Design Issues for SIGMA: Seamless IP diversity based Generalized Mobility Architecture

Pulak K Chowdhury, Abu S Reaz, Ta-Chun Lin, Mohammed Atiquzzaman

TR-OU-TNRL-06-101
February 2006

Telecommunication & Networks Research Lab

School of Computer Science

THE UNIVERSITY OF OKLAHOMA

200 Felgar Street, Room 159, Norman, Oklahoma 73019-6151
(405)-325-4042, atiq@ou.edu, www.cs.ou.edu/~atiq

Design Issues for SIGMA: Seamless IP diversity based Generalized Mobility Architecture

Pulak K Chowdhury, **Abu S Reaz**, **Ta-Chun Lin**, **Mohammed Atiquzzaman**
Telecommunication and Networks Research Lab
School of Computer Science
University of Oklahoma,
Norman, OK 73019-6151, USA
Email: {pulak, sayeem_reaz, djlin, atiq }@ou.edu

***Abstract*— In this paper, we propose and evaluate a novel transport layer mobility management scheme: Seamless IP diversity based Generalized Mobility Architecture (SIGMA). SIGMA provides seamless handover for mobile hosts and is based on SCTP, which is a new reliable transport protocol introduced by IETF to transport SS7 signaling messages over IP network. We also show that our handover scheme can greatly reduce the handover latency, packet loss, signaling costs and improve the whole system's throughput compared to the popular Mobile IP based handover schemes.**

I. Introduction

Mobile IP (MIP) [1] is the standard proposed by IETF to support mobile data communication in the IP networks. The mobility support is accomplished by making sure that the Mobile Host (MH) is reachable by its originally assigned IP address even when the MH leaves its home subnet. In order to reach the MH outside its home network area, MIP installs a Home Agent (HA) in the MH's native area to take care of all the packets sent to the MH. The Home Agent (HA) knows about the foreign location of the MH, and forwards all packets addressed for the MH to an agent in the foreign location (called as Foreign Agent (FA)) which finally delivers the packets to the MH. While the MH visits the foreign network, all the connections between the HA and the MH have to be transferred to the FA to keep the ongoing communications alive. This phenomena is referred to as handover. MIP suffers from a number of drawbacks in a mobile computing environment. The most important ones identified to date are high handover latency, high packet loss rate during handover [2], and requirement for change in Internet infrastructure. These drawbacks of MIP in handling handover have been extensively studied in the literature, and several improvements [2] [3] [4] [5] [6] [7] [8] [9] [10] of MIP based handover scheme have been proposed to solve the existing problems. In general, these schemes *focus* on improving three handover parameters, namely handover latency, packet loss, and signaling cost.

When an MH enters a new domain, it needs to perform an IP address reconfiguration at the new point of attachment. Besides, it is necessary for an MH to register its foreign location at the Home Agent (HA) while moving. Thus the handover latency in MIP is primarily due to two procedures: the address resolution at the foreign network and the new address registration with the HA. Accordingly, the handover latency is the overall time taken by an MH to configure a new network care-of address at the Foreign Agent (FA) and to register its new address with the HA. In order to reduce the handover latency, time taken during either or both of these procedures should be reduced.

Fast and Scalable Handover [3], also known as FSHWI, reduces the registration time by using hierarchical structure of the network. Similarly, Fast Handover [4], Proactive Handover [2] [5] are based on Hierarchical Mobile IP [11] which takes the advantages of hierarchical structure of the network to improve the performance of MIP handovers. Even with these enhancements, MIP still can not completely solve the high latency problem, and the resulting packet loss rate is still high [6].

S-MIP [12] proposes another handover scheme based on pure software based movement tracking technology which can reduce the time of both IP address reconfiguration and registration. The work reported in [10] and [13] are also similar to S-MIP, where the reduction of latency is obtained by tracking and predicting host mobility. Sharma et al. [8] proposes yet another novel scheme to overcome the inability of mobility software to sense the signal strengths of multiple access points when operating in an infrastructure-mode wireless LAN. This method reduces the handover latency by reducing the IP address reconfiguration time at the new point of attachment.

Packet or data loss is another issue that has to be dealt with during handover in MIP. Packet loss can be reduced by using complex caching and forwarding techniques between the previous location and the new location. Koodli et al. [7] introduce a technique for fast tunnel set up between the old and new attachment point of the MH as soon as layer 2 handover is detected. The tunnel avoids packet losses which are caused by path set up delay between the mobility domain.

Increased signalling costs during handover is also a prime concern in mobile environments. Micro mobility techniques described in [3], [11], [14], and [15] attempt to reduce the signaling cost by using per domain foreign agents (hierarchical approach). For example, Hierarchical Mobile IP [11] and Hierarchical Mobile IPv6 [14] are extensions of MIP that

The research reported in this paper was funded by NASA Grant NAG3-2922.

support a hierarchy of FAs between the MH and the HA. TeleMIP [15] adds some load balancing features in the basic principles of Hierarchical Mobile IP. Haverinen et al. [16] propose to use paging technology to reduce the total signaling cost. These schemes have been shown to reduce signaling costs.

In spite of above improvements to MIP, there are still unsolved problems during handover. Most of the state-of-the-art handover schemes are based on MIP which is known to have intrinsic flaws that do not fit in handover situations. Those problems include inevitable connection interruption, conflict with IPsec and non-scalable routing. Therefore, significant challenges exist in designing new handover schemes based on mobile IP. To address these problems, we propose a novel IP diversity-based mobility management scheme, called as SIGMA. This scheme minimizes handover latency and packet loss with minimum signaling overhead during handover by exploiting multiple IP addresses to achieve soft handover. Although current mobile devices do not incorporate two interfaces for IP diversity, Software Defined Radios may be used to make one network interface card (NIC) to work as two virtual NICs.

SCTP [17] is a new reliable transport protocol introduced by IETF to transport SS7 signaling messages over IP network and will be used as underlying protocol of SIGMA. It has multi-homing support, which opens new horizons for solving handover problems. By using SCTP and some of its currently proposed extensions, including SCTP Dynamic Address Reconfiguration [18] and Mobile SCTP [19], a seamless handover can be accomplished without any change in the network, but only assisted by the functions embedded in Mobile SCTP enabled servers. A multihomed node is an endpoint with more than one assigned IP address. Each IP address represents a path through the network towards that endpoint, and has separate congestion control variables. SCTP does not know if the paths are completely, partly or not at all distinct from each other, though it is desired that all paths are completely distinct. SCTP can also be used to support TCP based applications by having a middleware which will convert TCP applications to SCTP (changes needed: API calls, etc.). SCTP also supports an ”unreliable mode” which is very similar to UDP.r Although we illustrate SIGMA using SCTP, it is important to note that SIGMA can be used with other protocols that support IP diversity. It can also cooperate with IPv4 or IPv6 infrastructure without the support of MIP.

The objective of this paper is to discuss the design issues related to SIGMA, and based on the design issues, compare the performance of SIGMA with other handover schemes. It is essential that we compare the implementation and deployment issues of SIGMA with other transport layer mobility solutions. This will allow us to have a similar platform of comparison and most importantly, investigate the paradigms of possible improvements stemming from these solutions.

A number of transport layers solutions have been proposed in the literature. MSOCKS [20] uses TCP Splice [21] for connection migration. TCP Splice can be used to split a TCP connection at a proxy by dividing the host-to-host communication into host-proxy and proxy-host communications. Migrate [22] is a transparent mobility management scheme which is based on connection migration using Migrate TCP [22], and uses DNS for location management. Reception Control Protocol (R^2CP) [23], Mobile Multimedia Streaming Protocol (MMSP) [24] and Mobile SCTP (mSCTP) [25] are IP diversity based seamless handover solutions. Indirect TCP (I-TCP) [26] and its derivatives, Mobile TCP (M-TCP) [27] and Mobile UDP (M-UDP) [28], are mobility schemes that require gateways between the communication path of the CN and MH to enable mobility, where the mobility is managed through connection information exchange between these intermediate gateways.

To have a generic understanding of the strength and weakness of SIGMA over these solutions, we compare different aspects of research and implementation issues of SIGMA with these solutions.

The rest of the paper is organized as follows: Section II illustrates the SIGMA architecture and signalling timeline. Section III gives brief introduction to different transport layer solutions for mobility and their implementation issues. In section IV, we present different design issues of SIGMA and also compare the performance of SIGMA with other handover schemes based on these design issues. Section V includes the concluding remarks.

II. SIGMA Architecture

In this section, we give a brief description of SIGMA architecture with detailed handover procedure, location management and signalling diagram.

A. Detailed Handover Procedure of SIGMA

A typical mobile handover in SIGMA using SCTP as an illustration is shown in Fig. 1, where the Mobile Host (MH)is multi-homed node connected through two wireless access networks. Correspondent node (CN) is a single-homed node sending traffic to MH, which corresponds to the services like file downloading or web browsing by mobile users. The handover process of SIGMA can be described by the following five steps using the SCTP protocol.

STEP 1: Obtain new IP address
Refer to Figure 1 as an example, the handover preparation procedure begins when MH moves into the overlapping radio coverage area of two adjacent subnets. Once the MH receives the router advertisement from the new access router (AR2), it should begin to obtain a new IP address (IP2 in Fig. 1). This can be accomplished through several methods: DHCP, DHCPv6, or IPv6 stateless address auto-configuration (SAA) [29].

STEP 2: Add IP addresses into the association
After the MH obtained the IP address IP2 by STEP 1, MH should notify CN about the availability of the new IP address through SCTP Address Dynamic Reconfiguration option [18]. 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

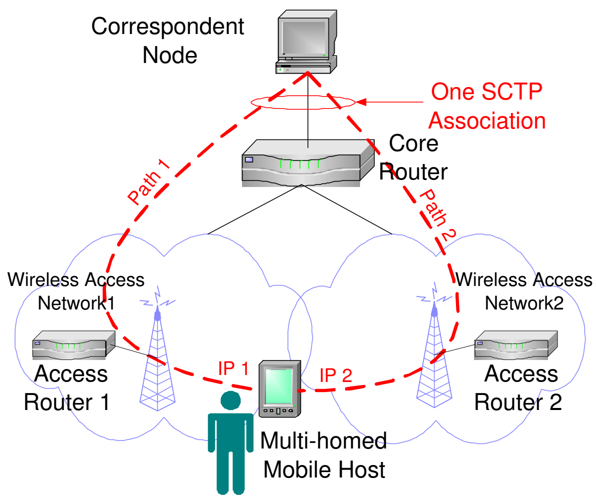

Fig. 1. An SCTP association with multi-homed mobile host

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 MHs IP2.

STEP 4: Update location manager (LM)

SIGMA supports location management by employing a location manager which maintains a database recording the correspondence between MHs identity and MHs 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). 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 CNs available destination IP list. A less aggressive way to prevent CN from sending data to IP1 is MH advertising a zero receiver window (corresponding to IP1) to CN. By deactivating, instead of deleting, the IP address, SIGMA can adapt more gracefully to MHs zigzag movement patterns and reuse the previously obtained IP address (IP1) as long as the IP1s lifetime is not expired. This will reduce the latency and signalling traffic caused by obtaining a new IP address.

B. Timing diagram of SIGMA

Figure 2 summarizes the signalling sequences involved in SIGMA. The numbers before the events correspond to the step numbers in Sec. II-A. Here we assume IPv6 SAA is used by MH to get new IP address. It should be noted that before the old IP is deleted at CN, it can always receive data packets (not shown in the figure) in parallel with the exchange of signalling packets.

C. Location management of SIGMA

SIGMA needs to setup a location manager which, unlike MIP, is not restricted to the same subnet as MHs home network (in fact, SIGMA has no concept of home or foreign network). This will make the deployment of SIGMA much more flexible than MIP. Location management can be achieved as shown by the sequences in Figure 3.

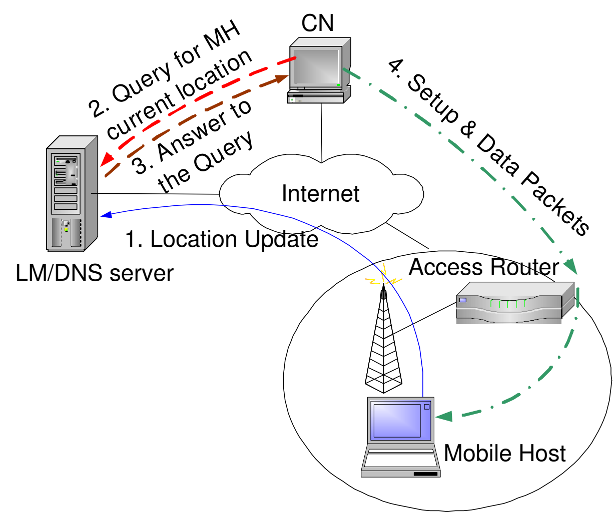

Fig. 3. Location Management in SIGMA

If we use the domain name as MHs 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]. 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 location servers for its registered mobile users. Compared to MIPs 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.

III. Transport Layer Solutions

In current networking system, IP address has *two* most important role: identification of end hosts and routing. In wireless networks, where the mobile host is changing its IP addresses, the real challenge is to make sure the on going connections are routed to the new IP address and the new connection requests can identify the MH using the new address. Most of the transport layer solutions focus on maintaining the ongoing connections rather than location update. The exception is Migrate, which proposes to use DNS for maintaining location information. Migrate is based on Migrate-TCP[22] which releases the current IP address, retrieves a new IP address and reestablishes the ongoing connection with the new IP address. This process of reconnecting to a new connection is called connection migration.

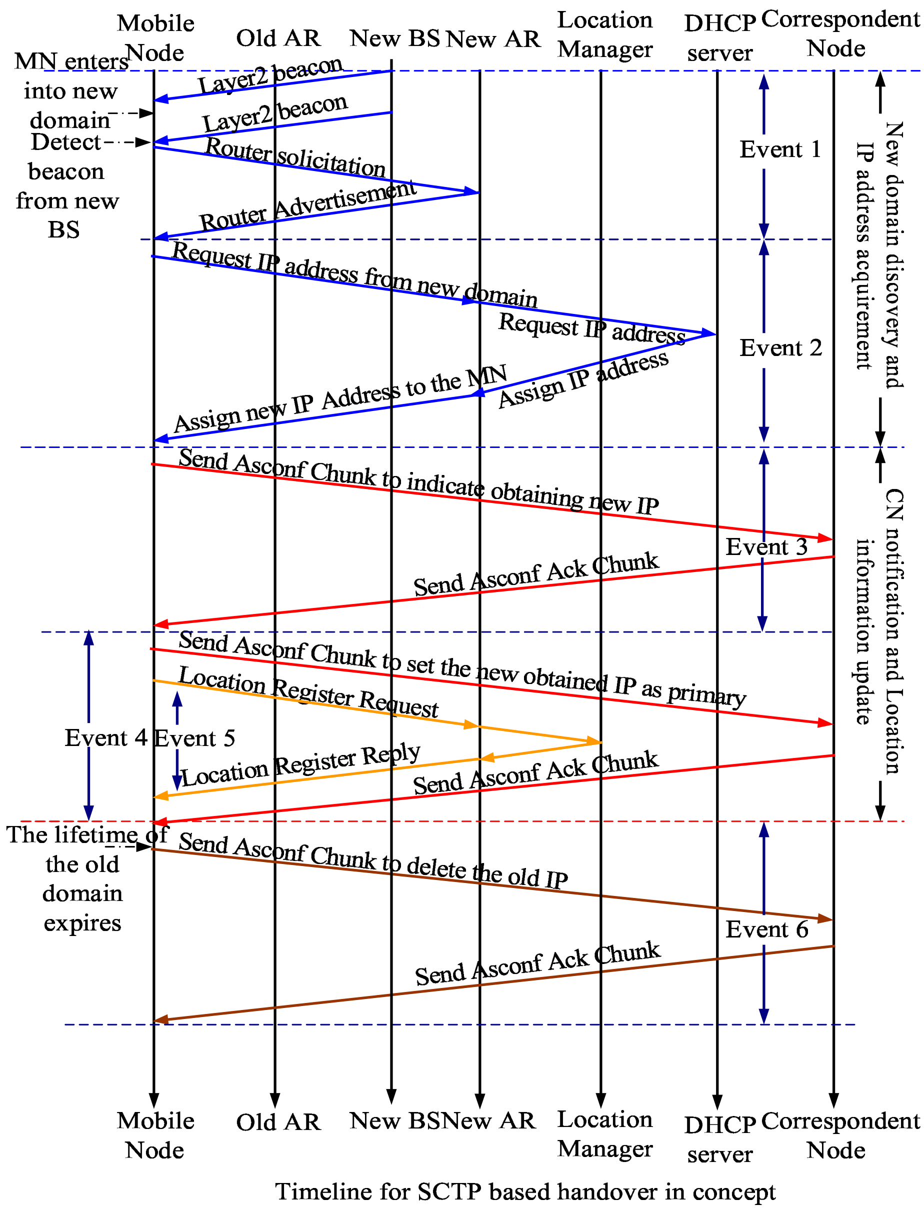

Fig. 2. Timeline of SIGMA

But solutions like R^2CP, MMSP and mSCTP can retrieve the new IP address while still having the old one, so switching connection from the old one to the new one is much seamless. This is possible because these solutions are based on IP diversity, a technique obtain and utilize multiple IP address on a single MH simultaneously. Each of these solutions has different ways to improve the quality of the wireless connection, but the fundamental principle is the same: an IP diversity based seamless transformation of connection between the IP addresses. But these solutions focus on maintaining on-going connections and does not include identification of MH whiles its changing its IP address.

There is another group of solutions that focuses on maintaining on-going connection when the MH changes its IP address. This group uses an intermediate entity between MH and CN to divide the connection. MSOCKS belongs to this group and used proxy as intermediate entity to divide the connection. It uses TCP-Splice [20] that splits the connection between MH and CN and keeps the proxy to CN part of the connection intact. Whenever MH has a new IP address, the connection between MH and proxy releases the old connection and reconnect using the new address. I-TCP and its derivatives M-TCP and M-UDP use gateways as intermediate entity to divide the connection. When the MH changes it IP address, it moves into the coverage of a new gateway. The old gateway sends the connection information to the new gateway to establish the connection with same parameters, so MH views it as the same connection.

All of these solutions are implemented at transport layer and has its pros and cons. Section IV discusses these benefits and limitations and compares them with SIGMA based on different implementation and design issues.

IV. Design Issues for SIGMA

While designing mobility solutions, it is important to follow some important design criteria. This section describes these fundamental design criteria and compare the performances of different mobility solutions based on these issues.

A. Throughput

Throughput refers to the total amount of data that are successfully received at the destination side. Compared to MIP, SIGMA can achieve much higher overall throughput for following two reasons: (1) elimination of triangle routing and (2) less association/session interruption.

1) Triangle Routing: Triangle routing means the packets between CN and MH must be routed along a triangular path longer than the optimal one, which definitely introduces higher latency and high network load in MIP. Both in MIP (in extensions) and MIPv6, the route optimization protocol is proposed to solve the triangle routing problem, which allows packets to be routed along an optimal path from CN to MH [31]. However, the route optimization protocol may cause high signaling and processing costs. In SIGMA, there is no triangle routing because the CN always sends the packets directly to the MH's current IP address.

2) Association/Session Interruption: The design of MIP is founded on the principle that connections based on TCP should survive subnet changes of the MH. However, this opinion is not always agreed due to end to end association interruption during handover. In MIP, MH uses the new CoA (Care of Address) to continue data communication, which might interrupt the previous session due to the delay in CoA registration to HA. The session broken during this time might cause data packet retransmission, and CWND can shrink severely. CWND is a variable that limits the data, in number of bytes, a sender can send to a particular destination transport address before receiving an acknowledgement.

In MIPv6, an MH may use more than one CoA at a time. To assist smooth handovers, an MH should retain its pervious CoA as a (non-primary) CoA, and should still accept packets at this address, even after registering its new primary CoA with its HA [31]. This reduces the chances of session or association interruption.

SIGMA can minimize the possibility of session break by employing the multihoming feature of SCTP. Specifically, SIGMA can always manage to switch to the new path (associated with the new IP address) before the old path failure.

The concept of transport layer handover solution is based on end to end data communication. Solutions like Migrate, MSOCKS, Indirect TCP and its derivatives migrate the connection at the end hosts during handover. It involves disconnecting from one existing connection and reconnecting with a new one; the new connection should be able to identify the old connection and continue. All the packets destined for the host via the old IP address will not be able to reach the destination, resulting in a drop of throughput. If the transport layer protocol is not aware of mobility, the congestion control mechanism would trigger a slow start, resulting in a further drop of throughput. On the other hand, solutions like MMSP, R2CP and mSCTP are based on IP diversity (multi-homing) based soft handover, very similar to SIGMA. This enables them to perform seamless transition from one connection to the next one without noticable drop in throughput.

B. Handover Latency

Handover latency is one of the most important benchmarks for evaluating handover schemes.

1) Handover Latency in MIP: In Section I, we have given the definition of handover latency in MIP-based handover schemes. To be specific, the handover latency of MIP consists of the delay in three parts: (1) address reconfiguration at the new point of attachment, (2) register the new address with the HA, and (3) tunnel setup between HA and FA.

It is important to note that, whenever address reconfiguration starts, ongoing communication is interrupted; resulting in increased handover latency.

2) Handover Latency in SIGMA: SIGMA also employs procedures like address reconfiguration and location update to location server. Moreover, MH also notifies CN about all IP address reconfiguration including Add IP, Set Primary and Delete IP. The definition of handover latency in SIGMA is, however, different because it employs special SCTP features during handover. Here, the handover latency is defined as the service disruption time at the application layer. Service disruption time is the difference between the time instant the first data packet is received by the MH after the handover and the time instant this same packet would be received if there were no handover. SIGMA, thus, only measures the interruption time experienced by the application layer.

In SIGMA, when MH moves from one subnet to another, it gets the knowledge of reaching the coverage of another network by information from layer2 beacons and layer3 advertisement. Layer2 beacon notification is the fastest mean to discover that MH has entered a new domain. The layer3 notification is employed in case if layer2 information is not available. In most scenarios, layer2 triggered new IP address reconfiguration is faster than layer3 triggered one. After getting the knowledge of the new network, the MH requests IP address from the new subnet. The MH then begins to establish a link to the new subnet in addition to the already existing link. It can be regarded as the new IP address (path) preparation for the handover.

Ideally, the Set Primary chuck should be sent timely so that CN can switch to the new IP address before the old IP address becomes unavailable. Therefore, the data stream between CN and MH will not be interrupted because the reliability behavior of transport protocol ensures that all data are sent over the second link in case of the failure of first link. That is the main reason why SIGMA can achieve much less handover latency over MIP based schemes.

On reaching the new subnet, MH may loose the link for its first IP address. There are two possible ways to detect the failure of the path associated with the old IP address,

TABLE I

KEY THROUGHPUT ISSUES

Scheme	Throughput Issues
MIP	• Triangle routing • TCP session interruption • CWND shrinks severely due to timeout or packet loss
SIGMA	• No triangle routing • Seldom association interruption due to multihoming
MIPv6	• No triangle routing if route optimization is used • session interruption may occur if multiple CoAs are not used
Transport Layer Solutions	• IP Diversity based solutions have less fall in the throughput • Migration based solutions has drop in throughput for connection reestablishment. • Transport layer solutions are mobility aware, thus avoid slow start due to loss/delay during handover and avoid further fall in throughput.

TABLE II

KEY HANDOVER LATENCY ISSUES

Scheme	Handover Latency Issues
MIP	• Address reconfiguration at the new point of attachment • Register the new address to the HA • Tunnel setup between HA and FA
SIGMA	• Less than the MIP because it is defined as service disruption time
MIPv6	• same issues as in MIP • several improvements are proposed
Transport Layer Solutions	• IP Diversity based solutions prepares the new path while still receiving data from the old path, so reduces delay during handover • Migration based solutions perform the steps in handover process sequentially, thus have more delay during handover period.

namely layer3 triggered and layer4 triggered detection. Layer3 triggered detection can provide information like unreasonable packet loss or increased RTT, which indicates possible failure of the old path. In layer4 triggered detection, information such as heart beat can be used to detect the failure of old path. Layer3 and layer4 triggering have similar speed in detecting the failure of the old path.

3) Handover Latency in MIPv6: As in MIP, three significant delays are involved during the handover: (a) Move Detection latency (b) Registration latency (c) Binding update latency [31]. To reduce the handover latency in MIPv6, Yegin et al. [7] has proposed two handover schemes: (1) Anticipated Fast Handover Protocol, where the MH or access router has predictive information about the handover which can be helpful to reduce handover latency, (2) Tunnel Based Fast Handover protocol which is similar to the previous one except that there is minimum involvement of layer 3 details. Multiple CoAs can also be used for smooth and seamless handover in MIPv6 [31].

4) Handover Latency in Transport Layer Solutions: The major contributors in handover latency are re-establishment of the new connection and the loss of the packets destined to the old IP address. The solutions that do not prepare the new path before handover would suffer this latency. As described in the previous section, migration based solutions detach from one connection and re-attach to another one. This consists of releasing obsolete address, acquiring new address, reconnecting using the new address and sending packets via the new address. This whole process adds up to the handover delay. IP diversity based solutions acquire the new address and establishes the new path and start communicating via the new path while still receiving data from the old one. Thus, these solutions reduce handover latency.

C. Packet Loss

In MIP, packet loss is caused mainly because of the failure of the old path before the completion of the registration. In MIP, the communication resumes only after the registration is finished. During the period of registration, data packets will be queued at HA. These packets will be dropped if the queue at HA is full or the packets are timed out.

In contrast, whenever there is a path available during the

TABLE III
KEY PACKET LOSS ISSUES

Scheme	Packet Loss Issues
MIP	• Packet loss due to the failure of the old link
SIGMA	• Less than MIP due to multihoming
MIPv6	• Less than MIP if multiple CoAs can be used • Fast handover schemes to reduce packet loss
Transport Layer Solutions	• IP Diversity based solutions can receive packets destined to old address, so reduces packet loss. • Migration based solutions looses on the fly packets.

handover, SIGMA manages to send data packets through the available path. In the ideal case, the Set Primary chunk should be sent timely so that the CN switches to the new IP address before the old IP address becomes unavailable. Therefore, packet loss is minimized by using the multihoming feature of SCTP during the handover. Of course, if there were no path available after leaving the old domain, the packet loss is inevitable.

MIPv6 can use multiple CoA to reduce packet loss during handover. While acquiring a new CoA, MIPv6 can still use old CoA to receive on-the-fly packets; thus reducing packet losses [31]. Fast handover schemes [7] can also be used in MIPv6 to reduce packet loss during handover.

IP Diversity enabled mobile devices can simultaneously utilize multiple IP addresses. Therefore, solutions like MMSP, R2CP and mSCTP reduce packet loss during handover by letting on-the-fly packets to be delivered using the soon-to-be-obsolete IP address while receiving the new packets via the new IP address. Migrate, MSOCKS and Indirect TCP have to release the old IP address before getting the new one. Thus packets already destined to the old IP address will be lost. This handover packet loss is inevitable for hard handover-based transport layer solutions.

D. Handover Signalling Costs

Signaling costs refer to the costs of sending control packets to perform the handover. In MIP the signaling costs includes: (a) Registration to HA and (b) Tunneling setup.

In SIGMA, the necessary signaling includes: (a) Add IP (ASCONF chunk), (b) Set Primary (ASCONF chunk), (c) Delete IP (ASCONF chunk), and (d) Location update.

SIGMA may have higher signaling cost than MIP. However, by introducing better location server structure, SIGMA can greatly reduce the signaling costs for location update. The location server structure will be the hieratical location server like the current DNS server. The main purpose of the improved location server structure is to make the signalling cost for registrations as less as possible.

Furthermore, if we can introduce better macro-mobility protocol for SIGMA to reduce numbers of IP address reconfiguration, the signaling used to Add IP, Set Primary and Delete IP can also be reduced. Macro-mobility will reduce IP address reconfiguration time when MH moves among different BS/AP in the same network domain.

When the mobile host (MH) detects a handover, it generates a new Care of Address (CoA) by IPv6 mechanisms. No FA is needed during this CoA generation. MH then registers its new primary CoA with its HA [31]. After updating its home registration, the MH then updates associated mobility bindings in CN so that CN can perform route optimization [31]. Thus in MIPv6, the signalling costs are similar as in SIGMA due to binding updates in CN.

For the solution based on the infrastructure in the middle, the mobility information is updated at the mobile device and the middle infrastructure. For MSOCKS, the signaling cost would be the cost of updating the proxy. For I-TCP and its derivatives, it would be updating the new gateway + new IP address at CN. For Migrate, the cost would be cost of updating the CN with the new address (request + connection establishment). For MMSP, mSCTP and R2CP, this cost is going to be the sum of first three costs of SIGMA. So in terms of signaling cost, SIGMA is more expensive than most of the transport layer solutions.

E. Deployment

Deployment refers to the required changes at the end hosts, infrastructure modification, and applications.

1) Changes in Applications: In MIP, the presence of firewall and Network Address Translation (NAT) in existing IP network challenge the introduction of home address and care of address. Similar problems also exist with MIPv6. although several improvements are proposed. IPv6 nodes (MH or CN) must also maintain binding update lists for route optimization [31]. In SIGMA, the applications must employ SCTP as underlying transport protocol. Besides, in order to enable the multistreaming feature of the SCTP, the application layer should be SCTP aware. SIGMA has no problem with firewall and NAT because the communication between the MH and CN are always direct and transparent.

Transport layer solutions are end-to-end, thus compatible with NAT and firewall. But some of the schemes like MSOCKS and I-TCP and its derivatives use connection redirection based on gateways. These schemes would conflict with internet security solutions like IPSec.

TABLE IV
KEY SIGNALLING COST ISSUES

Scheme	Signalling Overhead
MIP	• Registration to HA • Tunneling setup
SIGMA	• Add IP (ASCONF chunk) • Set Primary (ASCONF chunk) • Delete IP (ASCONF chunk) • Location update
MIPv6	• New CoA registration to HA • Binding updates to CN
Transport Layer Solutions	• MSOCKS: update proxy • Migrate: Request + connection reestablishment • I-TCP and Derivatives: updating the new gateway + new IP address at CN • MMSP, mSCTP and R2CP: Add IP + Set Primary + Delete IP

TABLE V
KEY DEPLOYMENT ISSUES

Scheme	Deployment Issues
MIP	• Introduction of HA • Introduction of FA
SIGMA	• Employing SCTP as underlying transport protocol • Introduction of the location server
MIPv6	• No FA is needed, although HA is still required • For route optimization, special supports are needed in the IPv6 nodes
Transport Layer Solutions	• MSOCKS: requires proxy • I-TCP and Derivatives: requires gateways for subnets and inter gateway communication • Migrate, MMSP, mSCTP and R2CP: requires respective transport layer protocols

2) Infrastructure modification: MIP introduces HA and the widely distributed FA to support host mobility; requiring changes in the existing Internet infrastructure. SIGMA only introduces the location server for better location management. Existing DNS server in the Internet can be used as location managers in SIGMA [32]. The basic problem to support mobility is namespace problem: endpoint with a "single" name should be reachable via different links/paths. One good solution for client mobility management is the application of SCTP (used in SIGMA) which can eventually solve the requirements of transport layer mobility in the Internet. In MIPv6, there is no need to deploy FAs as in MIP. MIPv6 can operate in any location without any special support required from the local router.

As described above, transport layer solutions are end-to-end, thus do not need any additional infrastructure. But MSOCKS uses connection redirection based on proxy, so do I-TCP and its derivatives based on gateways. So mobility aware proxies or gateways are added infrastructures in the network that would be required to implement mobility with these schemes.

All the transport layer solutions are required to have mobility aware transport layer at the MH. MSOCKS require mobility aware transport layer in the intermediate proxy but not at the CN. For I-TCP and its derivatives, the transport layer at gateways is mobility aware, but not at CN. For Migrate, both MH and CN has mobility aware transport layer.

3) Changes in the end hosts: In MIP, MH has to be aware of the mobility, while there requires no change to CN. SIGMA requires both MH and CN to use SCTP as underlying transport protocol. To support route optimization, mobile host and HA functionality, all IPv6 nodes must support home address destination option, type 2 routing header or the mobility header [31].

In summary, compared to MIP, SIGMA requires less modification to the current of IP network except the application which must employ SCTP as underlying transport protocol. By using SCTP and some of its currently proposed extensions a seamless handover can be fully accomplished in the mobile client without any changes in the network. SIGMA only needs to introduce some mechanisms to do the location registration.

F. Scalability, Availability and Fault Tolerance

The scalability refers to the ability of handling huge amount of simultaneous node mobilities. In MIP, the HA is often seen as the "heart" of the mobility system. Therefore, HA usually hosts a large number of subscribers that makes it vulnerable to the problem of single point. If the HA fails, MH does not learn about HA failure until re-registration. During HA faliure, no packet will be forwarded to MH and to make situation worse, the entire MIP mechanism will not work properly.

Similarly, the SIGMA's scalability bottleneck lies in the location server. The situation of location server in SIGMA is quite similar to HA in the MIP except that the chance of CN disconnection from the MH is greatly reduced by the multihoming feature of SCTP. Therefore, one of the future improvements to SIGMA will be the better location server structure. The candidate solutions are the hieratical location server with caching and distributed location server.

Using the hieratical location server structure with caching can reduce the amounts of location update and effectively reduce the load of the central location server.

The distributed location server can achieve the same effect as the hierarchical location server via distributing the location server network wide. Since the location server is distributed, the load is balanced among all the location servers. Also, the time to perform the location update is reduced because the MH can always register to the closest location server.

In MIPv6, though there is no need for FA, HA still acts as a broker to handle packets destine to MH. Therefore, the same problem exists. If HA is down, packets sent from new CN will get dropped in home network. MIPv6 supports multiple home agents so that when HA is down or unavailable for the reconfiguration of the home network, MH can use "dynamic home agent address discovery" to discover a new HA [31].

Transport layer schemes like MMSP, mSCTP or R2CP are handover protocols which handles handover at the end hosts, thus scalable to any network topology. I-TCP and its derivatives are dependant on the coverage of gateways as the mobility is handled via the connection handover across gateways. So, these schemes are bounded by the presence of gateways. MSOCKS uses proxy server to split a connection and mobility is handled under the coverage of the proxy. So this can be implemented only within a given domain under a single proxy server. These schemes are all dependant on their intermediate nodes; so if these nodes fails, the scheme would be inactive. Location management system in Migrate is very similar, thus if the MH acts as a server, the failure at the location server would allow the new connection requests from CN to fail, thus making this scheme vulnerable at the location server, provided that the MH is a server.

G. Security

1) Firewall: Firewalls block all classes of incoming packets that do not meet specified criteria. Furthermore, enterprise firewalls are typically configured to block packets from entering via the Internet that appear to be originated from internal computers.

Firewalls, in particular, cause difficulty for MIP because packets originating from the MH carry the MH's home address, and would thus be blocked by the firewall. Although this permits management of internal Internet nodes without great attention to security, it presents difficulties for MH wishing to communicate with other nodes within their home enterprise networks. Therefore, a great deal of attention is being focused on making MIP to coexist with the security features coming into use in the Internet. Gupta et al. [33] has proposed a firewall traversal solution for MIP.

SIGMA does not use the home address to identify the MH; requiring no capsulation of packets. Therefore, it will work harmonically with the firewall because there is always real end to end communication in SIGMA after the handover.

Though some of the firewall now can handle IPv6 packets, most of the firewalls do not know about the MIPv6 control message. So it is hard to complete the full registration and binding process since the firewall cannot interpret the information carried in the MIPv6 packet header. Even more, in MIPv6, the communication between MH and CN requires authentication and encapsulation by IPSec protocol. Most firewalls either block the IPSec protocol or do not support it. If the firewall is MIPv6-aware, Shen et al. [34] issues the new firewall detection and detection reply message which are used to indicate the existence of firewall before the binding update message.

All the transport layer solutions carry the IP address of the domain it is currently visiting. So no tunneling/encapsulation are required. Thus firewall should not be an issue with these solutions.

2) Ingress filtering: Ingress filtering means that the border routers discard packets coming from within the enterprise if the packets do not contain a source IP address configured for one of the enterprise's internal networks.

In MIP, complications are also presented by ingress filtering operations because MH would use their home address as the source IP address of the packets they transmit. Solutions to this problem in MIP typically involve tunneling outgoing packets from the CoA, but then the difficulty is how to find a suitable target for the tunneled packet from the MH. Montenegro et al. [35] has proposed the use of reverse tunnels to the HA to counter the restriction imposed by ingress filtering.

In SIGMA, whenever there is communication between the MH and CN, the packets will use source IP address configured for one of the enterprise's internal networks. Therefore, SIGMA will have no problem working with the ingress filtering.

MIPv6 mobile node uses care-of address as source address in foreign network. Correspondent node uses IPv6 routing header rather than IP encapsulation, so there is no ingress filtering problem in MIPv6 [36].

In transport Layer Solutions, the source IP address of the mobile node is always going to be the one that the MH has obtained from the domain it is in. So the ingress filter would not discard packets destined from the MH.

TABLE VI
SCALABILITY, AVAILABILITY AND FAULT TOLERANCE

Scheme	System Bottleneck Issues
MIP	• Rely on HA : single point of faliure • Tunneling
SIGMA	• Rely on Location Server
MIPv6	• No need for HA • HA still acts as broker • Supports multiple home agents
Transport Layer Solutions	• MSOCKS: rely on proxy - single point • I-TCP and Derivatives: rely on gateways - single point • Migrate: Rely on Location Server - single point

TABLE VII
KEY SECURITY ISSUES

Scheme	Security Issues
MIP	• Interferes with firewall • Interferes with ingress filtering
SIGMA	• cooperates with firewall • cooperates with ingress filtering
MIPv6	• Not fully firewall cooperative • Still problem exists with ingress filtering
Transport Layer Solutions	• Cooperation with firewall • Cooperation with ingress filtering.

H. Operation in Heterogeneous Environment

Both MIP and SIGMA do not exhibit any problem while working in heterogenous environment. SIGMA will operate well in heterogeneous network environment because it is a transport layer handover scheme. It should have no problem with all kinds of underlying access technology such as WLAN, Cellular network and satellite network.

For transport Layer Solutions, I-TCP and its derivatives depend on the gateways of the domain the MH is currently visiting to handle mobility. So, for mobility across heterogeneous networks, the gateways should be aware of each other and should be able to transfer the connection from one gateway to another. Usually different networks are deployed under different administration, so this solution is not feasible. MSOCKS depends on proxy server to split connection; so mobility across heterogeneous network is not possible as different networks would not be under a single proxy. Migrate, MMSP, mSCTP and R2CP are end-to-end schemes. So they can work under this kind of multiple-network scenario.

I. QoS

Quality of service (QoS) in network environment conventionally refers to four pivot attributesreliability, delay, jitter, and bandwidth [37].

1) **Reliability:** The rate and probability of packet loss and bit error during transit. It is achieved through the error detection and error correction methods to make sure that the packets received are correct. Otherwise, those damaged packets will be retransmitted until the receiver gets the correct ones. Services such as file transfer and e-mail require high reliability. On the other hand, for video streaming service, we can accept some dropped frame.
2) **Delay:** The amount of time requires forwarding one packet from one point to another. Some applications such as web surfing and voice over IP, which have higher real-time requirement, are delay sensitive. Algorithms utilize the idea of buffering can smooth the effect of delay.
3) **Jitter:** Packets arrive at the destination in irregular time intervals will cause jitter effect.
4) **Bandwidth:** Rate at which packets are transited in a network.

QoS in MIP has been developed with several techniques, including buffering, admission control, packet scheduling, reservation protocol (RSVP), and DiffServ. But different design issues arise for general MIP architecture. The route alone the path from MH to CN does not guarantee to provide

TABLE VIII
KEY QOS ISSUES

Scheme	QoS Issues
MIP	• Interruption during handover • Require QoS support in routing • Require QoS support in tunneling
SIGMA	• Lower handover latency
MIPv6	• Handover latency still exists
Transport Layer Solutions	Not much research yet

QoS. Andreas et al. [38] provides Mobile IP with Location Registers (MIP-LR), an alternative on MIP protocol, where sender queries the location of the mobile host before sending the packets. It not only reduces the triangle routing in MIP, but also results in longer initial latency. The third one is tunneling. In MIP, the tunneling technique is heavily used to send encapsulated packets from MH to CN. Both two end-points, FA and HA, have to agree on the QoS parameters and utilize them before the end-to-end connection is built. Shing et al. [39] addresses several problems applying RSVP over MIP, including tunneling, and provides a possible solution.

In SIGMA, the handover latency and packet loss rate is alleviated utilizing multihoming. On the other hand, since SIMGA utilizes SCTP protocol with multiple interfaces, the power management is a big issue if applied to handheld devices such as PDA.

To MIPv6, several enhancements have been proposed to reduce handover latency. Hierarchical MIPv6 (HMIPv6) [40] proposes a hierarchical mobile agent architecture to reduce the registration latency from MH to HA. Fast handovers for MIPv6 (FMIPv6)[9] introduces a mechanism to configure a new IP address before entering a new subnet, which reduces the handover latency. And the hybrid of the two gains better performance. However, the handover latency still exists [12].

QoS in transport layer is handled in different phases. In I-TCP[41], the mobile support router will establish the TCP connection to host on fixed network for mobile host, which solves the unreliability problem in wireless communication. Hari et al.[42] issued the snoop protocol which also uses mobile support router for handling packets and retransmission. Samaraweera[43] utilizes the round trip delay to specify non-congested packet loss from congested one, which helps to reduce the TCP error recovery time.

J. Implementation in Space Networks

NASA [44] is using Mobile IP based schemes (MIP and MIPv6) to build future space communication networks. The limitations of MIP based schemes discussed before in this paper are also applicable for space networks. SIGMA can be used in the space networks and provides smooth handover between space crafts [45]. Other transport layer based solutions have not shown their applicability in space networks.

V. CONCLUSION

This paper presents the design issues of SIGMA: a Seamless IP diversity based Generalized Mobility Architecture. After the introduction of SIGMA architecture, we compare the performance of SIGMA with different handover schemes in the literature. From the discussions, it can be concluded that for typical network parameters, SIGMA has a lower handover latency, lower packet loss rate and higher throughput than MIPv6 enhancements. It has been shown that SIGMA has a higher survivability than MIP thanks to its location management scheme. SIGMA can also easily interoperate with existing network security infrastructures such as Ingress filtering and IPSec. SIGMA can be deployed in the current IP network without much infrastructure change. SIGMA can also be applicable in space networks for performing inter-satellite handovers.

REFERENCES

[1] C. Perkins, “IP mobility support.” IETF RFC 3344, August 2002.

[2] K. Malki, P. Calhoun, T. Hiller, J. Kempf, P. McCann, A. Singh, H. Soliman, and S. Thalanany, “Low latency handover in mobile IPv4.” Internet Draft, Mobile IP Working Group, June 2002.

[3] R. C’aceres and V. N. Padmanabhan, “Fast and scalable handover for wireless internetworks,” *Proceedings of Mobicom 96*, Rye, New York, USA, pp. 56–66, 10-12 November 1996.

[4] K. El-Malki and H. Soliman, “Fast handovers in mobile IPv4.” Internet Draft, draft-elmalki-mobileip-fast-handovers-03.txt, September 2000.

[5] P. Calhoun, T. Hiller, J. Kempf, P. McCann, C. Pairla, A. Singh, and S. Thalanany, “Foreign agent assisted hand-off.” Internet Draft, draft-ietf-mobileip-proactive-fa-03.txt, November 2000.

[6] R. Hsieh and A. Seneviratne, “A comparison of mechanisms for improving mobile IP handoff latency for end-to-end TCP,” *ACM MobiCom*, San Diego, USA, pp. 29–41, September 2003.

[7] A. Yegin, C. Perkins, G. Tsirtsis, K. El-Malki, and M. Khalil, “Fast handovers for mobile ipv6.” Internet Draft, draft-ietf-mobileip-fast-mipv6-04.txt, March 2002.

[8] Srikant Sharma, Ningning Zhu, and Tzi cker Chiueh, “Low-latency mobile ip handover for infrastructure-mode wireless LANs,” *IEEE Journal on Selected Areas in Communication, Special issue on All IP Wireless Networks*, vol. 22, no. 4, pp. 643–652, May 2004.

[9] R. Koodli (editor), “Fast handovers for mobile ipv6.” Internet Draft, draft-ietf-mipshop-fast-mipv6-03.tx, October 2004.

[10] A. C. Snoeren and Hari Balakrishnan, “An end-to-end approach to host mobility,” *Sixth ACM/IEEE International Conference on Mobile Computing and Networking*, Boston, MA, USA, pp. 155–66, 6-11 August 2000.

[11] E. Gustafsson, A. Jonsson, and C. Perkins, “Mobile IP regional registration.” Internet Draft, draft-ietf-mobileipreg-tunnel-04.txt, March 2001.

[12] R. Hsieh, Z.-G. Zhou, and A. Seneviratne, “S-MIP: A seamless handover architecture for mobile IP,” *Proceedings of IEEE INFOCOM*, San Francisco, CA, USA, pp. 1774–1784, 30 March - 3 April 2003.

[13] William Su, Sung-Ju Lee, and Mario Gerla, "Mobility prediction in wireless networks," *Proceedings of IEEE MILCOM 2000*, Los Angeles, CA, USA, pp. 491–5 vol.1, 22-25 October 2000.

[14] H. Soliman, C. Castelluccia, K. Elmalki, and L. Bellier, "Hierarchical mobile IPv6." Internet Draft, draft-ietf-mobileip-hmipv6-04.txt, 2001 July.

[15] S. Das, A. Misra, P. Agrawal, and S.K. Das, "TeleMIP, telecommunication-enhanced mobile IP architecture for fast intradomain mobility," *IEEE Personal Communications*, vol. 7, no. 4, pp. 50–58, August 2000.

[16] H. Haverinen and J. Malinen, "Mobile IP regional paging." Internet Draft, draft-haverinen-mobileip-reg-paging-00.txt, June 2000.

[17] R. Stewart and Q. Xie, "Stream control transmission protocol." RFC 2960, October 2000.

[18] R. Stewart, M. Ramalho, and Q. Xie, "Stream control transmission protocol (SCTP) dynamic address reconfiguration." Internet Draft, draft-ietf-tsvwg-addip-sctp-08.txt, June 2004.

[19] M. Riegel and M. Tuexen, "Mobile sctp." Internet Draft, draft-riegel-tuexen-mobile-sctp-03.txt, August 2003.

[20] D. A. Maltz and P. Bhagwat, "MSOCKS: an architecture for transport layer mobility," *IEEE INFOCOM*, San Francisco, CA, pp. 1037 – 1045, March 29 - April 2, 1998.

[21] D. Maltz and P. Bhagwat, "TCP splicing for application layer proxy performance," IBM Research Report RC 21139, IBM T.J. Watson Research Center, March 1998.

[22] A. C. Snoeren and H. Balakrishnan, "An end-to-end approach to host mobility," *MOBICOM*, Boston, MA, pp. 155 – 166, August 6-11, 2000.

[23] H. Hsieh, K. Kim, Y. Zhu, and R. Sivakumar, "A receiver-centric transport protocol for mobile hosts with heterogeneous wireless interfaces," *ACM MobiCom*, San Diego, CA, pp. 1 – 15, September 14-19, 2003.

[24] H. Matsuoka, T. Yoshimura, and T. Ohya, "End-to-end robust IP soft handover," *IEEE ICC*, Anchorage, Alaska, pp. 532 – 536, May 11-15, 2003.

[25] S.J. Koh, M. J. Chang, and M. Lee, "mSCTP for soft handover in transport layer," *IEEE Communications Letters*, vol. 8, no. 3, pp. 189 – 191, March 2004.

[26] A. Bakre and B. R. Badrinath, "I-TCP: indirect TCP for mobile hosts," *IEEE International Conference on Distributed Computing Systems*, Vancouver, Canada, pp. 136 – 143, May 30 - June 2, 1995.

[27] Z. J. Haas and P. Agrawal, "Mobile-TCP: an asymmetric transport protocol design for mobile systems," *IEEE ICC*, Montreal, Canada, pp. 1054 – 1058, June 8 - 12, 1997.

[28] K. Brown and S. Singh, "M-UDP: UDP for mobile cellular networks," *Computer Communication Review*, vol. 26, no. 5, pp. 60 – 78, October 1996.

[29] S. Thomson and T. Narten, "IPv6 stateless address autoconfiguration." IETF RFC 2462, December 1998.

[30] B. Awerbuch and D. Peleg, "Concurrent online tracking of mobile users," *ACM SIGCOMM Symposium on Communications, Architectures and Protocols*, Zurich, Switzerland, p. 221233, September 1991.

[31] J. Arkko D. Johnson, C. Perkins, "Mobility support in IPv6." RFC 3775, June 2004.

[32] A. S. Reaz, M. Atiquzzaman, and S. Fu, "Performance of DNS as Location Manager," *IEEE Electro/Information Technology Conference*, Lincoln, NE, 22-25 May 2005.

[33] G. Montenegro and V. Gupta, "Sun's skip firewall traversal for mobile ip." RFC 2356, June 1998.

[34] Yang Shen, Zhang Sidong, Zhang Hongke, and Miao Fuyou, "Firewall traversal for mobile IPv6." Internet Draft, draft-miao-mip6-ft-00, September 2005.

[35] G. Montenegro and P. Nikander, "Protecting against bidding down attacks." Internet Draft, draft-montenegro-mipv6sec-bit-method-00.txt, October 2002.

[36] S. Thiruvengadam, H. Tschofenig, and F. Le, "Mobile IPv6 - NSIS interaction for firewall traversal." Internet Draft, draft-thiruvengadam-nsis-mip6-fw-02.txt, April 2005.

[37] *Computer Networks*, New Jersey, USA: Prentice Hall PTR, 4 ed., 2003.

[38] Andreas Terzis and Lixia Zhang, "Reservations for aggregate traffic: Experiences from an rsvp tunnels implementation," *International Workshop on QoS, IWQoS*, 1998.

[39] Shing-Jiuan Leu and Ruay-Shiung Chang, "Integrated service mobile internet: Rsvp over mobile ipv4&6," *Mobile Networks and Applications*, vol. 8, pp. 635–642, 2003.

[40] H. Soliman, C. Catelluccia, and K.E. Malki et al., "Hierarchical mobile ipv6 mobility management (hmipv6)." Internet Draft, draft-ietfmipshop-hmipv6-04.txt, December 2004.

[41] Ajay Bakre and B.R. Badrinath, "I-TCP: Indirect TCP for mobile hosts," *International Conference on Distributed Computing Systems*, 1994.

[42] Hari Balakrishnan, Srinivasan Seshan, Elan Amir, and Randy H. Katz, "Improving tcp/ip performance over wireless networks," *MobiCom*, 1995.

[43] N.K.G. Samaraweera, "Non-congestion packet loss detection for TCP error recovery using wireless links," *IEEE proceeding of Communication*, 1999.

[44] Kent Leung, Dan Shell, William Ivancic, David H. Stewart, Terry L. Bell, and Brain A. Kachmar, "Application of mobile-IP to space and aeronautical networks," *IEEE Aerospace Conference Proceedings*, Big Sky, MT, USA, pp. 1027–33, 10-17 March 2001.

[45] S. Fu and M. Atiquzzaman, "SIGMA: A transport layer mobility management scheme for terrestrial and space networks." book chapter to be published by Kluwer Academic Publishers, 2005. www.cs.ou.edu/~netlab.