e-Business and Telecommunication Networks

edited by

João Ascenso

ISEL,
Lisbon, Portugal

Luminita Vasiu

University of Westminster,
London, UK

Carlos Belo

IST/IT,
Lisbon, Portugal

and

Mónica Saramago

INSTICC,
Setúbal, Portugal

Springer

SIGMA: A TRANSPORT LAYER MOBILITY MANAGEMENT SCHEME FOR TERRESTRIAL AND SPACE NETWORKS*

Shaojian Fu and Mohammed Atiquzzaman
Telecommunications and Networks Research Lab
School of Computer Science, University of Oklahoma,
Norman, OK 73019-6151, USA
Email: {sfu,atiq}@ou.edu

Keywords: Internet Mobility, Mobility Management, Wireless Networks, Handoff management.

Abstract: Mobile IP has been developed to handle mobility of Internet hosts at the network layer. Mobile IP 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 the architecture of Seamless IP diversity-based Generalized Mobility Architecture (SIGMA) - a new mobility management scheme. SIGMA utilizes IP diversity to achieve seamless handover, and is designed to solve many of the drawbacks of Mobile IP, including requirement for changes in infrastructure. The survivability and security of SIGMA is evaluated and shown that SIGMA has a higher survivability than Mobile IP - thanks to its centralized location management scheme. SIGMA can interoperate with existing network security infrastructures such as Ingress filtering and IPSec fairly easily. We also show the application of SIGMA to manage satellite handovers in space networks.

1 INTRODUCTION

Mobile IP (MIP) (Perkins, 2002; Perkins, 1998) has been designed to handle mobility of Internet hosts at the network layer. It allows a TCP connection to remain alive and receive packets 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 (Malki, 2003), and requirement for change in Internet infrastructure. Mobile IP 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) register the new CoA with the HA, and (iv) forward 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 (Koodli, 2004), giving rise to a large handover latency and high packet loss rate.

*The research reported in this paper was funded by NASA Grants NAG3-2528 and NAG3-2922.

MIP is known to have conflict with network security solutions (Perkins, 1998). Base MIP does not cooperate well when the HA is behind a firewall and the MH is outside the firewall, unless firewall transversal solution (Montenegro and Gupta, 1998) is used. Moreover, base MIP has difficulty in the presence of a foreign network which implements ingress filtering, unless reverse tunnelling, where the HA's IP address is used as the exit point of the tunnel, is used to send data from the MH.

1.1 Recent Research on Improving Mobile IP

Many improvements to Mobile IP have been proposed to reduce handover latency and packet loss. IP micromobility protocols like Hierarchical IP (Gustafsson et al., 2001), HAWAII (Ramjee et al., 1999) and Cellular IP (Cambell et al., 1999) 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 (Malki, 2003) uses pre-registrations and post-registrations which are based on utilizing link layer event triggers to reduce handover latency.

J. Ascenso et al. (eds.), e-Business and Telecommunication Networks, 41–52.

Optimized smooth handoff (Perkins and Wang, 1999) not only uses a hierarchical FA structure, but also queues packets at the visited FA buffer and forward packets to MH's new location. To facilitate packet rerouting after handover and reduce packet losses, Jung et.al. (Jung et al., 2002) introduces a location database that maintains the time delay between the MH and the crossover node. Mobile Routing Table (MRT) has been introduced at the home and foreign agents in (Wu et al., 2002), and a packet forwarding scheme similar to (Perkins and Wang, 1999) is also used between FAs to reduce packet losses during handover. A reliable mobile multicast protocol (RMMP), proposed in (Liao et al., 2000), uses multicast to route data packets to adjacent subnets to ensure low packet loss rate during MH roaming. In (Fu and Atiquzzaman, 2003), Fu et al. use SCTP, a new transport layer protocol, to improve the performance of MIP by utilizing SCTP's unlimited SACK Gap Ack Blocks (Fu et al., 2005).

Mobile IPv6 (Johnson et al., 2004) 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 (Johnson et al., 2004). Fast Handovers for Mobile IPv6 (FMIPv6) (Koodli, 2004), aims to reduce handover latency by configuring a new IP address before entering a 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) (Soliman et al., 2004) also introduces a hierarchy of mobile agents to reduce the registration latency and the possibility of an outdated Collocated CoA (CCOA). FMIPv6 and HMIPv6 can be used together, as suggested in (Soliman et al., 2004), to improve the performance further (in this paper, we refer to this combination as FHMIPv6). 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. Even with the above enhancements, Mobile IP still can not completely remove the latency resulting from the four handover steps mentioned earlier, resulting in a high packet loss rate (Hsieh and Seneviratne, 2003).

1.2 Motivation of SIGMA

As the amount of real-time traffic over wireless networks keeps growing, the deficiencies of the network layer based Mobile IP, in terms of 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 (Maltz and Bhagwat, 1998) and connection migration solution (Snoeren and Balakrishnan, 2000). 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 resulting from handovers. As a result, the handover latency for these schemes is in the scale of seconds.

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 (Rappaport, 1996; Caire et al., 1998). Recently, increasing number of mobile nodes are equipped with multiple interfaces to take advantage of overlay networks (such as WLAN and GPRS) (Holzbock, 2003). The development of Software Radio technology (Glossner et al., 2003) 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 (Fu and Atiquzzaman, 2004). In the field of mobile and wireless communications, the performance of SCTP over wireless links (Fu et al., 2002), satellite networks (Fu et al., 2003; Atiquzzaman and Ivancic, 2003), and mobile ad-hoc networks (Ye et al., 2002) 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.

The *objective* of this paper is to describe the architecture, survivability, and security of a new scheme for supporting low latency, low packet loss mobility management scheme called Transport Layer Seamless Handover (SIGMA). We also show the applicability of SIGMA to mange handoffs in space networks. Similar in principle to a number of recent transport layer handover schemes (Koh et al., 2004; Xing et al., 2002; Li, 2002), the basic idea of SIGMA is to decouple location management from data transfer, and achieve seamless handover by exploiting IP diversity to keep the old path alive during the process of setting up the new path during handover. 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 IPv4 or IPv6 infrastructure without any support from Mobile IP.

1.3 Contributions of Current Research

The contributions of this paper are:

- Propose and develop transport layer based seamless handover (SIGMA). Here, "seamless" means low latency and low packet loss.
- Adapt SIGMA for satellite handovers in space networks.
- Evaluate the survivability and security of SIGMA, and compare with those of MIP.

1.4 Structure of this Paper

The rest of this paper is structured as follows: First, Sec. 2 describes the basic concept of SIGMA, including handover signalling procedures, timing diagram, and location management of SIGMA. We then apply the concept of SIGMA for satellite handovers in Sec. 3. The survivability and security issues of SIGMA are evaluated in Secs. 4 and 5, respectively. Finally, concluding remarks are presented in Sec. 6.

2 ARCHITECTURE OF SIGMA

In this section, we outline SIGMA's signalling procedure for mobility management in IP networks. The procedure can be divided into five parts which will be described below. The main idea of SIGMA is to decouple location management from data transfer, and achieve seamless handover by exploiting IP diversity to keep the old path alive during the process of setting up the new path during handover.

In this paper, we illustrate SIGMA using SCTP. SCTP's multi-homing allows an association between two end points to span multiple IP addresses or network interface cards. An example of SCTP multihoming is shown in Fig. 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 bottom. One of the links is designated as the primary while the other can be used as a backup link in the case of failure of the primary, 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 Fig. 2, where MH is a multi-homed node connected to two wireless access networks. Correspondent node (CN) is a node sending traffic to MH, representing services like file download or web browsing by mobile users.

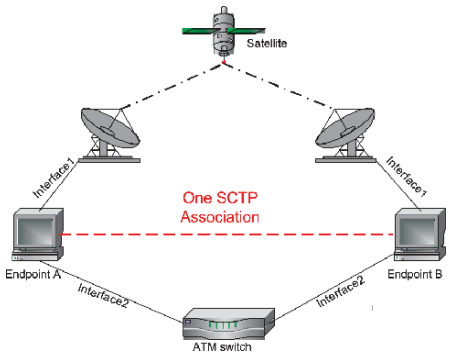

Figure 1: An SCTP association with multi-homed endpoints.

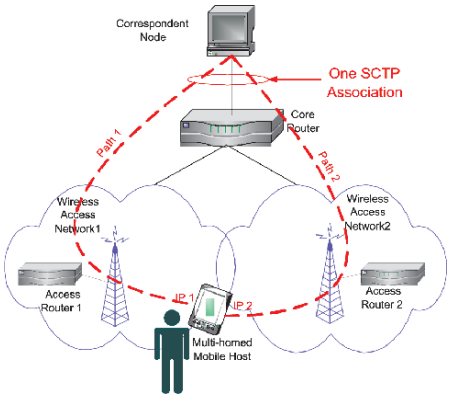

Figure 2: An SCTP association with multi-homed mobile host.

2.1 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 Fig. 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. As an example, in IEEE802.11 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 Fig. 2). This can be accomplished through several methods: DHCP, DHCPv6, or IPv6 stateless address auto-configuration (SAA) (Thomson and Narten, 1998).

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 (Stewart et al., 2004). 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 need to modify its local routing table to make sure the future outgoing packets to CN using new path through AR2.

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 his 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 can not 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 signalling traffic caused by obtaining a new IP address.

2.2 Timing Diagram of SIGMA

Figure. 3 summarizes the signalling sequences involved in SIGMA, the numbers before the events correspond to the step numbers in Sec. 2.1. Here we assume IPv6 SAA is used for MH to get new IP address. It should be noted that before the old IP is deleted at CN, it can receive data packets (not shown in the figure) in parallel with the exchange of signalling packets.

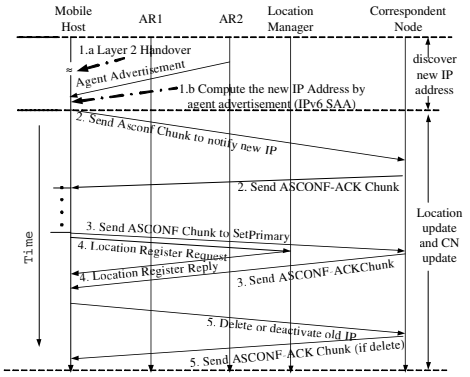

Figure 3: Timing diagram of SIGMA.

2.3 Location Management

As mentioned in STEP 4 of Sec. 2.1, 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

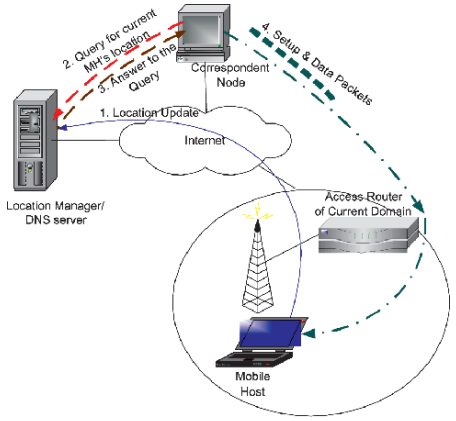

Figure 4: Location management in SIGMA.

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). The location of the LM does not have impact 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 Fig. 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 (Awerbuch and Peleg, 1991). 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 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. Moreover, the survivability of the whole system will also be enhanced as discussed in Sec. 4.

3 SIGMA-SN: SIGMA IN SPACE NETWORKS

Spacecrafts, such as satellites, communicate among themselves and with ground stations on the earth to enable space communications. Depending on the altitude, satellites can be classified into three types: Low Earth Orbit (LEO), Medium Earth Orbit (MEO) and Geosynchronous Earth Orbit (GEO). GEO satellites are stationary with respect to earth, but LEO and MEO satellites move around the earth, and are handed over between ground stations as they pass over different areas of the earth. This is analogous to mobile computers being handed over between access points as the users move in a terrestrial network.

The National Aeronautics and Space Administration (NASA) has been studying the use of Internet protocols in spacecrafts for space communications (Bhasin and Hayden, 2002). For example, the Global Precipitation Measurement (GPM) project is studying the possible use of Internet technologies and protocols to support all aspects of data communication with spacecraft (Rash et al., 2002b). The Operating Missions as Nodes on the Internet (OMNI) (NASA, Hogie et al., 2001) project at GSFC is not only involved in prototyping, but is also testing and evaluating various IP-based approaches and solutions for space communications. Other efforts in using Internet protocols for space communications have also been reported in the literature (Minden et al., 2002).

Some of the NASA-led projects on IP in space involve handoffs in space networks. Such projects include OMNI (Hallahan, 2002; NASA,), Communication and Navigation Demonstration on Shuttle (CANDOS) mission (Hogie, 2002), and the GPM project (Rash et al., 2002a). NASA has also been working with Cisco on developing a Mobile router (Leung et al., 2001). It is also anticipated that MIP will play a major role in various space related NASA projects such as Advanced Aeronautics Transportation Technology (AATT), Weather Information Communication (WINCOMM) and Small Aircraft Transportation Systems (SATS) (Leung et al., 2001). In this section, we will investigate the use of SIGMA in space networks to support IP mobility. First, the scenarios of network layer handoffs in satellite environment is identified. Then we introduce SIGMA-SN — the mapping of SIGMA in space network.

3.1 Handoffs in a Satellite Environment

LEO satellites have some important advantages over GEO satellites for implementing IP in space. These include lower propagation delay, lower power requirements both on satellite and user terminal, more efficient spectrum allocation due to frequency reuse between satellites and spotbeams. However, due to the non-geostationary nature and fast movement of LEO satellites, the mobility management in LEO is much more challenging than in GEO or MEO.

If one of the communicating endpoint (either satellite or user terminal) changes its IP address due to the movement of satellite or mobile user, a network layer handoff is required to migrate the connection of higher level protocol (e.g. TCP, UDP, or SCTP) to the new IP address. We describe below two scenarios requiring network layer handoff in a satellite environment.

1. *Satellite as a router* (Fig. 5): When a satellite does not have any on-board equipment which generates or consumes data, but is only equipped with on-board IP routing devices, the satellite acts as a router in the Internet. Hosts are handed over from one satellite to another as the hosts come under the footprint of different satellites due to the rotation of the LEO satellites around the Earth. Referring to Fig. 5, the Fixed Host/Mobile Host (FH/MH) needs to maintain a continuous transport layer connection with the correspondent node (CN) while their attachment points change from Satellite A to satellite B. Different satellites, or even different spot-beams within a satellite, can be assigned with different IP subnet addresses. In such a case, IP address change occurs during an inter-satellite handoff, thus requiring a network layer handoff. For highly dense service areas, a spot-beam handoff may also require a network layer handoff. Previous research (Nguyen et al., 2001; Sarikaya and Tasaki, 2001) have used Mobile IPv6 to support mobility management in LEO systems, where the FH/MH and Location Manager are mapped to Mobile IP's Mobile Node and Home Agent, respectively.
2. *Satellite as a mobile host* (Fig. 6): When a satellite has on-board equipment (such as earth and space observing equipment) which generates data for transmission to workstations on Earth, or the satellite receives control signals from the control center, the satellite acts as the endpoint of the communication, as shown in Fig. 6. Although the satellite's footprint moves from ground station A to B, the satellite should maintain continuous transport layer connection with its corespondent node (CN). A network layer handoff has to be performed if the IP address of the satellite needs to be changed due to the handover between ground stations.

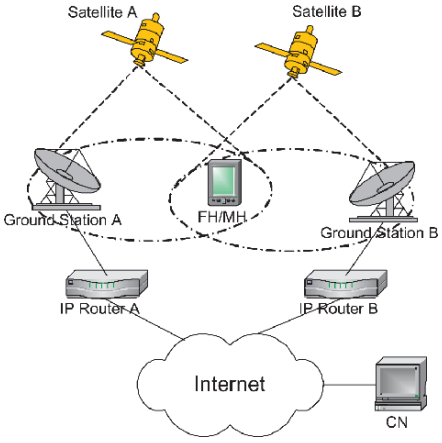

Figure 5: User handoff between satellites.

3.2 SIGMA-SN: Application of SIGMA in Space Networks

Having described our proposed SIGMA scheme and handoffs in space networks in Secs. 2 and 3.1, respectively, we describe below the mapping of SIGMA into a space handoff scenario, using satellites as examples of spacecrafts. We call this application and mapping of SIGMA to a space environment as SIGMA-SN.

1. *Satellite as a router:* Research results desribed in (Kwon and Sung, 2001) showed that the mean number of available satellites for a given FH/MH is at least two for latitudes less than 60 degrees. This means the FH/MH is within the footprint of two satellites most of the time, which makes SIGMA-SN very attractive for handoff management with a view to reducing packet loss and handoff latency. The procedure of applying SIGMA in this handoff scenario is straightforward; we just need to map the FH/MH and satellites in Fig. 5 to the MH and access routers, respectively, in the SIGMA scheme (see Fig. 2) as given below:
   - *Obtain new IP*: When FH/MH receives advertisement from Satellite B, it obtains a new IP address using either DHCP, DHCPv6, or IPv6 Stateless Address Autoconfiguration.
   - *Add new IP address to the association*: FH/MH binds the new IP address into the association (in addition to the IP address from Satellite A domain). FH/MH also notifies CN about the availability of the new IP address by sending an ASCONF chunk (Stewart et al., 2004) to the CN with the parameter type set as "Add IP Address".

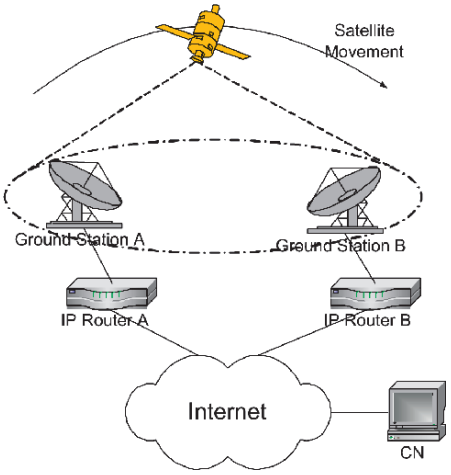

Figure 6: Satellite handoff between ground stations.

- *Redirect data packets to new IP address*: CN can redirect data traffic to the new IP address from Satellite B to increase the possibility of data being delivered successfully to the FH/MH. This task can be accomplished by sending an ASCONF chunk with the Set-Primary-Address parameter to CN, which results in CN setting its primary destination address to FH/MH as the new IP address.
- *Updating the Location manager*: `SIGMA-SN` supports location management by employing a location manager that maintains a database which records the correspondence between FH/MH's identity (such as domain name) and its current primary IP address. Once the Set-Primary-Address action is completed successfully, FH/MH updates the location manager's relevant entry with the new IP address. The purpose of this procedure is to ensure that after FH/MH moves from the footprint of Satellite A to that of Satellite B, further association setup requests can be routed to FH/MH's new IP address.
- *Delete or deactivate obsolete IP address*: When FH/MH moves out of the coverage of satellite A, FH/MH notifies CN that its IP address in Satellite A domain is no longer available for data transmission by sending an ASCONF chunk to CN with parameter type "Delete IP Address".

Due to the fixed movement track of the satellites in a space environment, FH/MH can predict the movement of Satellites A and B quite accurately. This a-priori information will be used to decide on the times to perform the set primary to the new IP address and delete the old IP address. This is much easier than in cellular networks, where the user mobility is hard to predict precisely.

2. *Satellite as a mobile host:* In this case, the satellite and IP Router A/B (see Fig. 6) will be mapped to the MH and access routers, respectively, of `SIGMA`. In order to apply `SIGMA-SN`, there is no special requirement on the Ground Stations A/B and IP routers A/B in Fig 6, which will ease the deployment of `SIGMA-SN` by not requiring any change to the current Internet infrastructure. Here, the procedure of applying `SIGMA-SN` is similar to the previous case (where the satellite acts as a router) if we replace the FH/MH by the satellite, in addition to replacing Satellites A/B by IP routers A/B.

   Since a satellite can predict its own movement track, it can contact Ground Station B while it is still connected to Ground Station A. There may be multiple new Ground Stations available to choose from due to the large footprint of satellites. The strategy for choosing a Ground Station can be influenced by several factors, such as highest signal strength, lowest traffic load, and longest remaining visibility period.

3.3 Vertical Handoff between Heterogeneous Technologies

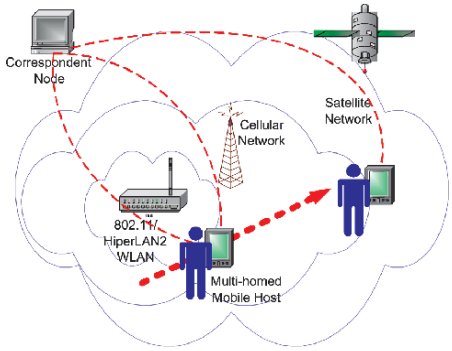

Figure 7: Vertical handoff using `SIGMA-SN`.

Different types of wireless access network technologies can be integrated to give mobile users a transparent view of the Internet. Handoff will no longer be limited to between two subnets in Wirless LAN (WLAN), or between two cells in a cellular network (horizontal handoff). In the future, mobile users will expect seamless handoff between heterogeneous

access networks (vertical handoff), such as WLANs and cellular networks.

MIP operates in Layer 3 and is independent of the underlying access network technology. Although it can be used for handoffs in a heterogeneous environment, there are a number of disadvantages in using MIP for vertical handoffs (Dixit, 2002). The disadvantages include complexity in routing, high signaling overhead, significant delay especially when CN is located in foreign network, difficulty in integrating QoS protocols such as RSVP with triangular routing and tunnelling.

SIGMA-SN is well suited to meeting the requirements of vertical handoff. Figure 7 illustrates the use of SIGMA-SN to perform vertical handoffs from WLAN to a cellular network, and then to a satellite network. A multi-homed mobile host in SIGMA-SN is equipped with multiple interface cards that can bind IP addresses obtained from different kinds of wireless network access technologies.

4 SURVIVABILITY COMPARISON OF SIGMA AND MIP

In this section we discuss the survivability of MIP and SIGMA. We highlight the disadvantages of MIP in terms of survivability, and then discuss how those issues are taken care of in SIGMA.

4.1 Survivability of MIP

In MIP, the location database of all the mobile nodes are distributed across all the HAs scattered at different locations (home networks). According to principles of distributed computing, this approach appears to have good survivability. However, there are two major drawbacks to this distributed nature of location management as given below:

- If we examine the actual distribution of the mobile users' location information in the system, we can see that each user's location and account information can only be accessible through its HA; these information are not truly distributed to increase the survivability of the system. The transparent replication of the HA, if not impossible, is not an easy task as it involves extra signaling support as proposed in (Lin and Arul, 2003).
- Even if we replicate HA to another agent, these HAs have to be located in the home network of an MH in order to intercept the packets sent to the MH. The complete home network could be located in a hostile environment, such as a battlefield, where the possibility of all HAs being destroyed is still relatively high. In the case of failure of the home networks, all the MHs belonging to the home network would no longer be accessible.

4.2 Centralized Location Management of SIGMA offers Higher Survivability

Referring to Fig. 4, SIGMA uses a centralized location management approach. As discussed in Sec. 2.1, the location management and data traffic forwarding functions in SIGMA are decoupled, allowing it to overcome many of the drawbacks of MIP in terms of survivability (see Sec. 4.1) as given below:

- The LM uses a structure which is similar to a DNS server, or can be directly combined with a DNS server. It is, therefore, easy to replicate the Location Manager of SIGMA at distributed secure locations to improve survivability.
- Only location updates/queries need to be directed to the LM. Data traffic do not need to be intercepted and forwarded by the LM to the MH. Thus, the LM does not have to be located in a specific network to intercept data packets destined to a particular MH. It is possible to avoid physically locating the LM in a hostile environment; it can be located in a secure environment, making it highly available in the network.

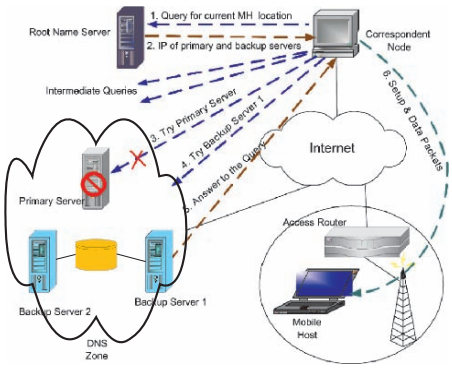

Figure 8: Survivability of SIGMA's location management.

Figure 8 illustrates the survivability of SIGMA's location management, implemented using DNS servers as location servers. Currently, there are 13 servers in the Internet (R. Bush et al., 2000) which constitute the root of the DNS name space hierarchy. There are also several delegated name servers in the DNS zone (Stevens, 1994), one of which is primary and

the others are for backup and they share a common location database. If an MH's domain name belongs to this DNS zone, the MH is managed by the name servers in that zone. When the CN wishes to establish a connection with the MH, it first sends a request to one of the root name servers, which will direct the CN to query the intermediate name servers in the hierarchy. At last, CN obtains the IP addresses of the name servers in the DNS zone to which the MH belongs. The CN then tries to contact the primary name server to obtain MH's current location. If the primary server is down, CN drops the previous request and retries backup name server 1, and so on. When a backup server replies with the MH's current location, the CN sends a connection setup message to MH. There is an important difference between the concept of MH's DNS zone in SIGMA and MH's home network in MIP. The former is a logical or soft boundary defined by domain names while the latter is a hard boundary determined by IP routing infrastructure.

If special software is installed in the primary/backup name servers to constitute a high-availability cluster, the location lookup latency can be further reduced. During normal operation, heart beat signals are exchanged within the cluster. When the primary name server goes down, a backup name server automatically takes over the IP address of the primary server. A query requests from a CN is thus transparently routed to the backup server without any need for retransmission of the request from the CN.

Other benefits SIGMA's centralized location management over MIP's location management can be summarized as follows:

- *Security*: Storing user location information in a central secure database is much more secure than being scattered over various Home Agents located at different sub-networks (in the case of Mobile IP).
- *Scalability*: Location servers do not intervene with data forwarding task, which helps in adapting to the growth in the number of mobile users gracefully.
- *Manageability*: Centralized location management provides a mechanism for an organization/service provider to control user accesses from a single server.

5 SECURITY OF SIGMA

In this section, we discuss the security issues of SIGMA and its interoperability with the current security mechanisms of the Internet.

5.1 Interoperability between MIP and Ingress Filtering

Ingress filtering is widely used in the Internet to prevent IP spoofing and Denial of Service (DoS) attacks. Ingress filtering is performed by border routers to enforce topologically correct source IP address. Topological correctness requires MH to use COA as the source IP address, since the COA is topologically consistent with the current network of the MH. On the other hand, TCP keeps track of its internal session states between communicating endpoints by using the IP address of the two endpoints and port numbers (Stevens, 1994). Therefore, applications built over TCP require the MH to always use its home address as its source address. The solution to this contradiction caused by combined requirements of user mobility, network security and transport protocols is *reverse tunnelling*, which works but lacks in terms of performance as illustrated below.

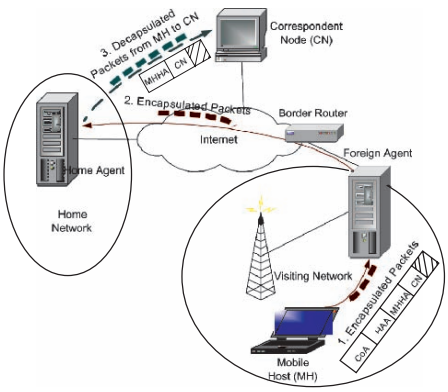

Figure 9: Interoperability between Mobile IP and Ingress Filtering.

Reverse tunnelling in MIP is shown in Fig. 9 which consists of the following components (Perkins, 2002):

1. *Encapsulation*: A data packet sent from the MH to the CN has two IP headers: the inner header has source IP address set to MH's home address (MHHA) and destination IP address set to CN's IP address; the outer header has its source IP address set to MH's CoA and destination IP address set to HA's IP address (HAA). Since the MH's CoA is topologically correct with the foreign network address, ingress filtering at foreign network's border routers allows these packets to pass through.
2. *Decapsulation*: The packets from the MH are routed towards the MH's HA because of the outer

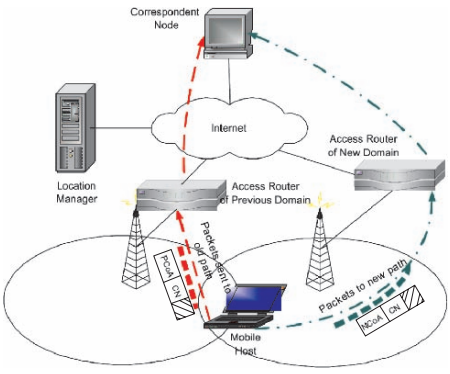

Figure 10: Interoperability between SIGMA and Ingress Filtering.

IP destination address. The HA decapsulates the packets, resulting in data packets with only one IP header (same as the previous inner header), which are then forwarded to their actual destination, i.e. the CN.

3. *Data Delivery*: When data packets arrive at the CN with the source and destination addresses being that of MH's home address and CN's address, respectively, they are identified by its TCP connection and delivered to the upper layer application.

Reverse tunnelling makes it possible for MIP to interoperate with Ingress filtering. However, the encapsulation and decapsulation of packets increase the end-to-end delay experienced by data packets, and also increase the load on the HA, which may become a performance bottleneck as the number of MHs increases.

5.2 Interoperability between SIGMA and Ingress Filtering

In SIGMA, the transport protocol uses IP diversity to handle multiple IP addresses bound to one association. The CN can thus receive IP packets from multiple source IP addresses belonging to an association, identify the association, and deliver the packets to the corresponding upper layer application. This improved capability of endpoint transport protocol permits smooth interoperability between SIGMA and Ingress Filtering.

As shown in Fig. 10, MH can use the CoA that belongs to the subnet which is responsible for sending data for the MH. In the new network, after the new CoA (NCoA) has been bound into the current association through ADDIP chunks (discussed in Sec. 2.1), the MH uses the NCoA to communicate directly with the CN. Since the NCoA is topologically correct with the subnet address, the border router of the foreign network allows packets with source IP set to the new CoA to pass. Thus, SIGMA does not require encapsulation and decapsulation as done in MIP. The transport protocol stack at the CN takes care of delivering packets coming from both previous CoA (PCoA) and NCoA to the upper layer application. SIGMA, therefore, interoperates well with ingress filtering without the need for reverse tunnelling.

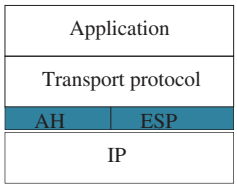

Figure 11: Use of IPSec with SCTP.

5.3 Enhancing the Security of SIGMA by IPSec

IPSec has been designed to provide an interoperable security architecture for IPv4 and IPv6. It is based on cryptography at the network layer, and provides security services at the IP layer by allowing endpoints to select the required security protocols, determine the algorithms to use, and exchange cryptographic keys required to provide the requested services. The IPSec protocol suite consists of two security protocols, namely Authentication Header (AH) and Encapsulating Security Payload (ESP). ESP provides data integrity, authentication, and secrecy services, while the AH is less complicated and thus only provides the first two services. The protocol stack, when IPSec is used with a transport protocol (SCTP in our case), is shown in Fig. 11.

SIGMA is based on dynamic address reconfiguration, which makes the association vulnerable to be hijacked, also called *traffic redirection attack*. An attacker claims that its IP address should be added into an established association between MH and CN, and further packets sent from CN should be directed to this IP address. Another kind of security risk is introduced by dynamic DNS update. An attacker can send a bogus location update to the location manager, resulting in all future association setup messages being sent to illegal IP addresses. The extra security risk introduced by SIGMA gives rise to the authentication problem: CN and LM need to determine whether the MH initiated the handover process. Since both AH and ESP support authentication, in general, we can

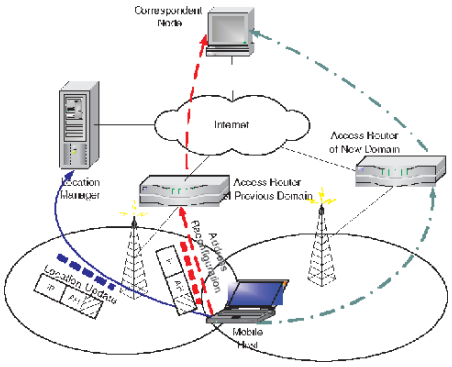

Figure 12: Interoperability between SIGMA and IPSec.

choose either of them for securing SIGMA. ESP has to be used if data confidentiality is required. Assume that we are only concerned with authentication of MH by CN and LM to prevent redirection attack and association hi-jacking. In this case, AH can be used as shown in Fig. 12. All address reconfiguration messages and location updates sent to CN and LM should be protected by IPSec AH header.

6 CONCLUSIONS

We have presented the architecture of Seamless IP diversity-based Generalized Mobility Architecture (SIGMA) to manage handovers of mobile nodes in the Internet architecture. We have shown the applicability of SIGMA to space networks for performing inter-satellite handovers, and presented the survivability and security of SIGMA. It has been shown that SIGMA has a higher survivability than MIP – thanks to its centralized location management scheme. SIGMA can also easily interoperate with existing network security infrastructures such as Ingress filtering and IPSec.

ACKNOWLEDGMENTS

We thank William Ivancic for the numerous discussion that greatly improved the quality of this paper.

REFERENCES

Atiquzzaman, M. and Ivancic, W. (2003). Evaluation of SCTP multistreaming over wireless/satellite links. In *12th International Conference on Computer Communications and Networks*, pages 591–594, Dallas, Texas.

Awerbuch, B. and Peleg, D. (1991). Concurrent online tracking of mobile users. In *ACM SIGCOMM Symposium on Communications, Architectures and Protocols*, pages 221–233.

Bhasin, K. and Hayden, J. L. (2002). Space Internet architectures and technologies for NASA enterprises. *International Journal of Satellite Communications*, 20(5):311–332.

Caire, G., Taricco, G., and Biglieri, E. (1998). Bit-interleaved coded modulation. *IEEE Transactions on Information Theory*, 44(3):927–946.

Cambell, A. T., Kim, S., and et al., J. G. (1999). Cellular IP. IETF DRAFT, draft-ietf-mobileip-cellularip-00.txt.

Dixit, S. (2002). Wireless IP and its challenges for the heterogeneous environment. *Wireless Personal Communications*, 22(2):261–273.

Fu, S. and Atiquzzaman, M. (2003). Improving end-to-end throughput of Mobile IP using SCTP. In *Workshop on High Performance Switching and Routing*, pages 171–176, Torino, Italy.

Fu, S. and Atiquzzaman, M. (2004). SCTP: State of the art in research, products, and technical challenges. *IEEE Communications Magazine*, 42(4):64–76.

Fu, S., Atiquzzaman, M., and Ivancic, W. (2002). Effect of delay spike on SCTP, TCP Reno, and Eifel in a wireless mobile environment. In *11th International Conference on Computer Communications and Networks*, pages 575–578, Miami, FL.

Fu, S., Atiquzzaman, M., and Ivancic, W. (2003). SCTP over satellite networks. In *IEEE 18th Annual Workshop on Computer Communications*, pages 112–116, Dana Point, California.

Fu, S., Atiquzzaman, M., and Ivancic, W. (2005). Evaluation of SCTP for space networks. *IEEE Wireless Communications*, 12(5):54–62.

Glossner, J., Iancu, D., Lu, J., Hokenek, E., and Moudgill, M. (2003). A software-defined communications baseband design. *IEEE Communications Magazine*, 41(1):120–128.

Gustafsson, E., Jonsson, A., and Perkins, C. (2001). Mobile IP regional registration. IETF DRAFT, draft-ietf-mobileip-reg-tunnel-04.txt.

Hallahan, F. (2002). Lessons learned from implementing Mobile IP. In *The Second Space Interent Workshop*, Greenbelt, MD.

Hogie, K. (2002). Demonstration of Internet technologies for space communication. In *The Second Space Internet Workshop*, Greenbelt, Maryland.

Hogie, K., Criscuolo, E., and Parise, R. (2001). Link and routing issues for Internet protocols in space. In *IEEE Aerospace Conference*, pages 2/963–2/976.

Holzbock, M. (2003). IP based user mobility in heterogeneous wireless satellite-terrestrial networks. *Wireless Personal Communications*, 24(2):219–232.

Hsieh, R. and Seneviratne, A. (2003). A comparison of mechanisms for improving Mobile IP handoff latency for end-to-end TCP. In *ACM MobiCom*, pages 29–41, San Diego, USA.

Johnson, D., Perkins, C., and Arkko, J. (2004). Mobility support in IPv6. IETF RFC 3775.

Jung, M., Park, J., Kim, D., Park, H., and Lee, J. (2002). Optimized handoff management method considering micro mobility in wireless access network. In *5th IEEE International Conference on High Speed Networks and Multimedia Communications*, pages 182–186.

Koh, S. J., Lee, M. J., Ma, M. L., and Tuexen, M. (2004). *Mobile SCTP for Transport Layer Mobility*. draft-sjkoh-sctp-mobility-03.txt.

Koodli, R. (2004). Fast handovers for Mobile IPv6. IETF DRAFT, draft-ietf-mipshop-fast-mipv6-03.txt.

Kwon, Y. and Sung, D. (2001). Analysis of handover characteristics in shadowed LEO satellite communication networks. *International Journal of Satellite Communications*, 19(6):581–600.

Leung, K., Shell, D., Ivancic, W., Stewart, D., Bell, T., and Kachmar, B. (2001). Application of Mobile-IP to space and aeronautical networks. *IEEE Aerospace and Electronic Systems Magazine*, 16(12):13–18.

Li, L. (2002). PKI based end-to-end mobility using SCTP. In *MobiCom 2002*, Atlanta, Georgia, USA.

Liao, W., Ke, C., and Lai, J. (2000). Reliable multicast with host mobility. In *IEEE Global Telecommunications Conference (GLOBECOM)*, pages 1692–1696.

Lin, J. and Arul, J. (2003). An efficient fault-tolerant approach for Mobile IP in wireless systems. *IEEE Transactions on Mobile Computing*, 2(3):207–220.

Malki, K. E. (2003). Low latency handoffs in Mobile IPv4. IETF DRAFT, draft-ietf-mobileip-lowlatency-handoffs-v4-07.txt.

Maltz, D. A. and Bhagwat, P. (1998). MSOCKS: An architecture for transport layer mobility. In *INFOCOM*, pages 1037–1045, San Francisco, USA.

Minden, G., Evans, J., Baliga, S., Rallapalli, S., and Searl, L. (2002). Routing in space based Internets. In *Earth Science Technology Conference*, Pasadena, CA.

Montenegro, G. and Gupta, V. (1998). Sun's SKIP firewall traversal for Mobile IP. IETF RFC 2356.

NASA. Omni: Operating missions as nodes on the internet. ipinspace.gsfc.nasa.gov.

Nguyen, H., Lepaja, S., Schuringa, J., and Vanas, H. (2001). Handover management in low earth orbit satellite IP networks. In *GlobeCom*, pages 2730–2734.

Perkins, C. (1998). Mobile Networking Through Mobile IP. *IEEE Internet Computing*, 2(1):58–69.

Perkins, C. and Wang, K. (1999). Optimized smooth handoffs in Mobile IP. In *IEEE International Symposium on Computers and Communications*, pages 340–346.

Perkins, C. E. (2002). IP Mobility Support. IETF RFC 3344.

Ramjee, R., Porta, T., and et al., S. T. (1999). IP micromobility support using HAWAII. IETF DRAFT, draft-ietf-mobileip-hawaii-00.txt.

Rappaport, T. S. (1996). *Wireless Communications Principles and Practice*. Prentice Hall, Upper Saddle River, NJ.

Rash, J., Casasanta, R., and Hogie, K. (2002a). Internet data delivery for future space missions. In *NASA Earth Science Technology Conference*, Pasadena, CA.

Rash, J., Criscuolo, E., Hogie, K., and Praise, R. (2002b). MDP: Reliable file transfer for space missions. In *NASA Earth Science Technology Conference*, Pasadena, CA.

Bush, R., Karrenberg, D., Kosters, M., and Plzak, R. (2000). Root name server operational requirements. IETF RFC 2870.

Sarikaya, B. and Tasaki, M. (2001). Supporting node mobility using mobile IPv6 in a LEO-satellite network. *International Journal of Satellite Communications*, 19(5):481–498.

Snoeren, A. C. and Balakrishnan, H. (2000). An end-to-end approach to host mobility. In *ACM MobiCom*, pages 155–166, Boston, MA.

Soliman, H., Catelluccia, C., and et al., K. M. (2004). Hierarchical Mobile IPv6 mobility management (HMIPv6). IETF DRAFT, draft-ietf-mipshop-hmipv6-04.txt.

Stevens, W. R. (1994). *TCP/IP Illustrated, Volume 1 (The Protocols)*. Addison Wesley.

Stewart, R., Ramalho, M., and et al., Q. X. (2004). Stream control transmission protocol (SCTP) dynamic address reconfiguration. IETF DRAFT, draft-ietf-tsvwg-addip-sctp-09.txt.

Thomson, S. and Narten, T. (1998). IPv6 stateless address autoconfiguration. IETF RFC 2462.

Wu, I., Chen, W., Liao, H., and Young, F. (2002). A seamless handoff approach of Mobile IP protocol for mobile wireless data networks. *IEEE Transactions on Consumer Electronics*, 48(2):335–344.

Xing, W., Karl, H., and Wolisz, A. (2002). M-SCTP: Design and prototypical implementation of an end-to-end mobility concept. In *5th Intl. Workshop on the Internet Challenge: Technology and Applications*, Berlin, Germany.

Ye, G., Saadawi, T., and Lee, M. (2002). SCTP congestion control performance in wireless multi-hop networks. In *MILCOM2002*, pages 934–939, Anaheim, California.