessions in conferences. He was technical co-chair of HPSR 2003 and the SPIE Quality of Service over Next-Generation Data Networks Conference (2001, 2002, and 2003). He also serves on the technical program committee of many national and international conferences including IEEE INFOCOM, IEEE GLOBECOM, and IEEE International Conference on Computers and Communication Networks. His current research interests are in wireless, satellite, and mobile networks, QoS for next-generation Internet, broadband networks, multimedia over high-speed networks, TCP/IP over ATM, multiprocessor systems, and image processing. He is a coauthor of the book TCP/IP over ATM networks. He has taught many short courses to industry in the area of computer and telecommunication networking. His research has been supported by state and federal agencies like NSF, NASA, U.S. Air Force, Ohio Board of Regents, and DITARD (Australia). He has over 130 refereed publications in the above areas, most of which can be accessed at http://www.cs.ou.edu/~atiq

Liran Ma received the B.E. degree in Electrical Engineering from the Hunan University, Changsha, China, in 1999 and the M.E. degree in Communication and Information System from Northern Jiaotong University, Beijing, China, in 2003. He is currently working towards his Ph.D. computer science at The George Washington University. His research interests include wireless networking and approximation algorithm design.



Yong-Jin Lee received the B.S. and M.S. degrees in Industrial Engineering from Korea University, Seoul, his Korea in 1981 and 1983, respectively. From 1984 to 1986, he worked as a research fellow in KAIST. He received the Ph.D. in Computer Science from Korea University, in 1995. Since 1995, he has been worked in the Department of Computer and Information Science at Woosong University, Taejon, Korea, where he is currently an associate professor. He has also worked as a visiting research scholar from 2003 to 2004 in the School of Computer Science at University of Oklahoma. His research interests include mobile wireless network, transport layer mobility, and Internet QoS.