UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

SURVIVABILITY, SCALABILITY AND SECURITY OF MOBILITY PROTOCOLS

A DISSERTATION
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
DOCTOR OF PHILOSOPHY

By

MD SHOHRAB HOSSAIN
Norman, Oklahoma
2012

SURVIVABILITY, SCALABILITY AND SECURITY OF
MOBILITY PROTOCOLS

A DISSERTATION APPROVED FOR THE
SCHOOL OF COMPUTER SCIENCE

BY

Dr. Mohammed Atiquzzaman, Chair

Dr. Changwook Kim

Dr. John K. Antonio

Dr. Dean Hougen

Dr. Zahed Siddique

Dedication

This dissertation is dedicated to my wife.

Acknowledgements

This work has been possible because of a number of individuals. First of all, I would like to express my sincere gratitude to my faculty advisor Dr. Mohammed Atiquzzaman for his expert advice, patience and continuous supervision throughout my PhD studies, which made this work possible.

I am truly grateful to my committee members: Dr. John K. Antonio, Dr. Changwook Kim, Dr. Dean Hougen and Dr. Zahed Siddique for their valuable advice and time to review the dissertation. Their comments and suggestions have been very helpful to improve the quality of this dissertation.

I am also thankful to all the faculty and staff members of the School of Computer Science at the University of Oklahoma for providing a supportive environment for my study. Moreover, numerous discussions with my fellow researchers Dr. Abu Zafar Shahriar, Mr. Sazzadur Rahman, Syed Maruful Huq and Yasmin Jahir, and their technical expertise led to the improvements of this work.

The author would like to acknowledge the financial support of National Aeronautics and Space Administration (NASA) to carry out this project.

Last but not least, I would like to thank my wife Nusrat and my daughter Tasneem for accompanying and supporting me through all the tough times, and my parents and sister for providing moral support.

Table of Contents

Appendix A

Appendix B

List of Tables

List of Figures

Abstract

Today mobile computing has become a necessity and we are witnessing explosive growth in the number of mobile devices accessing the Internet. To facilitate continuous Internet connectivity for nodes and networks in motion, mobility protocols are required and they exchange various signaling messages with the mobility infrastructure for protocol operation. Proliferation in mobile computing has raised several research issues for the mobility protocols. First, it is essential to perform cost and scalability analysis of mobility protocols to find out their resource requirement to cope with future expansion. Secondly, mobility protocols have survivability issues and are vulnerable to security threats, since wireless communication media can be easily accessible to intruders. The third challenge in mobile computing is the protection of signaling messages against losses due to high bandwidth requirement of multimedia in mobile environments. However, there is lack of existing works that focus on the quantitative analysis of cost, scalability, survivability and security of mobility protocols.

In this dissertation, we have performed comprehensive evaluation of mobility protocols. We have presented tools and methodologies required for the cost, scalability, survivability and security analysis of mobility protocols. We have proposed a dynamic scheduling algorithm to protect mobility signaling message against losses due to increased multimedia traffic in mobile environments and have also proposed a mobile network architecture that aims at maximizing bandwidth utilization. The analysis presented in this work can help network engineers compare different mobility protocols quantitatively, thereby choose one that is reliable, secure, survivable and scalable.

Chapter 1

Introduction

Next generation networks are gradually converging towards the all-IP networks which can enable true global mobility and Internet connectivity to mobile devices. Today mobile computing has become a necessity, and we are witnessing the explosive growth in the number of mobile devices accessing the Internet. This has been possible due to the advance of wireless access technologies and miniaturization of mobile devices, driven by people's desire to get connected to one another anytime anywhere through online social networking or other means of communication.

1.1 Introduction

Internet Protocol (IP) is the underlying communication protocol that allows an end host to get connected to other hosts over the public Internet. However, IP cannot work in mobile environment since the IP address of a mobile device changes with the change of its location, and such a change of IP address causes the termination of an ongoing connection. Therefore, to facilitate continuous Internet connectivity for mobile nodes, Internet Engineering Task Force (IETF) proposed Mobile IPv6 [1], an IP-based mobility protocol.

Collection of IP-enabled devices, such as laptops, PDAs, IP-cameras or networks of sensors (deployed in vehicles) moving together in a bus, train, aircraft, ship can form mobile networks and the mobility management can be performed in an aggregated way, rather than by each mobile device of the mobile network. This aggregated mobility management can significantly reduce signaling requirement and power consumption. Only one of the components in a mobile network, therefore, needs to be equipped with high-power transceivers to manage mobility of all the nodes of the mobile network. IETF standardized NEtwork MObility (NEMO) [2] to provide mobility support to multiple IP-enabled devices in a mobile network.

However, Mobile IPv6 [1] and NEMO basic support protocol [2] suffer from the high latency, packet loss, inefficient routing resulting in degradation of the overall performance. To address the above mentioned drawbacks of Mobile IPv6 and NEMO, SIGMA [3] and SINEMO [4], Seamless IP-diversity based host and network mobility protocols, were proposed by the researchers at the Telecommunications and Networks Research Lab (TNRL) of the University of Oklahoma with collaboration of NASA. SIGMA [3] and SINEMO [4] are suitable for providing seamless mobility support for terrestrial and space networks.

1.2 Motivation and Problem Statement

In a mobile computing environment, a number of *network parameters* (such as network size, mobility rate, traffic rate) influence the signaling costs related to mobility management. With the rapid growth and popularity of mobile and wireless networks, increasingly larger number of IP-enabled mobile devices now require support from the *mobility management entities*, many of which were not initially designed to handle such enormous signaling load. This ever-increasing load on mobility management entities may result in the performance degradation of mobility protocols. Hence, mobility protocols are required to be analyzed with respect to the signaling overhead on its key entities to find out their resource requirement for their smooth operation.

Mobility protocols have their applications in buses, trains, aircrafts, military vans, satellites. The increased load on the critical (infra-structural) components of mobility protocols may sometimes lead to serious consequences, especially for critical applications, such as in-flight communications for air traffic signaling in a commercial aircraft or in a battlefield where continuous connectivity with peers is crucial. Therefore, it is essential to choose a mobility protocol that is scalable with respect to the future growth of networks and user demands. Thus, it is essential to analyze the scalability of mobility protocols and its key components with respect to various system parameters.

Another challenge in mobile computing is the protection of control messages from being ignored due to excessive amount of audio-video streaming data in mobile environment. Today more people are surfing the Internet and accessing multimedia data from their mobile devices. Due to such high data traffic, the access networks are often overloaded with packets of different priorities. The access routers have to act quickly enough to avoid loss of connection of mobile nodes' (or mobile networks') ongoing sessions. Therefore, it is important to have a dynamic scheduling algorithm in the access routers to facilitate such priority scheduling of the control messages, thereby improving the performance of mobility protocol.

Survivability of mobility protocols is another important issue. Survivability is a crucial aspect for any kind communication. A survivable network has the ability to withstand malicious attacks and to continue to work properly even in the presence of natural or man-made disturbances. Wireless and mobile networks have the challenge of survivability, since users are mobile and the communication channels are accessible to anyone. Mobility signaling message is the crucial control message for a mobility protocols and if these messages are intercepted or get lost, it may lead to complete outage of the system, leading to serious consequences for critical applications, such as air-traffic control system or in military battlefield. Hence, it is essential to analyze the survivability of mobility protocols and employ improved survivability measures for critical applications.

Finally, mobility protocols can be vulnerable to security threats. This is because the communication media is wireless (therefore, easily accessible by attackers) and the mobility protocols require data to be delivered to a node which are moving and can have multiple point-of-attachments. These special features in mobile computing have introduced several security issues. Any malicious agent can try to fool mobility agents by sending spoofed control messages and redirecting traffic away from the victim nodes, hijacking ongoing sessions, or even modifying the contents if proper protection mechanisms are not enforced. Therefore, it is very important to identify possible security threats for mobility protocols and analyze defense mechanisms to prevent or mitigate these threats.

1.3 Objectives of this Research

The *objectives* of this research are as follows:

- The first objective of this research is to perform a comprehensive cost and scalability evaluation of the host and network mobility protocols. The idea is to investigate how much signaling overhead is imposed on different components of the network that are responsible to ensure IP-connectivity to mobile devices. Moreover, we intend to investigate their scalability as mobile computing is proliferating.

- The second objective of this research is the quantitative evaluation of survivability of the mobility infrastructure and the associated components. We want to improve survivability of mobility protocols by protecting the crucial mobility signaling packets over bulk real-time data, thereby ensuring improved availability of the mobile device and the mobile network. To achieve this goal, we have proposed multi-homed NEMO architecture and performed survivability analysis both through simulations and experimental test-bed.

- The third objective of this research is to protect control or signaling messages of mobility protocols from being ignored as a consequence of excessive amount of audio-video streaming data in mobile Internet. We intend to propose a dynamic scheduling algorithm for wireless access networks to facilitate priority scheduling of the control messages, thereby ensuring their faster response for mobility protocols.

- The fourth objective of this research is to protect mobility protocols from security threats. Since mobile devices are connected through the wireless networks which are more vulnerable to attacks by malicious agents, we focus on identifying potential attack scenarios and propose protective mechanism to prevent or mitigate such attacks on mobility protocols.

- Finally, mobility protocols require a realistic mobility model that can mimic the movement pattern of nodes in motion. Our fifth objective is to derive all the stochastic properties of a realistic mobility model for nodes roaming in vehicles around the city.

1.4 Contributions of the Dissertation

The *contributions* of the dissertation are summarized as follows:

- Perform entity-wise cost evaluation of host and network mobility protocols.

- Perform quantitative scalability analysis of host and network mobility protocols.

- Perform multi-class queuing analysis and propose a dynamic scheduling algorithm to protect crucial control messages (of mobility management) against losses.

- Propose a multi-band mobile router architecture for NEMO to ensure maximum utilizations of available bandwidth.

- Perform survivability analysis of mobility protocols and propose a seamless NEMO architecture with improved performance.

- Identify security threats for mobility protocols, critically analyze defense mechanisms, and propose an effective security measure for SIGMA.

- Develop a mathematical model to derive various stochastic properties of a realistic mobility model for mobile networks (onboard a vehicle).

1.5 Organization of the Dissertation

The rest of the dissertation is organized as follows. Chapter 2 presents a review of host and network mobility protocols. Chapter 3 presents entity-wise cost analysis of different mobility protocols, followed by their scalability analysis in Chapter 4. In Chapter 5, we perform multi-class queuing analysis to measure various performance metrics for different classes of data and signaling traffic in the access router. In Chapter 6, we propose a multi-band mobile router architecture for NEMO to ensure maximum possible utilizations of the available bandwidth. Chapter 7 presents survivability analysis for NEMO, followed by the experimental evaluation of a proposed multi-homed NEMO architecture in Chapter 8. In Chapter 9, we explain with illustrative examples all the major security threats and critically analyze existing defense mechanisms for IP-mobility protocols, along with security solutions for SIGMA protocol. In Chapter 10, we present a mathematical model to derive various stochastic properties of a realistic mobility model for the mobile network (onboard a vehicle) and apply the model in mobility protocol analysis. Finally, Chapter 11 has the concluding remarks.

Chapter 2

Literature Review

Mobility protocols are required to facilitates Internet connectivity to nodes or networks in motion. In this chapter, a review of different host and network mobility protocols is provided to help readers better understand the rest of the dissertation.

2.1 Mobility Protocols

Next generation networks are gradually converging towards the all-IP networks which can enable true global mobility and Internet connectivity to mobile devices. Though IP is the underlying communication protocol to get connected to the public Internet, it cannot work in mobile environment. This is because the IP address of a mobile device changes with the change of its location, and such a change of IP address causes the termination of an ongoing connection. Therefore, to facilitate continuous Internet connectivity for mobile nodes, IETF proposed Mobile IPv6 [1], an IP-based mobility protocol to maintain session continuity during handover.

Collection of IP-enabled devices, such as, laptops, PDAs, IP-cameras moving together in a bus, train, aircraft, ship can form mobile networks and the mobility management can be performed in an aggregated way, rather than by each mobile device. This aggregated mobility management can significantly reduce signaling requirement and power consumption. Only one of the components in a mobile network can, therefore, be equipped with high-power transceivers to manage mobility of all the nodes of the mobile network. IETF standardized NEtwork MObility (NEMO) [2] to provide mobility support to multiple IP-enabled devices in a mobile network.

The rest of the chapter is organized as follows. In Section 2.2, we explain different host mobility protocols. In Section 2.3, we explain the NEMO architecture and review two network mobility protocols. Finally, we conclude in Section 2.4.

2.2 Host Mobility Protocols

The objective of this type of mobility protocols is to provide mobility support to a single mobile node. In this subsection, we explain briefly four host mobility protocols: Mobile IPv6, Hierarchical MIPv6 (IETF standard protocols), and SIGMA and Hierarchical SIGMA (proposed by the TNRL lab).

2.2.1 Mobile IPv6

IETF proposed Mobile IPv6 [1] which aims at solving two problems at the same time. First, it allows transport layer sessions (TCP or UDP) to continue even if the underlying hosts are roaming and changing their IP addresses. Second, it allows a host to be reached through a static IP address (home address).

The architecture of Mobile IPv6 is shown in Fig. 2.1. Each Mobile Host (MH) has a home network where it is registered with a router called the Home Agent (HA). Each MH has a static Home Address (HoA) by which it is identified, regardless of its current point of attachment to the Internet.

While away from its home, an MH is also associated with a Care-of Address (CoA), which provides its current location information. Whenever an MH acquires a new CoA from a foreign network, it must inform its HA about this new CoA through a Binding Update (BU). The HA accepts the BU and updates it binding cache (a table maintained by the HA for all its MHs), and sends a binding acknowledgement.

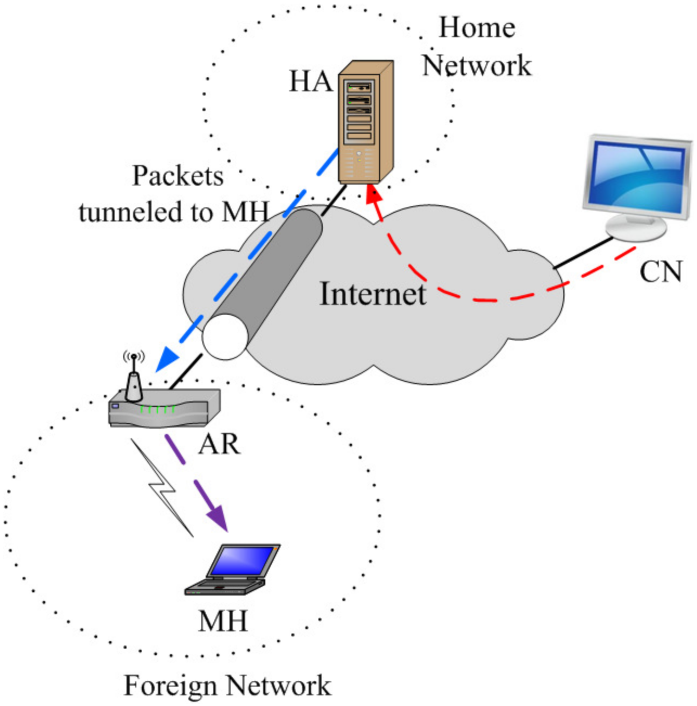


Figure 2.1: Mobile IPv6 architecture.

In Mobile IPv6, every data packet destined to an MH's HoA is intercepted by the HA. The HA checks its binding cache to obtain the MH's CoA and forwards the data packet to the new location using IPv6 encapsulation. Data packets from the Correspondent Node (CN) follows an un-optimized route to the MH (CN –> HA –> MH) (as shown in Fig. 2.1) instead of direct route (CN –> MH). This degrades the performance of Mobile IPv6, introducing large handover delay and packet loss.

2.2.2 Hierarchical Mobile IPv6

For high mobility of nodes, location information are sent very frequently to the HA, resulting in bandwidth wastage and possible network congestion. Moreover, the CN might get stale information about the MHs current location if the HA is far away from the MH, resulting in possible connection termination. Therefore, it is very important to reduce the delay of time-critical handover process of the MH to

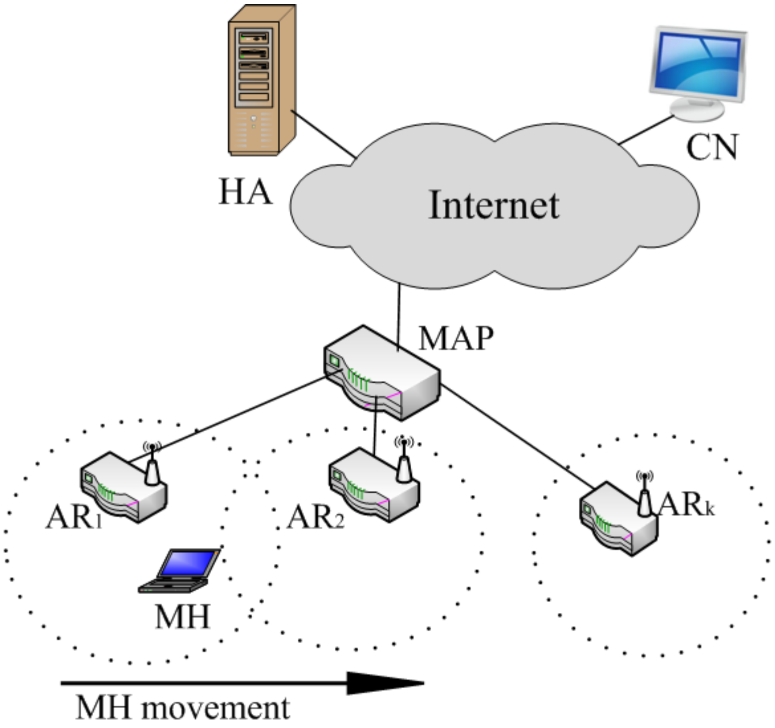


Figure 2.2: Hierarchical Mobile IPv6 architecture.

improve its performance. Hierarchical Mobile IPv6 (HMIPv6) [5], an enhancement of MIPv6 [1], aims at reducing handover delay by introducing a new network element, called Mobility Anchor Point (MAP) in addition to the HA.

HMIPv6 is designed to reduce the signaling cost of the base MIPv6 and it has the lowest signaling cost in all versions of MIPv6 enhancements. The architecture of HMIPv6 is shown in Fig. 2.2. Each MAP is essentially a local HA and covers several subnets under its domain. Upon arrival in a new MAP-domain, an MH discovers MAP's global address that is stored in the Access Routers (AR) and communicated to the MH via router advertisements. The MH then updates the HA with an address assigned by the MAP, called Regional Care-of-Address (RCoA), as its current location. The MAP intercepts all packets sent to the MH, encapsulates, and forwards them to the MH's current address. While moving within a MAP-domain, the MH

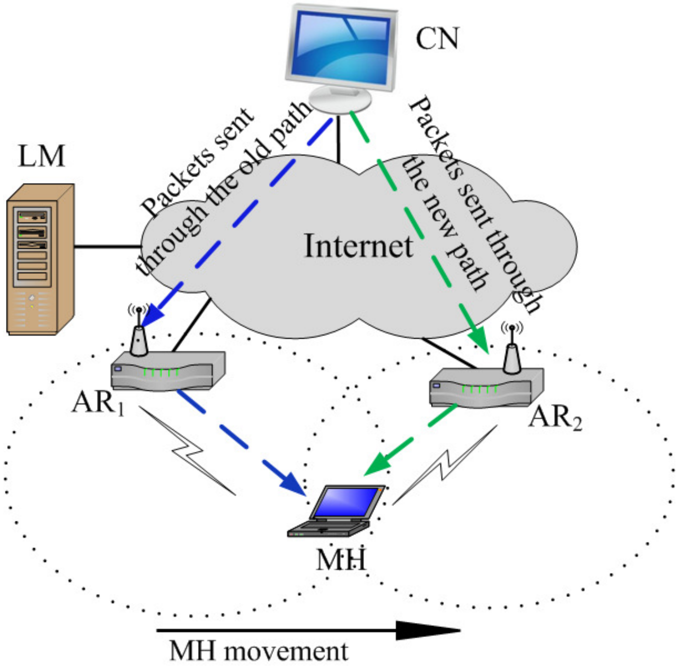

CN
LM
Packets sent through the old path
Packets sent through the new path
Internet
AR$_1$
AR$_2$
MH
MH movement

Figure 2.3: SIGMA architecture.

is not required to send updates to the HA, rather it uses the Local Care-of-Address (LCoA) within a MAP-domain. Thus, HMIPv6 reduces signaling requirement of the MH by introducing a local mobility agent (MAP).

2.2.3 SIGMA

SIGMA (Seamless IP-diversity based Generalized Mobility Architecture) [3] was designed to resolve the major drawbacks of Mobile IP (which has a large handover latency and packet loss). SIGMA exploits IP-diversity (i.e., having multiple network interfaces) of the MH to achieve seamless handover.

The architecture of SIGMA is shown in Fig. 2.3. The Location Manager (LM) is responsible for keeping up-to-date location information of the mobile hosts. Whenever any CN wants to send data to a MH, it must first send a query message to the LM to obtain its current IP address.

The main difference of SIGMA with MIPv6 protocol is the decoupling of data communication from location management. Unlike MIPv6, data packets between CN and the MH are not routed through the LM in SIGMA. After getting IP address from the LM, the CN sends data packets directly to the MH. In every handover, the MH must update its new IP address (after acquiring it from the new access network) to the LM. At the same time, the MH needs to send binding update to the communicating CN so that any subsequent packets of the ongoing session between MH and CN can be sent directly to the new address of the MH after the handoff. Thus, the LM is not responsible for data traffic to be routed to the MH from the CN, resulting in less overhead on the LM as well as the the complete system.

SIGMA achieves seamless handover by using the multi-homing feature of Stream Control Transport Protocol (STCP) [6] and uses *make-before-break strategy.* This means that the MH in SIGMA protocol obtains a secondary IP address from the new Access Router (AR_2) when it approaches to the overlapping area of two access routers (see Fig. 2.3). As the result of having two IP addresses (multi-homing approach) in the SCTP association, the MH can continue with the data transmission without any disruption, even though one of the access network (AR_1) becomes unreachable just after the handover. This is making a connection with the new AR_2 before breaking the existing connection with the old AR_1. Thus, the use of IP-diversity and make-before-break strategy results in the least handover delay and packet loss for SIGMA protocol [7].

2.2.4 Hierarchical SIGMA

Hierarchical SIGMA (HiSIGMA) [8] is an extension of SIGMA that promotes localized mobility approach. In SIGMA, the LM might be far away from the MH which may increases the delay between sending binding update and receiving binding acknowledgements. Hierarchical approach can relax the requirement of sending every binding updates to the LM (which might be located at a distant place).

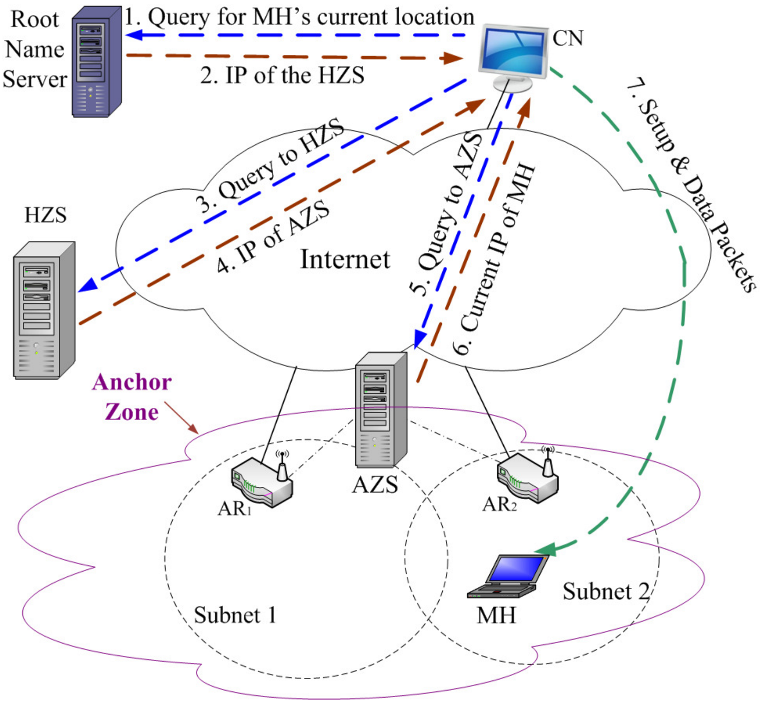


Figure 2.4: Hierarchical SIGMA.

Fig. 2.4 shows the Hierarchical SIGMA architecture where location management is done in multiple levels. A local location manager, namely Anchor Zone Server (AZS) is introduced in the hierarchy to keep records of movement within an AZS-domain which covers several subnets. This entity is similar to the MAP in HMIPv6 architecture. However, the main difference between this new mobility entity, AZS with HMIPv6's MAP is that AZS does not deal with any data packet destined to the MH since HiSIGMA decouples location management from data transmission. Similar is the case for the Home Zone Server (HZS) which deals with macro-mobility of MHs.

When the MH moves within an AZS-domain, there is no need to send signaling traffic outside of the AZS-domain, thereby reducing signaling requirement with the external networks over the bandwidth-limited air interface. For movement across AZS-domains, BU is sent to the HZS to keep record of the new AZS where the MH has recently moved. This hierarchical approach of location management reduces the handover latency and packet loss while improving the accuracy of the location management [8].

It can be noted that DNS can be used as the location manager of SIGMA [9] where domain name is used for the identification of MH. Therefore, location query (for the MH) in HiSIGMA can be processed in an approach similar to hierarchical DNS lookup. When CN wants to set up a new association with MH, CN sends a query message (with MH's domain name) to the root name server which replies with HZS's IP address. The CN then queries the HZS which replies with the IP address of current AZS where MH resides. Finally, the CN queries the AZS which replies with the current IP address(es) of MH.

2.3 Network Mobility Protocols

In this section, we briefly explain two network mobility protocols: NEMO (an IETF standard protocol) and SINEMO (TNRL proposed protocol).

2.3.1 NEMO architecture

Mobile networks can be formed with IP-enabled devices including laptops, PDAs, IP-cameras or networks of sensors deployed in vehicles, such as aircrafts, buses, and trains, etc. Fig. 2.5 shows the architecture of a mobile network. Mobile Router (MR) acts as the gateway for the nodes inside the mobile network, each of which is called a Mobile Network Node (MNN). Different types of MNNs are: Local Fixed Nodes (LFN) that do not move with respect to MN, Local Mobile Nodes (LMN) that usually reside in MN and can move to other networks, and Visiting Mobile Nodes (VMN) that get attached to the MN from another network. LMNs and VMNs are MIPv6 capable. The MR attaches to the Internet through ARs. A mobile network is usually connected to a network called the home network where an MR is registered

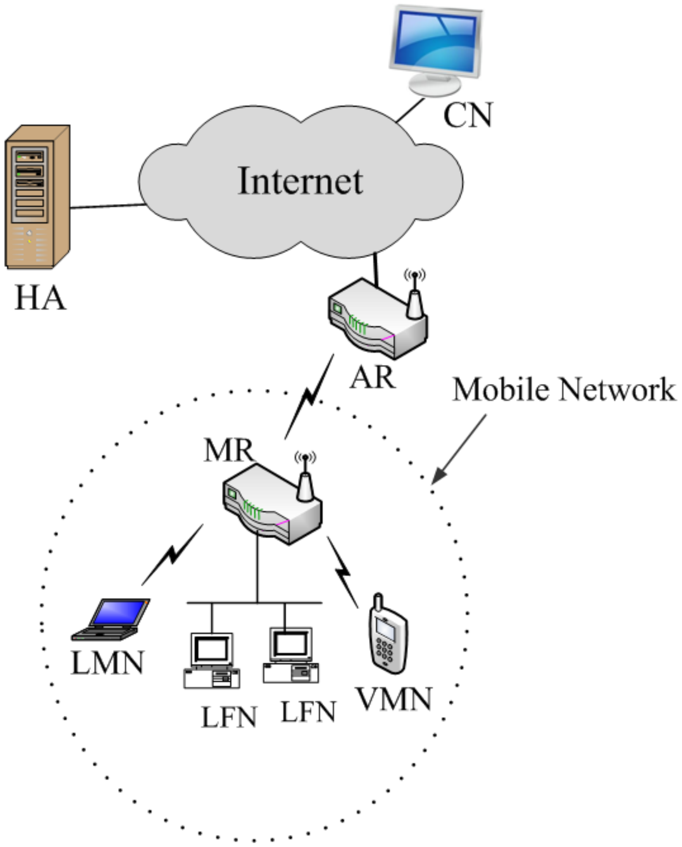


Figure 2.5: NEMO architecture.

with a router called the Home Agent. The HA is notified of the location of the MR and the HA redirects packets sent by the CN to MNNs.

2.3.2 NEMO BSP

IETF proposed NEtwork MObility Basic Support Protocol (NEMO BSP) [2] to facilitate continuous Internet connectivity of hosts moving together. In NEMO BSP [2], the MR ensures connectivity of all hosts inside the mobile network when the MR changes its point of attachment to the Internet while moving from a home network to a foreign network. An MR has its unique IP address and one or more Mobile

Network Prefixes (MNP) that it advertises to the hosts attached to it. MR establishes a bidirectional tunnel with the HA of MHs to pass all the traffic between its MHs and the CNs. When MR changes its point of attachment, it acquires a new care-of-address from the visited foreign network. It then sends a binding update to its HA which creates a cache entry (a mapping of the MRs home address to its care-of-address) and creates a bidirectional tunnel between HA and MR. When a CN sends a packet to a host, the packet is routed to the HA of the corresponding MR. HA looks at its cache entry and forwards the packet to the MR using the bidirectional tunnel. Finally, MR receives the packet, decapsulates it, and forwards it to the host inside the MN.

2.3.3 SINEMO architecture

Fig. 2.6 shows the architecture of SINEMO [4]. SINEMO is based on the concept of SIGMA and exploits IP-diversity feature of the MR. This implies that the mobile network consists of a multi-homed Mobile Router which can be connected to two wireless networks exploiting IP-diversity. The MR acts as a gateway between the hosts and the ARs for Internet access. A CN sends traffic to an MNN. A Central Location Manager (CLM) maintains the IP addresses of MR in a mobile network. A Local Location Manager (LLM), usually co-located with the MR, is used to keep the IP addresses of the hosts inside the MN. When an MN moves into one subnet, MR obtains its own public IP address and one or more address prefixes. MR provides and reserves an IP address for each host which only uses private addresses for connectivity.

In SINEMO, the hosts are not aware of their public IP addresses; they use only the private IP addresses for connectivity. After handover, only the public addresses are modified at MR; the private IP addresses of the hosts remain unchanged. MR thus hides mobility from the hosts. The readers can refer to [4] for more details of SINEMO handover and location management.

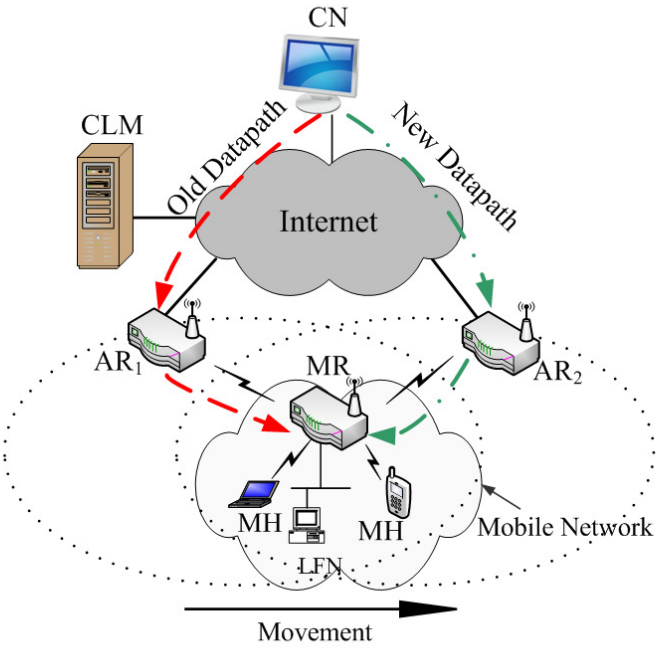


Figure 2.6: SINEMO architecture.

2.4 Summary

Host and network mobility protocols are required to provide Internet connectivity to nodes and networks in motion. In this chapter, we have reviewed the IETF standard host and network mobility protocols: MIPv6, HMIPv6 and NEMO basic protocol. We have also explained the architecture and protocol operations of SIGMA and SINEMO, two seamless IP-diversity-based host and network mobility protocols, respectively, proposed in the TNRL lab. In the next chapter, we perform cost analysis of these mobility protocols to measure the amount of resources needed for each mobility management entities for protocol operation.

Chapter 3

Cost Analysis of Mobility Protocols

Host and network mobility protocols are required to provide Internet connectivity to nodes and networks in motion. With the rapid growth and popularity of mobile and wireless networks, increasingly larger number of IP-enabled mobile devices now require mobility support. However, mobility protocols were not initially designed to handle such enormous signaling load. This ever-increasing load on mobility infrastructure may result in its performance degradation. In this chapter, we perform a comprehensive cost evaluation of the host and network mobility protocols to measure the amount of resources needed for their operations. This work can help us in choosing a mobility protocol that incurs the least cost on the mobility infrastructure.

3.1 Introduction

Mobility management protocols are used to facilitate delivery of data packets to hosts and networks that are in motion. IETF proposed Mobile IPv6 [1] and Hierarchical Mobile IPv6 (HMIPv6) [5] to support host-mobility and NEtwork MObility Basic Support Protocol (NEMO BSP) [2] for networks in motion. But these protocols have high handover latency, packet loss, and inefficient routing path, giving rise to deployment issues. To address these drawbacks, SIGMA [3] and SINEMO [4], Seamless IP-diversity-based host and network mobility protocols were proposed with reduced handover delay and packet loss.

In a mobile computing environment, a number of network parameters (such as network size, mobility rate, traffic rate) influence the costs relating to mobility. These include costs incurred in updating location manager about the change of location, sending updates to hosts with ongoing communication, and processing and lookup costs by various mobility agents, etc. With the rapid growth and popularity of mobile and wireless networks, increasingly larger number of IP-enabled mobile devices now require support from the mobility management entities (e.g., location manager, home agent, and mobile router, etc.), many of which were not initially designed to handle such enormous load. This ever-increasing load on mobility management entities may result in the performance degradation of mobility protocols. Hence, mobility protocols are required to be analyzed with respect to the overheads (on various mobility management entities) to measure the amount of resources required for its operations. This can help in choosing a mobility protocol that incurs the least cost on the mobility infrastructure.

There have been earlier attempts for cost analysis [10–15] of mobility protocols. Xie et al. [16] performed cost analysis of Mobile IP to minimize the signaling cost while introducing a regional location management scheme. Fu et al. [10] analyzed the signaling costs of SIGMA and HMIPv6. Reaz et al. [11] performed the signaling cost analysis of SINEMO but did not consider all possible costs of all the entities. Makaya et al. [12] presented an analytical model for the performance and cost analysis of IPv6-based mobility protocols. Munasinghe et al. [13] presented an analytical signaling cost model for vertical handoffs in a heterogeneous mobile networking environment. Lee et al. [14] analyzed the performance of Proxy Mobile IPv6 in terms of signaling cost and packet delivery cost. Xie et al. [15] analyzed various handoff scenarios for a dual stack mobile node roaming in a mixed IPv4/IPv6 environment. However, the above studies [10–15] did not consider all possible costs for mobility management, e.g., costs related to query messages by CN, refreshing binding updates, messages required to secure ongoing sessions, etc. Moreover, they did not compute the costs for various mobility management entities of the protocol. Hence, those analyses are incomplete.

The objective of the work presented in this chapter is to perform a comprehensive cost analysis of mobility entities of SIGMA, HiSIGMA and SINEMO, compare them with IETF standard protocols (HMIPv6 and NEMO BSP, respectively). We have chosen HMIPv6 in our analysis (to compare it with our lab-proposed SIGMA and HiSIGMA protocols) because HMIPv6 is designed to reduce the signaling cost of the base MIPv6 and it has the lowest signaling cost in all versions of MIPv6 enhancements. On the other hand, NEMO was chosen since it is the IETF-proposed base network mobility protocol [2].

The contributions of the work presented in this chapter are: (i) developing mathematical models to estimate total costs and efficiencies of various mobility management entities of HMIPv6, SIGMA, HiSIGMA, NEMO and SINEMO, and (ii) analyzing the impact of network size, mobility rate, traffic rate, and data volume on these costs and percentage overhead on the mobility entities.

The analytical cost models developed in this chapter covers all possible costs required for mobility management and will help in estimating the actual resources (bandwidth, processing power, and transmission power) required by key entities of the network. We have performed entity-based scalability analysis for the host and network mobility protocols since these entities are the key components and are subject to resource limitations in a mobility environment.

Our results show that SIGMA, HiSIGMA and SINEMO incur much lower overhead on their key entities and yield higher efficiency than HMIPv6 and NEMO, irrespective of session lengths, network size and mobility rate. This is because unlike HMIPv6 (and NEMO), SIGMA and SINEMO use an optimal route in data delivery between the mobile node and its correspondent node.

The rest of this chapter is organized as follows. First, the assumptions, notations and traffic model are listed in Section 3.2. Analytical cost models for HMIPv6, SIGMA, HiSIGMA, NEMO and SINEMO are presented in Sections 3.3, 3.4, 3.5, 3.6, and 3.7, respectively. In Section 3.8, we define two novel performance metrics for the mobility protocols and mobility management entities. In Section 3.9, we present numerical results. Section 3.10 has the concluding remarks.

3.2 Modeling Preparation

The assumptions, notations and arrival traffic model are explained in the following subsections.

3.2.1 Assumptions

Following are assumptions for cost analysis:

- Session arrival rate for each mobile host is equal.
- Each session length (data file size) is equal.
- Costs relating to standard IP switching are ignored.
- Uniform distribution of mobile hosts over the region of the network is assumed.
- Binary search is used to search location database.

The above assumptions are similar to the assumptions made by previous works [14, 15] and they make the model analytically tractable.

3.2.2 Notations

The notations used in this chapter are listed in this section.

N_m Number of Mobile Hosts,

N_c Average number of CNs per MH,

N_f Number of LFNs in the mobile network,

δ_Q Per hop transmission cost for query message,

δ_R Per hop transmission cost for registration message,

δ_L Per hop transmission cost for location update,

δ_B Per hop transmission cost for Binding Update,

δ_{AL} Per hop transmission cost for aggregated location update message,

δ_D Per hop transmission cost for DHCP request/reply message,

δ_{DP} Per hop transmission cost for average data packet,

δ_{DA} Per hop transmission cost for data Ack packet,

δ_{RR} Per hop transmission cost for Return Routability (RR) message,

h_p Average number of hops between Internet to arbitrary CN or CLM or AR,

h_{in} Average number of hops in the Internet,

σ Proportionality constant of wireless link over wired link,

ψ Linear coefficient for lookup cost,

T_r Subnet residence time,

T_{lf} Binding entry lifetime,

λ_s Average session arrival rate for each mobile host,

κ Maximum transmission unit,

α Average session length (in file size),

ξ Encapsulation cost,

β_P Processing cost at an entity P,

δ_{TH} Transmission cost for extra IP header used in tunneling,

k Number of access routers under a MAP or AZS,

m Number of MAPs (HMIPv6) or AZSs (HiSIGMA), where $m = xy/k$,

Υ Cost term for HMIPv6,

Γ Cost term for SIGMA,

Φ Cost term for HiSIGMA,

Λ Cost term for NEMO,

Δ Cost term for SINEMO.

3.2.3 Traffic Model

We assume that session arrival follows Poisson process [14, 15] with the following probability distribution function:

$$f_{sa}(n) = \frac{e^{-\lambda_s} \lambda_s^n}{n!}. \tag{3.1}$$

In other words, the inter-arrival times are exponentially distributed. The session length process that denotes size of data (file) in each session follows Pareto distribution. The mean session length is assumed to be α.

3.3 Cost Model for HMIPv6

In this section, we have performed entity-wise cost evaluations for different key mobility entities of HMIPv6: Home Agent, MAP, MH, and the complete system.

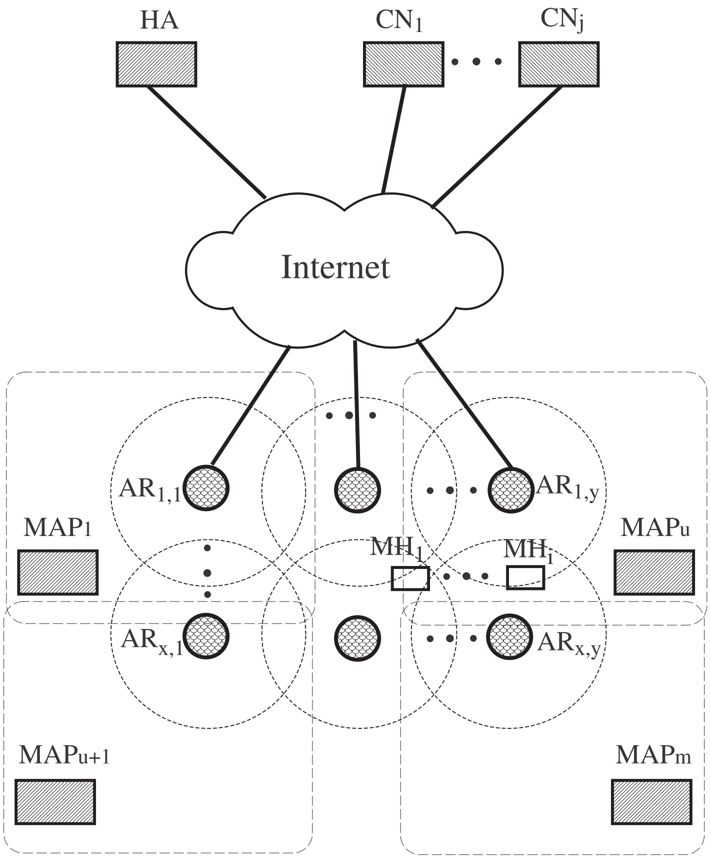


Figure 3.1: Hierarchical network structure for HMIPv6.

The cost terms for HMIPv6 protocol is influenced by the number of regional registration that happens in every move out of a MAP region. Let us first find out the expected number of moves causing an MH regional registration. In the topology (shown in Fig. 3.1), there are xy ARs in the foreign network. The MH can move from the coverage area of one AR to any other in one move. As each MAP covers k ARs, the probability that the mobile host will be within the coverage area of the previous MAP after a movement is $p = \frac{k}{xy}$. Conversely, the probability that MH will reach a new AR is $q = 1 - p = \frac{xy-k}{xy}$. So the probability that the MH moves out of a MAP domain in ith movement is $P_i = p^{i-1}q$. Hence, the expected number of moves for a MAP domain move-out can be obtained as follows:

$$\begin{aligned} M &= \sum_{i=1}^{\infty} iP_i = q(1 + 2p + 3p^2 + 4p^3 + ...) \\ &= \frac{1}{1-p} = \frac{xy}{xy-k}. \end{aligned} \tag{3.2}$$

We now derive the expressions for the total cost on the MAP, the HA, and the complete network for HMIPv6. Each MH communicates with a number of CNs and each such active communication period is termed as a session.

3.3.1 Mobility Anchor Point

The total cost on the MAP is due to exchange of RCoA and LCoA registration messages, RR messages and tunneling of packets from HA to MH, and vice versa.

LCoA registration messages: Every subnet crossing by the MH (that happens every T_r sec) within a MAP region, triggers an on-link CoA (LCoA) registration message to the MAP. This involves transmission cost of $2\delta_R$ and the processing cost of β_{MAP} for each MH. Hence, the cost incurred at MAP is as follows:

$$\Upsilon_{MAP}^{LC} = \frac{N_m}{m} \times \frac{2\delta_R + \beta_{MAP}}{T_r}. \tag{3.3}$$

Return routability messages: In order to ensure that binding update message is authentic and is not originated from malicious MH, RR procedure is performed before each BU. This process makes use of four messages: Home Test Init (HoTI), Home Test (HoT), Care-of Test Init (CoTI) and Care-of Test (CoT) [17]. Therefore, the cost incurred at the MAP for RR messages is as follows:

$$\Upsilon_{MAP}^{RR} = \frac{N_m}{m} \times \frac{4\delta_{RR} N_c}{MT_r}. \tag{3.4}$$

RCoA registration messages: The MAP receives registration requests from every MH entering the MAP domain. Since there are m MAPs and the MHs are uniformly distributed, there will be N_m/m MHs under an MAP on the average. The MAP processes the request and assigns an RCoA to the MH. This involves the transmission cost of $2\delta_R$ and the processing cost of β_{MAP} for each MH at an MAP. Each MH sends such RCoA registration requests in every MT_r seconds.

$$\Upsilon_{MAP}^{RC} = \frac{N_m}{m} \times \frac{2\delta_R + \beta_{MAP}}{MT_r}. \tag{3.5}$$

Data delivery cost: MAP acts as a local HA for the MH, receives all packets on behalf of the MH from the HA, decapsulates the packet, and then encapsulates it to forward it to MH's current location using the translation table of RCoA to LCoA. Thus, for every packet sent from CN to the MH, transmission and processing costs are incurred at the MAP. As the average session length is α, and maximum transmission unit is κ, there will be $\lceil \frac{\alpha}{\kappa} \rceil$ number of packets, and the packet rate can be obtained by $\lceil \frac{\alpha}{\kappa} \rceil \times \lambda_s$. The transmission cost for each packet is $(\delta_{DP} + \delta_{DA})$ due to the data packet and corresponding acknowledgement. As we have assumed a uniform distribution of MHs in the network, the number of MHs under an MAP-domain is $\frac{N_m k}{xy}$. The cost for IP routing table (with entries of k ARs) lookup is proportional to $\log_2 k$. Thus, the packet tunneling cost at MAP is given by

$$\Upsilon_{MAP}^{DD} = \frac{N_m}{m} \times N_c\lambda_s \left\lceil \frac{\alpha}{\kappa} \right\rceil \Bigg((\delta_{DP} + \delta_{DA}) + \psi \log_2 \left(\frac{N_m k}{xy} \right) + \psi \log_2 k + 2\xi \Bigg). \tag{3.6}$$

Total cost on MAP: Thus, the total cost on each MAP can be obtained by adding Eqns. (3.3), (3.4), (3.5), and (3.6):

$$\Upsilon_{MAP} = \Upsilon_{MAP}^{LC} + \Upsilon_{MAP}^{RR} + \Upsilon_{MAP}^{RC} + \Upsilon_{MAP}^{DD}. \tag{3.7}$$

3.3.2 Home Agent

The total cost on the HA is due to the exchange of location query messages with CNs, RR messages, RCoA registration messages with MH and MAP, refreshing BU message with the MHs and tunneling of packets from CN to MAP, and vice versa.

Query messages: For each association between MH and CN, query and reply messages are exchanged between CN and HA. The HA has to search a database of size proportional to the number of mobile hosts under its domain and the lookup cost is $\psi\lambda_s \log_2(cN_m)$. Here we assume that the HA has a total of cN_m number of hosts under its domain. Hence, the cost on HA for query messages is as follows:

$$\Upsilon_{HA}^{QR} = N_m N_c (2\delta_Q \lambda_s + \psi\lambda_s \log_2(cN_m)). \tag{3.8}$$

Return routability messages: Before each BU message, RR messages are exchanged among the MH, HA and CN. The HA receives the Home Test Init (HoTI) message sent by the MH and forwards it to the CN. It also receives the Home Test (HoT) message sent by the CN and sends it back to MH. This happens for every MT_r seconds and for every MH-CN pair under the HA. Therefore, the cost on HA for RR messages are as follows:

$$\Upsilon_{HA}^{RR} = N_m N_c \times \frac{4\delta_{RR}}{MT_r}. \tag{3.9}$$

RCoA registration messages: For every region crossing between MAPs (happens every MT_r seconds), MH needs to register the RCoA with HA. Therefore,

$$\Upsilon_{HA}^{RC} = N_m \frac{2\delta_R + \beta_{HA}}{MT_r}. \tag{3.10}$$

Refreshing updates MHs send periodic refreshing updates to the HA so that the binding entries are not removed from the location database after the binding lifetime. Let the lifetime of the entries in the location database be T_{lf}. Therefore, $\lfloor \frac{T_r}{T_{lf}} \rfloor$ refreshing updates are sent to HA within time T_r. Thus, the frequency of sending periodic refreshing updates is $\eta_r = \lfloor \frac{T_r}{T_{lf}} \rfloor / T_r$. Therefore, the cost on the HA for refreshing BU packets are as follows:

$$\Upsilon_{HA}^{RB} = \eta_r N_m \big(2\delta_B + \beta_{HA}\big). \tag{3.11}$$

Data delivery cost: For every packet sent from CN to MH, transmission and processing costs (for location database lookup and encapsulation) are incurred at the HA. These costs are similar to that incurred at MAP, except that HA does not have to decapsualte the packet from the CN. Thus, the cost on the HA due to packet tunneling is as follows:

$$\Upsilon_{HA}^{DD} = \quad N_m N_c \lambda_s \left\lceil \frac{\alpha}{\kappa} \right\rceil \Big((\delta_{DP} + \delta_{DA}) + \psi \log_2(cN_m) + \xi \Big). \tag{3.12}$$

Total cost on HA: Thus, the total cost on each HA can be obtained by adding Eqns. (3.8), (3.9), (3.10), (3.11), and (3.12):

$$\Upsilon_{HA} = \Upsilon_{HA}^{QR} + \Upsilon_{HA}^{RR} + \Upsilon_{HA}^{RC} + \Upsilon_{HA}^{RB} + \Upsilon_{HA}^{DD}. \tag{3.13}$$

3.3.3 Mobile Host

The mobility signaling overheads on each MH are due to the exchange of RCoA and LCoA registration request/reply messages with MAP and HA.

RCoA Registration Messages: Each MH entering a MAP domain receives router advertisements and registers with a MAP sending (receiving) RCoA registration request (reply). This event happens every MT_r seconds. So the overhead on each MH associated with these registration events is given by

$$\Upsilon_{MH}^{RC} = \frac{2\sigma\delta_R}{MT_r}. \tag{3.14}$$

LCoA registration messages: Every subnet crossing by the MH (happens every T_r sec) within a MAP region triggers an LCoA registration message to the MAP. Therefore,

$$\Upsilon_{MH}^{LC} = \frac{2\sigma\delta_R}{T_r}. \tag{3.15}$$

Return routability messages: RR messages are exchanged among the MH, HA and CN before the MH informs HA about its moving out of a MAP-region. This happens for every MT_r seconds and for every MH-CN pair. Therefore, the cost on MH for RR messages is as follows:

$$\Upsilon_{MH}^{RR} = N_c \times \frac{4\delta_{RR}}{MT_r}. \tag{3.16}$$

Refreshing updates MHs send periodic refreshing updates to the HA to keep its binding entry valid. Therefore, the cost on the MH for sending refreshing BU packets is as follows:

$$\Upsilon_{MH}^{RB} = 2\eta_r\delta_B. \tag{3.17}$$

Packet delivery cost: Each MH communicates with N_c correspondent nodes through MAP and HA. If in a session, a file of size α is transferred from the CN to the MH, then the total number of packets in a session is $\left\lceil \frac{\alpha}{\kappa} \right\rceil$. If the average number of retransmission of each data (and Ack) is χ, there will be $(\chi+1)$ attempts of transmission for each data (and Ack) packets. Hence, the packet delivery cost for each MH per second is:

$$\Upsilon_{MH}^{PD} = N_c \left\lceil \frac{\alpha}{\kappa} \right\rceil \sigma \lambda_s (\delta_{DP} + \delta_{DA}). \tag{3.18}$$

Total cost on each MH: Thus, the total signaling overhead on each HA can be obtained by adding Eqns. (3.14), (3.15), (3.16), (3.17) and (3.18) as follows:

$$\Upsilon_{MH} = \Upsilon_{MH}^{RC} + \Upsilon_{MH}^{LC} + \Upsilon_{MH}^{RR} + \Upsilon_{MH}^{RB} + \Upsilon_{MH}^{PD}. \tag{3.19}$$

3.3.4 Complete network

In order to compute the total cost on the network, we have considered resources (bandwidth, processing power, etc.) consumed due to HMIPv6 protocol.

Query message: As each MH has an average of N_c CNs, the total number of CNs for all the MHs is $N_m N_c$. The CN and HA are h_w ($= h_p + h_{in} + h_p$) wired-hops away. The transmission cost for all the query and reply messages towards the HA is $N_c N_m (2h_w \delta_Q)\lambda_s$. The searching cost in the HA is $N_c N_m (\psi \lambda_s \log_2(cN_m))$. Hence, the cost of the network for the query messages from the CNs is as follows:

$$\Upsilon_{Net}^{QR} = N_m N_c \lambda_s (2\delta_q h_w + \psi(\log_2 cN_m)). \tag{3.20}$$

LCoA registration messages: Every subnet crossing by the MH within a MAP region triggers an LCoA registration message to be sent to the MAP. Let the distance between MH and MAP be one wireless hop and $h_p - 1$ wired hops. Hence, LCoA registration involves transmission cost of $2\delta_R$ in each of the $h_p - 1$ wired hops and one wireless hop. Due to frame retransmissions and medium access contentions at the data link layer of wireless links, transmission cost of a wireless hop is higher than that of a wired hop; we denote this effect by a proportionality constant, σ. LCoA registration also incurs processing cost at MAP. So

$$\Upsilon_{Net}^{LC} = N_m \frac{2\delta_R(h_p - 1 + \sigma) + \beta_{MAP}}{T_r}. \tag{3.21}$$

Return routability messages: The RR messages are sent every MT_r second by the MH to HA which forwards them to CN. The HoTI message follows the path between MH and HA which is of $h_w (= h_p + h_{in} + h_p)$ hops with one wireless hop and the path between HA and CN which is of h_w wired hops. Similar cost is incurred for each HoT message. Each CoTI message is sent directly to CN from the MH which uses h_w hops that include one wireless hop. Therefore, the cost on the network for RR messages is:

$$\Upsilon_{Net}^{RR} = \frac{N_m N_c}{MT_r} 2\delta_{rr} \Big(3h_w - 2 + 2\sigma\Big). \tag{3.22}$$

Refreshing binding updates To keep binding entry valid in the binding cache, MHs send periodic refreshing updates to the HA. Since the MH is h_w hops (including one wireless hop) away, the cost on the network for refreshing BU packets is:

$$\Upsilon_{Net}^{RB} = \eta_r N_m \big(2\delta_B(h_w - 1 + \sigma) + \beta_{HA}\big). \tag{3.23}$$

RCoA registration messages: The MAP processes the RCoA request and assigns an RCoA to the MH. As the MAP is h_p hops (that include one wireless hop) away from the MH, this RCoA registration incurs a transmission cost of $2\delta_R(h_p - 1 + \sigma)$ and a processing cost β_{MAP} at the MAP. The MAP informs the HA about this new RCoA registration that requires a transmission cost of $2\delta_R \Phi_{mh}$ and a processing cost of β_{HA} at the HA. Thus the RCoA registration cost for the network is

$$\Upsilon_{Net}^{RC} = N_m \frac{2\delta_R(h_p - 1 + \sigma) + \beta_{MAP}}{MT_r} + N_m \frac{2\delta_R h_w + \beta_{HA}}{MT_r}. \tag{3.24}$$

Data delivery cost: CN sends every data packet to MH through HA and then MAP. The cost required for the data packet to reach HA is $\delta_{DP}h_w$. Similar cost of $\delta_{DA}h_w$ is required for each ACK packet. The HA receives the data packets, encapsulates them and sends them to the MAP. Thus the cost of $(\delta_{DP} + \delta_{DA})h_w + 2\xi$ is required for this. The MAP receives the data packet on behalf of the MH from the HA, decapsulates the packet, and then encapsulates it to forward it to MH's current location using the translation table of RCoA to LCoA. Hence, it costs $(\delta_{DP} + \delta_{DA})(h_p - 1 + \sigma) + 4\xi$ for each data and Ack packet. In addition, the visitor list lookup at MAP costs $\psi \log_2 \frac{N_m k}{xy}$, and the IP routing table lookup costs another $\psi \log_2 k$. So tunneling each data packet and corresponding ACK packet from MAP to the MH costs $(\delta_{DP} + \delta_{DA})(h_p - 1 + \sigma) + 4\xi + \psi \log_2 \frac{N_m k}{xy} + \psi \log_2 k$. Since the total number of MHs in the network is N_m and we have assumed a uniform distribution of MHs in the network, the number of MHs under an MAP-domain is $\frac{N_m k}{xy}$. Thus, the cost related to data delivery is given by

$$\begin{aligned} \Upsilon_{Net}^{DD} = N_m N_c \lambda_s \left\lceil \frac{\alpha}{\kappa} \right\rceil \Bigg(\Big(& (\delta_{DP} + \delta_{DA})h_w + (\delta_{DP} + \delta_{DA})h_w + 2\xi \\ & + (\delta_{DP} + \delta_{DA})(h_p - 1 + \sigma) \Big) + 4\xi + \psi \log_2 \frac{N_m k}{xy} + \psi \log_2 k \Bigg). \end{aligned} \tag{3.25}$$

Total cost on the network: Therefore, the total cost on the complete network due to HMIPv6 protocol can be obtained by adding Eqns. (3.20), (3.21), (3.22), (3.23), (3.24), and (3.25):

$$\Upsilon_{Net} = \Upsilon_{Net}^{QR} + \Upsilon_{Net}^{LC} + \Upsilon_{Net}^{RR} \Upsilon_{Net}^{RB} + \Upsilon_{Net}^{RC} + \Upsilon_{Net}^{DD}. \tag{3.26}$$

3.4 Cost Model for SIGMA

In this section, we perform cost analysis of different mobility management entities of SIGMA: Location Manager, Mobile Host, and the complete network.

3.4.1 Location Manager

The total cost on the Location Manager is due to the location update messages sent by the MHs, and the query message about the current location of the MH exchanged with CNs.

Location updates: In SIGMA, every subnet crossing (that happens every T_r seconds) by an MH triggers a location update message to be sent to LM. The LM receives each LU message, processes (updates) the corresponding entry, and sends the acknowledgement back to the MH. Thus, each location update requires transmission cost ($2\delta_L$) and processing cost (β_{LM}) at LM. Hence, the cost on the LM due to location updates is as follows:

$$\Gamma_{LM}^{LU} = N_m \frac{2\delta_L + \beta_{LM}}{T_r}. \tag{3.27}$$

Refreshing updates MHs send periodic refreshing updates to the HA to keep its binding entry valid. Therefore, the cost on the MH for sending refreshing BU packets is as follows:

$$\Gamma_{LM}^{RB} = N_m(2\eta_r\delta_B + \beta_{LM}). \tag{3.28}$$

Query message: For each association between MH and CN, a query message about the current location of MH is sent to the LM and the LM sends the reply to that query after looking up the table. This involves transmission cost and lookup cost. The LM has to search a database of size proportional to number of mobile hosts. Since the session arrival rate is λ_s, each association lookup cost is $\psi\lambda_s \log_2 N_m$.

Now for all the sessions between the MHs and CNs, the signaling load on the LM is as follows:

$$\Gamma_{LM}^{QR} = N_m N_c (2\delta_Q \lambda_s + \psi \lambda_s \log_2 N_m). \tag{3.29}$$

Total cost on LM: The total cost on the LM can be obtained by adding Eqns. (3.27), (3.28) and (3.29).

$$\Gamma_{LM} \quad = \quad \Gamma_{LM}^{LU} + \Gamma_{LM}^{RB} + \Gamma_{LM}^{QR}.$$

3.4.2 Mobile Host

The signaling costs on the mobile host are due to the query messages exchanged with ARs, binding update messages exchanged with CNs, and location update and refreshing update messages sent to LM, and packet delivery.

DHCP message: When the MH approaches the radio coverage area of a new AR, it acquires a new IP address (IP2) from that AR by exchanging DHCP request/reply messages that involve transmission cost of $2\sigma\delta_D$ through the wireless media. Therefore, the cost incurred at the MH for query message is as follows:

$$\Gamma_{MH}^{DH} = \frac{2\sigma\delta_D}{T_r}. \tag{3.30}$$

Binding update message: After acquiring the IP address, the MH notifies CN about the availability of the new IP address through SCTP address dynamic reconfiguration option [18] whose transmission cost is $2\sigma\delta_S$. When MH moves further into the radio coverage area of the new AR, it sends ASCONF message [18] to CN, using which CN sets its primary destination (for MH) to IP2 (cost of $2\sigma\delta_D$). The MH also updates its local routing table to make sure that future outgoing packets are sent through new AR and this requires processing cost of β_{AR}. Therefore, the cost incurred at the MH for BU message is as follows:

$$\Gamma_{MH}^{BU} = N_c \frac{2\sigma\delta_S + 2\sigma\delta_S + 2\sigma\delta_S + \beta_{AR}}{T_r}. \tag{3.31}$$

Location update message: Every MH in SIGMA is required to send a location update message (costing $2\sigma\delta_L$) to LM after each subnet crossing. These LU and BU messages are transmitted through the wireless media. When the MH moves further and goes away from the coverage area of the previous AR, it sends another ASCONF message to CN to delete the previous IP address (IP1) which costs $2\sigma\delta_S$. Therefore, the cost incurred at the MH for location update message is as follows:

$$\Gamma_{MH}^{LU} = \frac{2\sigma\delta_L}{T_r}. \tag{3.32}$$

Refreshing updates MHs send periodic refreshing updates to the LM to keep its binding entry valid. Therefore, the cost on the MH for sending refreshing BU packets are as follows:

$$\Gamma_{MH}^{RB} = 2\eta_r\delta_B. \tag{3.33}$$

Data delivery cost: Each MH communicates with N_c correspondent nodes. If in a session, a file of size α is transferred from the CN to the MH, then the total number of packets in a session is $\left\lceil \frac{\alpha}{\kappa} \right\rceil$. Hence, the packet delivery cost for each MH per second is:

$$\Gamma_{MH}^{DD} = N_c \left\lceil \frac{\alpha}{\kappa} \right\rceil \sigma\lambda_s(\delta_{DP} + \delta_{DA}). \tag{3.34}$$

Total cost on MH: Thus, the total cost on each MH can be obtained by adding Eqns. (3.30), (3.31), (3.32), (3.33), and (3.34) as

$$\begin{aligned} \Gamma_{MH} &= \Gamma_{MH}^{DH} + \Gamma_{MH}^{BU} + \Gamma_{MH}^{LU} + \Gamma_{MH}^{RB} + \Gamma_{MH}^{DD} \\ &= \frac{2\sigma\delta_D}{T_r} + N_c \frac{6\sigma\delta_S + \beta_{AR}}{T_r} + \frac{2\sigma\delta_L}{T_r} + 2\eta_r\delta_B + N_c \left\lceil \frac{\alpha}{\kappa} \right\rceil \sigma\lambda_s(\delta_{DP} + \delta_{DA}). \end{aligned} \tag{3.35}$$

3.4.3 Complete network

In order to compute the total cost on the network as a whole, we will consider resources (such as bandwidth, processing power, etc) consumed due to the SIGMA mobility protocol.

Query message: For each association between MH and CN, query messages are exchanged between CN and LM. Each MH has an average of N_c number of CNs; therefore, total number of CNs for all the mobile hosts are $N_m N_c$. As the session arrival rates for each MH are assumed to be equal (λ_s), the transmission cost for all the query messages towards the LM is $N_c N_m (2h_w \delta_Q)\lambda_s$. The searching costs for the query messages are $N_c N_m (\psi \lambda_s \log_2 N_m)$. Hence, the cost of the network for the query messages from the CNs is,

$$\Gamma_{Net}^{QR} = N_m N_c \lambda_s (2\delta_Q h_w + \psi \log_2 N_m). \tag{3.36}$$

Location update message: Each subnet crossing by the MH triggers location update message to be sent to LM which processes the message and sends back acknowledgement to MH. The location update cost is proportional to the distance (in hops) between the MH and LM (note that there is one wireless link, and transmission cost in wireless link is higher than that of wired link by a factor of σ). Therefore, the resources (bandwidth and processing cost) used in the network for location updates are as follows:

$$\Gamma_{Net}^{LU} = N_m \frac{2(h_w - 1 + \sigma)\delta_L + \beta_{LM}}{T_r}. \tag{3.37}$$

Refreshing binding updates: To keep binding entry valid in the binding cache, MHs send periodic refreshing updates to the LM. Since the MH is h_w hops (including one wireless hop) away, the cost on the network for refreshing BU packets is as follows:

$$\Gamma_{Net}^{RB} = \eta_r N_m \big(2\delta_B(h_w - 1 + \sigma) + \beta_{LM}\big). \tag{3.38}$$

DHCP message: Whenever a MH comes within the coverage area of a new AR, DHCP request message is sent to the AR, which processes the request, and sends back the DHCP reply. So the costs associated with these messages are

$$\Gamma_{Net}^{DH} = N_m \frac{2\sigma\delta_D + \beta_{AR}}{T_r}. \tag{3.39}$$

Binding updates: After each subnet crossing of each MH, binding updates (ASCONF message with Add IP, Set primary, Delete IP messages) are sent to all the correspondent nodes. The binding update is proportional to the distance (in hops) between the MH and CN.

$$\Gamma_{Net}^{BU} = N_m N_c \frac{6(h_w - 1 + \sigma)\delta_S + \beta_{CN}}{T_r}. \tag{3.40}$$

Data delivery cost: After getting the IP address of the MH from the LM, the CN send data packets directly to the MH through $h_w - 1$ wired hops and one wireless hop. So the transmission cost for each data packet is $(h_w - 1 + \sigma)\delta_{DP}$, and a transmission cost of $(h_w - 1 + \sigma)\delta_{DA}$ for each acknowledgement packet. Hence the packet delivery cost for all the communications in the network can be obtained using the following equation:

$$\Gamma_{Net}^{DD} \quad = \quad N_m N_c \lambda_s \left\lceil \frac{\alpha}{\kappa} \right\rceil (\delta_{DP} + \delta_{DA})(h_w - 1 + \sigma). \tag{3.41}$$

Total cost on the network: Thus, the total cost on the complete network due to SIGMA protocol can be obtained by adding Eqns. (3.36), (3.37), (3.38), (3.39), (3.40) and (3.41):

$$\Gamma_{Net} = \Gamma_{Net}^{QR} + \Gamma_{Net}^{LU} + \Gamma_{Net}^{RB} + \Gamma_{Net}^{DH} + \Gamma_{Net}^{BU} + \Gamma_{Net}^{DD}. \tag{3.42}$$

3.5 Cost Model for HiSIGMA

In this section, we present entity-wise cost analysis for HiSIGMA. Specifically, we analyzed total cost for AZS, HZS and the whole network. The AZS and HZS – two key mobility components of HiSIGMA – have been chosen for entity-wise evaluation since resource consumptions at these entities are expected to be high, and may affect the performance of the whole network.

3.5.1 HiSIGMA network structure

Figure 3.2 shows the two dimensional environment where xy number of ARs are arranged in a way similar to that of HMIPv6 (see Fig. 3.1). There are m AZSs, each of which covers k subnets (let). The HZS is responsible for keeping location information of all the MHs moving in this environment though the number of MHs under HZS-domain is assumed to be higher.

The expected number of moves that causes MH's AZS-domain move-out can be obtained as $M = \frac{xy}{xy-k}$ using similar approach used in Eqn. (3.2).

3.5.2 Anchor Zone Server

As mentioned earlier in Sec 2.2.4, the AZS only deals with location management (not data transmission). The main tasks of AZS are processing 1) query messages from CNs, and 2) registration messages from MHs. The following paragraphs estimates the cost terms of AZS.

Query-lookup cost: Each association between MH and CN (that happens in every $1/\lambda_s$ sec) requires query message and corresponding reply to be exchanged between CN and AZS (see Fig. 2.4). On the average, each AZS has N_m/m MHs under its domain. So each lookup cost is proportional to $\log_2(N_m/m)$. Therefore, cost on AZS for query message is as follows:

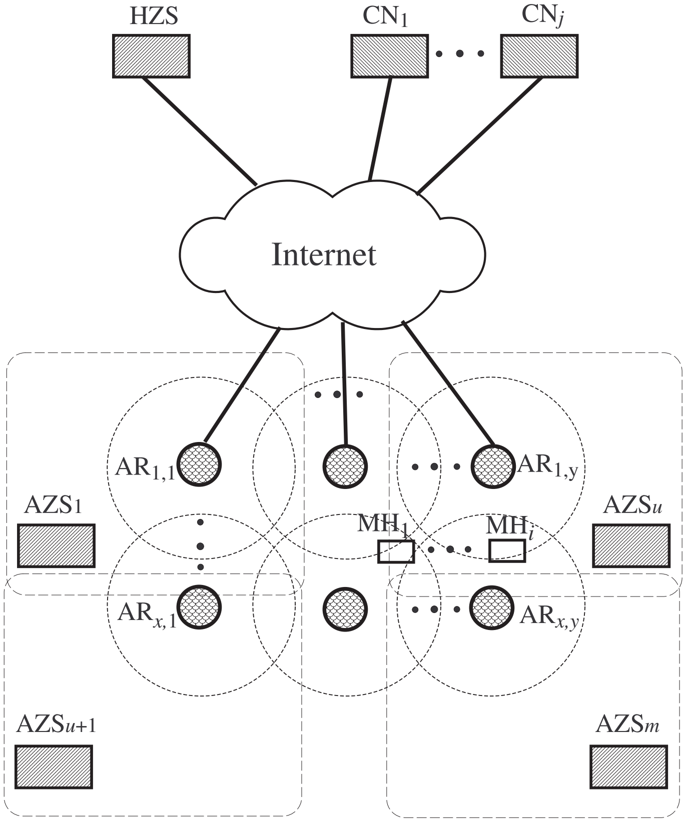


Figure 3.2: Network structure for HiSIGMA.

$$\Phi_{AZS}^{QR} = \lambda_s N_c \frac{N_m}{m} \big(2\delta_Q + \psi \log_2(N_m/m)\big). \tag{3.43}$$

Registration cost for micro and macro mobility: In HiSIGMA, an MH does not need to register with the HZS until the MH moves out of the region covered by an AZS, instead it only registers with the AZS. Therefore, every subnet crossing within a AZS (happens every T_r seconds) will trigger a registration message (and corresponding ACK) message to (from) the AZS, which incurs transmission cost of $2\delta_D$) and processing cost (β_{AZS}) at each AZS. Moreover, in case of the availability of multiple IP addresses, the MH should notify AZS with the primary IP address through dynamic address reconfiguration message which incurs transmission cost of δ_a at AZS. Since all the MHs are uniformly distributed among the m AZS domains, each AZS will have N_m/m number of MHs. Therefore, cost at AZS for registration message due to micro-mobility can be obtained as follows:

$$\Phi_{AZS}^{RG1} = \frac{N_m}{m} \times \frac{2\delta_L + \beta_{AZS} + 2\delta_D}{T_r}. \tag{3.44}$$

For every region crossing between the AZSs, transmission and processing costs are incurred (for registration request-reply messages) at the AZS which is given by,

$$\Phi_{AZS}^{RG2} = \frac{N_m}{m} \times \frac{2\delta_L + 2\beta_{AZS}}{MT_r}. \tag{3.45}$$

Total cost on AZS: Therefore, the total cost on each AZS can be obtained by adding Eqns. (3.43), (3.44), and (3.45):

$$\Phi_{AZS} = \Phi_{AZS}^{QR} + \Phi_{AZS}^{RG1} + \Phi_{AZS}^{RG2}. \tag{3.46}$$

3.5.3 Home Zone Server

In HiSIGMA, the HZS mainly processes 1) query messages from the CN, 2) registration messages due to MH macro-mobility, and 3) return routability test messages.

Query-Lookup cost: For each association (that happens in every $1/\lambda_s$ sec) between MH and CN, query message (and corresponding reply) about the current location of MH are exchanged between CN and HZS (see Fig. 2.4). The HZS replies with the IP address of the AZS which replies with the MH's current IP address(es). This incurs transmission cost at the HZS along with looking up a table whose size is proportional to number of MHs it (HZS) covers. Therefore, query-lookup cost on the HZS can be computed as follows:

$$\Phi_{HA}^{QR} = \lambda_s N_m N_c \big(2\delta_Q + \psi \log_2(cN_m)\big). \tag{3.47}$$

where, c is the ratio of number of MHs under HZS to N_m.

Registration cost for macro-mobility: For the inter-region movement of the MHs, transmission and processing costs are incurred at the HZS which is given by,

$$\Phi_{HZS}^{RG} = N_m \times \frac{2\delta_L + \beta_{HZS}}{MT_r}. \tag{3.48}$$

Refreshing updates MHs send periodic refreshing updates to the HZS to keep its binding entry valid. Therefore, cost on the HZS for refreshing BUs are as follows:

$$\Phi_{HZS}^{RB} = N_m(2\eta_r\delta_B + \beta_{HZS}). \tag{3.49}$$

Return routability messages: To avoid session hijacking, RR messages are exchanged among the MH, HZS and CN before MH sends BU message to the CN. The Home Test Init (HoTI) and Home Test (HoT) messages are sent through the HZS for RR procedure. This happens for every MT_r seconds and for every MH-CN pair under the HA. Therefore, the cost on HZS for RR messages is as follows:

$$\Phi_{HZS}^{RR} = N_m N_c \times \frac{4\delta_{RR}}{MT_r}. \tag{3.50}$$

Total cost on the HZS: Thus, the total cost on the HZS can be obtained by adding Eqns. (3.47), (3.48), (3.49), and (3.50):

$$\Phi_{HZS} = \Phi_{HZS}^{QR} + \Phi_{HZS}^{RG} + \Phi_{HZS}^{RB} + \Phi_{HZS}^{RR}. \tag{3.51}$$

3.5.4 Complete network

To compute the total cost on the whole system, we will consider resources consumed due to HiSIGMA protocol. These include transmission, lookup and processing costs incurred in the whole network along the route of various mobility protocol messages as well as the data packets. These costs are explained in the following paragraphs.

Query-lookup cost: The transmission cost for all the query and reply messages towards the HZS is $N_cN_m(2\delta_Q h_w)\lambda_s$. Per entry searching cost at HZS is $\psi\lambda_s \log_2(cN_m)$. Similar estimation can be done m AZSs of the network structure. Hence, the cost of the network for the query messages is given by,

$$\Phi_{Net}^{QR} = N_m N_c \lambda_s \big(2\delta_Q h_w + \psi \log_2(cN_m) + 2\delta_Q h_w + \psi \log_2(N_m/m)\big). \tag{3.52}$$

Registration costs for macro and micro mobility: As the MH only require to register with the HZS only if it moves out of the AZS-region. Otherwise, in every T_r sec, MH registers with the AZS which is h_p hops away that includes one wireless hop. MHs also exchange address reconfiguration messages with the AZS during this micro-mobility. Hence, the cost incurred at the network for MH's micro-mobility is as follows:

$$\Phi_{Net}^{RG1} = N_m \times \frac{2(\delta_L + \delta_D)(h_p - 1 + \sigma) + \beta_{AZS}}{T_r}. \tag{3.53}$$

On the other hand, for every region crossing (in every MT_r sec), MH registers with the HZS which is h_w hops (that includes one wireless hop) away from the MH. This incurs transmission and processing costs at the HZS, and processing cost at the AZS(s). Since there is only one location update per subnet crossing, no matter how many CNs an MH is communicating with, the number of CNs does not have any impact on the location update cost. Therefore, the cost incurred at the network for the macro-mobility of all the MHs can be obtained as follows:

$$\Phi_{Net}^{RG2} = N_m \frac{2\delta_L(h_w - 1 + \sigma) + \beta_{HZS}}{MT_r} + N_m \frac{2\delta_L(h_p - 1 + \sigma) + 2\beta_{AZS}}{MT_r}. \tag{3.54}$$

Refreshing binding updates: To maintain valid binding entry, MHs send periodic refreshing updates to the HZS. As the MH is h_w hops (including one wireless hop) away from the HZS, the cost on the network for refreshing BU packets are as follows:

$$\Phi_{Net}^{RB} = \eta_r N_m \big(2\delta_B(h_w - 1 + \sigma) + \beta_{HZS}\big). \tag{3.55}$$

Return routability messages: The RR messages are sent every MT_r sec by the MH to HZS which forwards them to CN. The HoTI message follow the path between MH and HZS (which is of h_w hops with one wireless hop) and the path between HZS and CN of h_w wired hops. Each HoT message incurs similar cost. Each CoTI message is sent directly to CN from the MH which uses h_w hops (that includes one wireless hop). Therefore, cost on the network for RR messages is as follows:

$$\Phi_{Net}^{RR} = \frac{N_m N_c}{MT_r} 2\delta_{RR}\big(h_w + h_w + h_w - 2 + 2\sigma\big). \tag{3.56}$$

Binding update cost: For each CN communicating with an MH, the MH needs to send a binding update after each handover. Therefore, the binding update cost per second in the whole system can be calculated by multiplying the number of MHs, the average number of communicating CNs per MH, and the average cost per binding update as follows:

$$\Phi_{Net}^{BU} = N_m N_c \frac{2(h_w - 1 + \sigma)\delta_B}{T_r}. \tag{3.57}$$

Packet delivery cost: Similar to the analysis in [16], we have considered data packet transmission cost while estimating the total cost on the whole system. HiSIGMA is free of packet encapsulation or decapsulation. Packets from CN to MH follows direct route with h_w hops including one wireless hop. As the average session length is α, and maximum transmission unit is κ, there are $\lceil \frac{\alpha}{\kappa} \rceil$ number of packets, and the packet rate can be obtained by $\lceil \frac{\alpha}{\kappa} \rceil \times \lambda_s$. The transmission cost for data packet and corresponding acknowledgement are $(\delta_{DP} + \delta_{DA})$. Thus, the packet delivery cost of the whole network can be expressed as:

$$\Phi_{Net}^{PD} = N_m N_c \lambda_s \left\lceil \frac{\alpha}{\kappa} \right\rceil (\delta_{DP} + \delta_{DA})(h_w - 1 + \sigma). \tag{3.58}$$

Total cost of the network: The total cost on the network due to HiSIGMA protocol can be obtained by adding Eqns. (3.52), (3.53), (3.54), (3.55), (3.56), (3.57), and (3.58) as follows:

$$\Phi_{Net} = \Phi_{Net}^{QR} + \Phi_{Net}^{RG1} + \Phi_{Net}^{RG2} + \Phi_{Net}^{RB} + \Phi_{Net}^{RR} + \Phi_{Net}^{BU} + \Phi_{Net}^{PD}. \qquad (3.59)$$

3.6 Cost Model for NEMO

In this section, we present the analytical cost model for NEMO and its mobility entities: HA, MR and complete network. It can be noted that in a mobile network, LMN, VMN and MRs are mobile nodes and MIPv6-capable. Thus, number of mobile nodes inside the mobile network is sum of number of LMN, VMN and MRs.

3.6.1 Home Agent

In NEMO, the HA keeps the location database of the mobile network. In fact, the location information of MR, LFNs and LMNs are kept in the HA whereas that of VMNs are kept in corresponding HAs since they belong to some other networks. The main tasks of HA are processing 1) query messages from CNs, 2) LU messages from MRs, 3) RR test messages, 4) BU messages to CNs, and 5) data delivery cost.

Query message: Every CN send query message to the HA at the beginning of every session. This requires a lookup at the HA which is proportional to the logarithm of the number of entries in the lookup table. So the lookup cost at HA is $\Psi_{HA}^{LK} = \psi \log_2(N_f + N_m)$. In addition, transmission cost is incurred for query-reply messages at the HA. Hence, the cost relating to query messages at HA is given by the following equation:

$$\Lambda_{HA}^{QR} = N_c(N_m + N_f)\lambda_s\Big(2\delta_Q + \Psi_{HA}^{LK}\Big). \qquad (3.60)$$

Location update messages: When the mobile network crosses subnets, MR sends LU message to the HA and the location database is modified by the HA which sends back acknowledgement to LU message. This happens in every T_r seconds. In addition, MRs and mobile nodes send periodic refreshing updates to the HA so that the entries are not removed from the the location database after the binding lifetime and the total frequency of sending LU and refreshing LU is $\eta_t = \Big(1 + \lfloor \frac{T_r}{T_{lf}} \rfloor\Big)/T_r$,

Each LU and corresponding Acknowledgement messages exchanged with HA incurs transmission and processing cost. The LU messages from mobile nodes go through one level of encapsulation which cost additional transmission cost of δ_{TH} and a processing cost of γ_t whereas the LU messages from the MR goes without encapsulation. In both cases, a lookup cost of Ψ_{HA}^{LK} is required. So the cost related to the LU and refreshing LU messages can be computed as follows:

$$\Lambda_{HA}^{LU} = \eta_t\Big(2\delta_L + \Psi_{HA}^{LK}\Big) + \eta_r\Big(2(\delta_L + \delta_{TH} + \gamma_t) + \Psi_{HA}^{LK}\Big). \tag{3.61}$$

Return routability messages: NEMO employs RR test before sending BU to the HA similar to the mechanism employed in route optimization of MIPv6 [17]. Before each BU message, RR messages are exchanged among the MR, HA and CN. The HA receives the Home Test Init (HoTI) message sent by the MR and forwards it to the CN. HA also receives the Home Test (HoT) message from the CN and sends it back to MR. This happens for every T_r seconds. The HA receives these RR messages for all CNs that are communicating with LMN. Therefore, the cost on HA for RR messages is as follows:

$$\Lambda_{HA}^{RR} = N_c(N_m + N_f)\frac{4\delta_{RR}}{T_r}. \tag{3.62}$$

Binding updates to CNs: To continue ongoing sessions with the CNs, mobile nodes inside the mobile network sends refreshing BUs to the CNs by tunneling them through the HA. The HA has to lookup the table, tunnel and transmit those BUs. Hence, cost incurred at the HA due to these BUs is given by,

$$\Lambda_{HA}^{BU} = 2N_c(N_m + N_f)\eta_r\Big(\delta_B + \delta_{TH} + \gamma_t + \Psi_{HA}^{LK}\Big). \tag{3.63}$$

Data delivery cost: In NEMO BSP, all data traffic to the mobile network are transmitted through the HA. In each session between the CN and MNN, an average of $\lceil \frac{\alpha}{\kappa} \rceil$ data packets are sent from the CN to MNN or vice versa. The successful reception of each data packet is confirmed by a corresponding ACK packet from the receiver. Therefore, the packet arrival rate is $\lambda_p = \lambda_s \lceil \frac{\alpha}{\kappa} \rceil$. As all the data traffic goes through the HA, it costs transmission cost data and ACK packets, extra IP-header processing and transmission cost as well as lookup cost. Therefore, the data delivery cost on the HA is given by,

$$\Lambda_{HA}^{DD} = N_c(N_m + N_f)\lambda_p\Big(\delta_{DT} + \delta_{DA} + 2\Big(\delta_{TH} + \gamma_t + \Psi_{HA}^{LK}\Big)\Big). \tag{3.64}$$

Total cost on HA: Thus, the total cost of the HA can be obtained by adding Eqns. (3.60), (3.61), (3.62) (3.63), and (3.64):

$$\Lambda_{HA} = \Lambda_{HA}^{QR} + \Lambda_{HA}^{LU} + \Lambda_{HA}^{RR} + \Lambda_{HA}^{BU} + \Lambda_{HA}^{DD}. \tag{3.65}$$

3.6.2 Mobile Router

In NEMO, the main tasks of the MRs are 1) IP address and prefix acquisition, 2) sending LU messages to HA, 3) sending binding updates to the CNs, 4)processing RR messages, and 5) processing data (ACK) packets to and from MNNs,

Acquiring IP address and prefixes: MRs acquire IP address from access router in the foreign network during each handoff by exchanging DHCPv6 request-reply messages through the wireless media which costs the following:

$$\Lambda_{MR}^{Acq} = \frac{2\sigma\delta_{DH}}{T_r}. \tag{3.66}$$

Location updates: After each handoff, each MR sends a LU message to the HA. In addition, periodic refreshing updates are also sent by the MRs and the mobile nodes through MR. Thus the cost on the MRs due to LU messages is as follows:

$$\Lambda_{MR}^{LU} \;=\; 2\sigma\eta_t\delta_L + 2\eta_r N_m\Big(\sigma(\delta_L + \delta_{TH}) + \gamma_t\Big). \qquad (3.67)$$

Binding updates to CNs: Mobile nodes send periodic refreshing BUs to the CNs through the MR updating the current address to continue ongoing sessions. This requires transmission of BU message through the wireless media with extra IP-header (encapsulation), and processing cost due to tunneling. Thus, the cost on the MRs for these BU messages is as follows:

$$\Lambda_{MR}^{BU} \;=\; 2\eta_r N_c(N_m + N_f)\Big(\sigma(\delta_B + \delta_{TH}) + \gamma_t\Big). \qquad (3.68)$$

Return routability messages: To ensure that the ongoing session is not hijacked by some malicious agent, before sending binding updates to the HA, it is essential to perform RR test to verify that the node can actually respond to packets sent to a given CoA [17]. Thus, the MR have to process and transmit RR messages on behalf of the mobile nodes under its domain which incurs the following cost:

$$\Lambda_{MR}^{RR} \;=\; \frac{4\sigma N_m\delta_{RR}}{T_r}. \qquad (3.69)$$

Data delivery cost: Data packet delivery incurs transmission cost through the wireless media (with extra IP-header), and processing cost for the MR. Therefore, the data delivery cost at the MRs is given by,

$$\Lambda_{MR}^{DD} \;=\; \lambda_p N_c(N_m + N_f)\Big(\sigma(\delta_{DT} + \delta_{DA} + \delta_{TH}) + \gamma_t\Big). \qquad (3.70)$$

Total cost on the MR: Therefore, total cost of each MR can be obtained by adding Eqns. (3.66), (3.67), (3.68), (3.69), and (3.70),

$$\Lambda_{MR} = \Lambda_{MR}^{Acq} + \Lambda_{MR}^{LU} + \Lambda_{MR}^{BU} + \Lambda_{MR}^{RR} + \Lambda_{MR}^{DD}. \qquad (3.71)$$

3.6.3 Complete network

In order to compute the signaling load on the network as a whole, we consider all the resources (such as, bandwidth, processing power, etc.) consumed in all network entities. The cost of the network due to the operation of NEMO BSP include query messages exchanged between HA and CN, RR messages, location update messages, binding updates to CNs, and data delivery to CN.

Query message: At the beginning of each session between a MNN and a CN, query messages are exchanged between CN and HA. As the session arrival rates for each MNN are assumed to be equal (λ_s), the transmission cost for all the query and reply messages towards the HA is $2N_c(N_m+N_f)(h_p+h_{in}+h_p)\delta_Q\lambda_s$. Let us assume h_w ($= h_p + h_{in} + h_p$). The searching cost in the HA is $N_c(N_m+N_f)\psi\lambda_s \log_2(N_m+N_f)$. Hence, the cost of the network for the query messages from the CNs is,

$$\Lambda_{Net}^{QR} = \lambda_s N_c(N_m + N_f)\Big(2h_w\delta_Q + \psi \log_2(N_m + N_f)\Big). \tag{3.72}$$

Return routability messages: The RR messages are sent every T_r second by the MRs (on behalf of the MNNs) to HA which forwards them to CN. The HoTI message follow the path between MR and HA which consists of h_w wired hops with one wireless hop (between the MR and the AR). The path between HA and CN contains h_w wired hops. Similar cost is incurred for each HoT message. Each CoTI message is sent directly to CN from the MR which uses h_w wired hops and one wireless hop. Therefore, cost on the network for RR messages is as follows:

$$\Lambda_{Net}^{RR} = \frac{N_c(N_m + N_f)}{T_r} \times 2\delta_{rr}\Big((h_w + \sigma) + h_w + (h_w + \sigma)\Big). \tag{3.73}$$

Location updates: After each handoff, the MRs and the mobile nodes send LU to the HA informing the newly acquired IP address and prefixes. As the HA is $(h_w + 1)$ hops (including h_p wireless hop) away from the MR, each LU from MR (and corresponding Ack) message incurs a transmission cost of $\delta_L(h_w + \sigma)$, and a lookup cost of Ψ_{HA}^{LK} at the HA. The LU messages from LMNs (or VMNs) travel one more wireless hop than the MR with additional transmission cost for tunneling header and corresponding processing cost. Thus, the cost of LU messages on the network is given by,

$$\begin{aligned}\Lambda_{Net}^{LU} &= 2\delta_L\eta_t(h_w+\sigma) + 2N_m\eta_r\Big((\delta_L+\delta_{TH})(h_w+2\sigma)+\gamma_t\Big) \\ &\quad + (\eta_t+\eta_r N_m)\Psi_{HA}^{LK}.\end{aligned} \tag{3.74}$$

Binding updates to CNs: To maintain continuous connectivity with the CNs that are communicating with the mobile nodes, binding updates informing the care-of-address are sent to the CNs. These BU messages goes through and h_w wired hops and two wireless hop, on the average, to reach a CN. Thus, cost required to send BUs to CNs is given by,

$$\Lambda_{Net}^{BU} = 2N_c(N_m+N_f)\eta_r\Big((h_w+2\sigma)(\delta_B+\delta_{TH})+\gamma_t\Big). \tag{3.75}$$

Data delivery cost: All the data and corresponding Ack) packets, that is, goes through HA. The path between a MNN and the HA contains h_w wired links and 2 wireless links whereas the path between HA and CN contains h_w wired links. In addition, data packets incur table lookup in the HA. Thus, the costs related to data delivery and processing by the network are given by

$$\Lambda_{Net}^{DD} = \lambda_p N_c(N_m+N_f)\Bigg(\Big((h_w+2\sigma)+h_w\Big)(\delta_{DT}+\delta_{DA}+2\delta_{TH})+2\gamma_t+2\Psi_{HA}^{LK}\Bigg). \tag{3.76}$$

Total cost of the network: Therefore, the total cost of the complete network due to NEMO protocol can be obtained by adding Eqns. (3.72), (3.73), (3.74), (3.75), and (3.76):

$$\Lambda_{Net} = \Lambda_{Net}^{QR} + \Lambda_{Net}^{RR} + \Lambda_{Net}^{LU} + \Lambda_{Net}^{BU} + \Lambda_{Net}^{DD}. \tag{3.77}$$

3.7 Cost Model for SINEMO

In this section, we perform an entity-wise cost analysis of SINEMO protocol. We have chosen CLM and MR, the two crucial entities of SINEMO. This is because CLM is involved in every session between a CN and a MNN, and all communications with the mobile network are carried out through the MR.

3.7.1 Central Location Manager

The CLM is an important mobility management entity for SINEMO as it is responsible for recording up-to-date location information of the mobile network. The total cost on the CLM should be excessive which may result in unavailability of several MNNs or the whole network.

In SINEMO, the CLM keeps the location database of the mobile network and has the tasks of 1) processing LU messages from MRs, 2) processing query messages from CNs, 3) searching the location database, 4) processing refreshing BU messages and 5) processing RR messages.

Query message: At the beginning of every session between CNs and MNNs, query (and corresponding reply) messages exchanged between CLM and the CNs. This incurs transmission cost of $2N_c\delta_Q\lambda_s$. In addition, a table lookup at CLM is required that is proportional to the logarithm of the number of MNNs, that is, $\log_2(N_m + N_f)$. Therefore, cost on CLM for query message is as follows:

$$\Delta_{CLM}^{QR} \quad = \quad 2N_c(N_m + N_f)\delta_Q\lambda_s + N_c\psi\lambda_s \log_2(N_m + N_f). \qquad (3.78)$$

Return routability messages: We assume that SINEMO employs RR test to prevent session hijacking similar to the mechanism employed in route optimization of MIPv6 [17]. This test verifies that the node (sending BU) can actually respond to packets sent to a given CoA. Before each BU message, RR messages are exchanged among the MR, CLM and CN. The CLM receives the Home Test Init (HoTI) message sent by the MR and forwards it to the CN. CLM also receives the Home Test (HoT) message from the CN and sends it back to MR. This happens for every T_r seconds and for every MH-CN pair under the MN. Therefore, cost on CLM for RR messages can be obtained as follows:

$$\Delta_{CLM}^{RR} = N_c(N_m + N_f)\frac{4\delta_{RR}}{T_r}. \tag{3.79}$$

Location update messages: When MN crosses subnets, MR acquires new IP address from the foreign network and notifies the CLM using LU message. The LU contains the new address of the LLM and the public addresses of the MHs inside the MR's domain. The MR sends such aggregated LU to the CLM requiring a transmission cost of δ_{AL} and the CLM sends back the acknowledgement requiring a transmission cost of δ_L. The CLM has to process the LU message and update the location database of all the nodes inside the MN. Thus, the cost (transmission and processing) on the CLM due to the LU messages is given by,

$$\Delta_{CLM}^{LU} = \frac{(\delta_{AL} + \delta_L) + \beta_{CLM}}{T_r}. \tag{3.80}$$

Refreshing update messages: During the subnet residence time, the MR sends refreshing BU to the CLM and all the CNs so that the binding entry is not expired. Thus, the cost of CLM is as follows:

$$\Delta_{CLM}^{RBU} = \eta_r(\delta_{AL} + \delta_L). \tag{3.81}$$

Total cost on the CLM: Thus, the total cost of the CLM can be obtained by adding Eqns. (3.78), (3.79) (3.80), and (3.81):

$$\Delta_{CLM} = \Delta_{CLM}^{QR} + \Delta_{CLM}^{RR} + \Delta_{CLM}^{LU} + \Delta_{CLM}^{RBU}. \tag{3.82}$$

3.7.2 Mobile Router

In a mobile network, the MR is the default gateway of all the MNNs. All the essential mobility management signaling go through this entity. Therefore, we evaluate the total cost incurred at the MR.

In SINEMO, the main tasks of each MR are 1) IP address and prefix acquisition, 2) updating the public IP addresses in the Network Address Translation (NAT) table, 3) sending updates to the CNs, 4) sending LUs to the CLM, 5) registration of MHs, 6) sending refreshing BU messages, 7) processing data (ACK) packets to and from MNNs, and 8) processing RR messages.

Acquiring IP address and prefixes: The MR acquire IP addresses and prefixes from the AR in the foreign network during each handoff by exchanging DHCPv6 request-reply messages through the wireless media. After acquiring the IP addresses for the nodes inside the MN, MR reserves public IP addresses for the MNNs and modifies the NAT table whose size is proportional to $(N_m + N_f)$. Since each entry of the NAT table will be updated after each handoff, the cost is proportional to $(N_m + N_f)\log_2(N_m + N_f)$. Therefore,

$$\Delta_{MR}^{Acq} = \frac{2\sigma\delta_{DH} + \psi(N_m + N_f)\log_2(N_m + N_f)}{T_r}. \tag{3.83}$$

Return routability messages: To prevent session hijacking RR messages are exchanged through MR. The cost on the MR associated with the RR message can be obtained as follows:

$$\Delta_{MR}^{RR} = \frac{4\sigma(N_m + N_f)N_c\delta_{RR}}{T_r}. \tag{3.84}$$

Updating sessions table and sending BU to CNs: To maintain continuous connectivity with the CNs that are communicating with the MNNs, the MR keeps a table known as Sessions table that records the CN-MNN pair of the ongoing sessions. Each entry of the sessions table is a triple with CN's IP address, MNN's current public address and MNN's private IP-address. After acquiring the IP address and prefixes at each handoff, the MR uses the newly assigned public addresses (to the MNNs in the NAT table) to modify the session table of size proportional to number of sessions. If we assume, each CN has one ongoing session with a MNN, then number of sessions is equal to $N_c(N_m + N_f)$. Thus updating the session table have a cost proportional to $N_c(N_m + N_f) \log_2 N_c(N_m + N_f)\big)$. In addition, the MR sends BUs to (and receive binding acknowledgement from) the CNs and the transmission cost associated with these BUs is $2\sigma\delta_B N_c$ in every handoff. Therefore, the cost on the MR regarding the update of the session table and transmission of the BU messages is given as:

$$\Delta_{MR}^{BU} \;=\; \frac{N_c(N_m + N_f) \log_2 N_c(N_m + N_f) + 2\sigma\delta_B N_c(N_m + N_f)}{T_r}. \qquad (3.85)$$

Location updates to CLM: After each handoff, the MR sends LU to CLM informing newly acquired IP address and prefixes. This is done by using one LU message containing the domain name (identification) and Care of Address (CoA) tuples of the MR as well as all the MNNs under its domain. Thus, the cost of the MR to transmit such LU message is as follows:

$$\Delta_{MR}^{LU} \;=\; \frac{\sigma(\delta_{AL} + \delta_L)}{T_r}. \qquad (3.86)$$

Refreshing update messages: MR sends refreshing BU to the CLM and the CNs with a frequency of η_r which costs the following for the MR:

$$\Delta_{MR}^{RBU} = \sigma\eta_r(\delta_{AL} + \delta_L)\big(1 + N_c(N_m + N_f)\big). \qquad (3.87)$$

Data delivery cost: In every CN-MNN session, $\lceil \frac{\alpha}{\kappa} \rceil$ data packets are sent along with corresponding ACK. Total data/Ack packet arrival rate to a MR is $\lambda_p = \lambda_s \lceil \frac{\alpha}{\kappa} \rceil$. Each data packet arriving from the CN is intercepted by MR which modifies the destination address by private IP address searching the NAT table of size proportional to $(N_m + N_f)$. The opposite is done for reverse path, that is, private IP is replaced by public IP address looking up the NAT table. Moreover, transmission cost is incurred through the wireless media. Therefore, data delivery cost at the MR is given by,

$$\Delta_{MR}^{DD} = \lambda_p N_c (N_m + N_f)\Big(\psi \log_2(N_m + N_f) + \sigma(\delta_{DP} + \delta_{DA})\Big). \tag{3.88}$$

Total cost on the MR: Therefore, the total cost of the MR can be obtained by adding Eqns. (3.83) through (3.88):

$$\Delta_{MR} = \Delta_{MR}^{Acq} + \Delta_{MR}^{RR} + \Delta_{MR}^{BU} + \Delta_{MR}^{LU} + \Delta_{MR}^{RBU} + \Delta_{MR}^{DD}. \tag{3.89}$$

3.7.3 Complete network

In order to compute the total cost of the network as a whole, we consider all the resources (such as, bandwidth, processing power etc) consumed in all network entities. This includes cost incurred for query messages exchanged between CLM and CN, local registration of MHs, RR messages, LU messages, BUs to CNs, and data delivery cost.

Query message: At the beginning of each session, query messages are exchanged between CN and CLM which incurs transmission and lookup costs. The CN and CLM are h_w ($= h_p + h_{in} + h_p$) wired hops away. Therefore, the transmission costs for the query-reply messages between CN and CLM are $N_c(N_m + N_f)(2h_w\delta_Q)\lambda_s$ whereas the lookup cost in the CLM is $N_c(N_m + N_f)\psi\lambda_s \log_2(N_m + N_f)$. Hence, the cost of the network for the query messages from the CNs is,

$$\Delta_{Net}^{QR} = 2\lambda_s N_c (N_m + N_f) h_w \delta_Q + \psi\lambda_s N_c (N_m + N_f)\log_2(N_m + N_f). \tag{3.90}$$

NAT translation: In a foreign network, the MR acquires IP address for the MNNs and reserves public IP addresses for the MNNs and modifies the NAT table whose size is proportional to $(N_m + N_f)$. This happens at every handoff. Therefore,

$$\Delta_{Net}^{NAT} = \frac{\psi(N_m + N_f)\log_2(N_m + N_f)}{T_r}. \tag{3.91}$$

Return routability messages: The RR messages are sent every T_r second by the MRs to CLM which forwards them to CN. Each HoTI and HoT messages follow a path of h_w wired hops and one wireless hop between MR and CLM whereas the path between CLM and CN contains h_w wired hops. Similar cost is incurred for each HoT message. Each CoTI message is sent directly to CN from the MR which contains h_w wired hops and one wireless hop. Therefore, cost for RR messages is as follows:

$$\begin{aligned}\Delta^{RR}_{Net} &= \frac{N_c(N_m+N_f)}{T_r} \times 2\delta_{RR}\big((h_w+\sigma)+h_w+(h_w+\sigma)\big) \\ &= 2N_c(N_m+N_f)\delta_{RR}(3h_w+2\sigma)/T_r.\end{aligned} \tag{3.92}$$

Location updates: After each handoff, the MR send LUs to the CLM informing the newly acquired IP address and prefixes. As the CLM is h_w wired hops and one wireless hop away from the MR, each LU and corresponding ACK message incurs a transmission cost of $(\delta_{AL}+\delta_L)(h_p+h_{in}+h_p+\sigma)$, and a processing cost β_{CLM} at the CLM. Thus, the cost of the network for LU message is given by,

$$\Delta^{LU}_{Net} \;=\; \frac{(\delta_{AL}+\delta_L)(h_w+\sigma)+\beta_{CLM}}{T_r}. \tag{3.93}$$

Binding updates to CNs: To maintain continuous connectivity with the CNs that are communicating with the MNNs, BUs informing the newly assigned IP address are sent to the CNs by the MRs. BUs by the MR go through one wireless hop, and h_w wired hops to reach a CN. In addition, processing cost of $N_c(N_m+N_f)\log_2 N_c(N_m+N_f)$ are incurred at the MR. Moreover, the MR sends η_r $(=\lfloor\frac{T_r}{T_{lf}}\rfloor)$ refreshing BUs to CLM and all CNs in every T_r. Thus, cost for BUs and refreshing BUs to CNs is given by,

$$\begin{aligned}\Delta^{BU}_{Net} = \frac{1}{T_r} &\times \Big(2\delta_B N_c(N_m+N_f)(h_w+\sigma)+N_c(N_m+N_f)\log_2 N_c(N_m+N_f)\Big)+ \\ \frac{\eta_r}{T_r} &\times \Big((\sigma+h_w)(\delta_{AL}+\delta_L+2N_c(N_m+N_f)\delta_B\Big).\end{aligned} \tag{3.94}$$

Data delivery cost: In SINEMO, CN sends every data packet to MNN using direct route unlike NEMO. The data and ack packets travel directly through h_w wired and one wireless hops to reach the MR which updates destination address and forward it to MNN. Thus, the data delivery cost is as follows:

$$\Delta_{Net}^{DD} = N_c(N_m + N_f)\lambda_p\Big(\psi \log_2(N_m + N_f) + (h_w + \sigma)(\delta_{DP} + \delta_{DA})\Big). \tag{3.95}$$

Total cost on the network: Therefore, total cost on complete network due to SINEMO protocol can be obtained by adding Eqns. (3.90), (3.91), (3.92), (3.93), (3.94) and (3.95):

$$\Delta_{Net} = \Delta_{Net}^{QR} + \Delta_{Net}^{NAT} + \Delta_{Net}^{RR} + \Delta_{Net}^{LU} + \Delta_{Net}^{BU} + \Delta_{Net}^{DD}. \tag{3.96}$$

3.8 Performance Metrics

Since no performance metrics of mobility protocols exists in terms of signaling costs, we define two performance metrics to evaluate the effectiveness of the protocols. They are normalized signaling overhead and efficiency.

3.8.1 Normalized overhead of an entity

Normalized overhead of any mobility entity can be defined as the percentage overhead per unit (net) data transmission cost. This can be computed by the ratio of total cost of any entity per unit data transmission cost.

The net data delivery cost does not include overhead and can be computed as follows:

$$\Psi_{Net-DD} = N_m N_c \lambda_s \left\lceil \frac{\alpha}{\kappa} \right\rceil \big(\delta_{DP} + \delta_{DA}\big). \tag{3.97}$$

Therefore, normalized overhead of a mobility entity can be computed by dividing its total cost with Ψ_{Net-DD}.

3.8.2 Efficiency of mobility protocols

We define another performance metric to evaluate the efficacy of mobility protocols in terms of signaling costs as there exists no such metric. Efficiency of a mobility protocol is defined as the ratio of data delivery cost (when an optimal route is used) to the total cost (that includes signaling and data delivery costs) required for the mobility protocol.

3.8.2.1 Host mobility protocols

For the host-mobility protocols, the net data delivery cost of the network is as follows:

$$\Psi^{Host}_{Net-DD} = N_m N_c \lambda_s \left\lceil \frac{\alpha}{\kappa} \right\rceil (\delta_{DP} + \delta_{DA})(h_w - 1 + \sigma). \tag{3.98}$$

Therefore, the efficiency of a host protocol can be obtained by dividing Ψ^{Host}_{Net-DD} with the total cost of the whole system. For example, the efficiency of HMIPv6 protocol can be computed as follows:

$$\eta^{HMIPv6} = \frac{\Psi^{Host}_{Net-DD}}{\Upsilon_{Net}}. \tag{3.99}$$

3.8.2.2 Network mobility protocols

The net data delivery cost of NEMO/SINEMO can be obtained as follows:

$$\Psi^{Network}_{Net-DD} = (N_m + N_f) N_c \lambda_s \left\lceil \frac{\alpha}{\kappa} \right\rceil \big(\delta_{DP} + \delta_{DA}\big)(h_w + 2\sigma). \tag{3.100}$$

Hence, efficiency of a network mobility can be computed by dividing $\Psi^{Network}_{Net-DD}$ with the total cost of the whole system. For example, the efficiency of SINEMO protocol can be computed as follows:

$$\eta^{SINEMO} = \frac{\Psi^{Network}_{Net-DD}}{\Delta_{Net}}. \tag{3.101}$$

Table 3.1: Values of parameters used in the numerical analysis.

Parameter	**Value**	**Parameter**	**Value**
N_m	400	N_c	5
N_f	100	k	12
x	51	y	34
δ_L	0.6	δ_B	0.6
δ_Q	0.6	δ_R	1.4
δ_{DP}	5.72	δ_{DA}	0.6
δ_{RR}	0.6	δ_R	0.6
σ	10	ψ	0.3
ξ	0.4	δ_{TH}	0.4
T_r	70 sec	T_{be}	90 sec
α	10240 bits	κ	512 bits
h_{in}	10	h_p	2
λ_s	0.01	c	10
β_{CLM}	0.3	β_{AR}	0.30
β_{CN}	0.3	β_{HA}	0.30
β_{MH}	0.3	β_{MR}	0.30
β_{AZS}	0.3	β_{HZS}	0.30
β_{MAP}	0.3	β_{LM}	0.30

3.9 Results

In this section, we present numerical results demonstrating the impact of network size, mobility rate, traffic rate and data volume on the total cost of host and network mobility protocols and its entities, along with the comparison between their efficiencies and normalized overhead.

The parameter values used in numerical analysis are derived using approaches similar to those used in [10, 15, 16]; each cost metric is a relative quantity and is based on the specific packet size (unit cost for 100 bytes [15, 16]). For example, if a signaling packet is 60 bytes long, the corresponding transmission cost is 0.6.

The values of parameters are listed in Table 3.1. In order to obtain the general trend/impact of different network parameters, we have varied their values in a wide range. For example, we varied number of MHs between 200 to 700; number of CNs per MH between 1 to 10; session arrival rate between 0.01 to 0.1; average session size between 10 Kb to 100 Kb; Session to Mobility Ratio (SMR) between 0.50 to 4 (SMR is defined as $T_r \times \lambda_s$).

3.9.1 HMIPv6 vs. SIGMA

First, we compare the total cost, normalized overhead and efficiency of SIGMA with that of HMIPv6.

3.9.1.1 LM vs. HA

Fig. 3.3 shows the total cost of LM (SIGMA) and HA (HMIPv6) as a function of number of MHs for different subnet residence times. Total cost of HA increases a much higher rate than that of LM since HA is heavily involved in data transmission unlike LM (of SIGMA). In SIGMA, the location management and data delivery are decoupled from each other, thereby reducing total cost on its key mobility entity (LM).

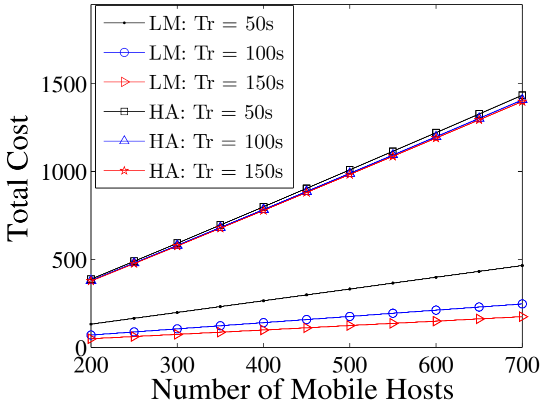


Figure 3.3: Impact of number of MHs on total cost of HA (HMIPv6) and LM (SIGMA).

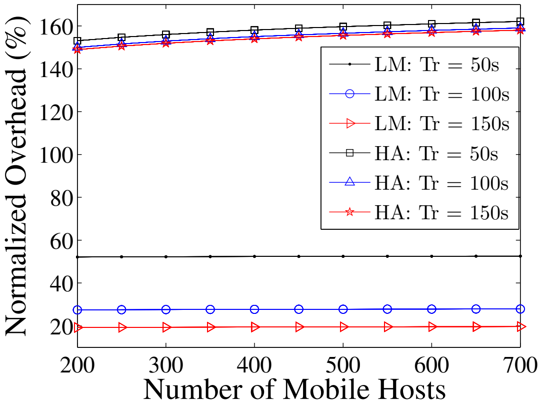


Figure 3.4: Normalized overhead of the HA (HMIPv6) and LM (SIGMA) vs. number of MHs.

Figure 3.4 shows the impact of number of MHs on the normalized overhead of LM (SIGMA) and HA (HMIPv6). Again, we find that the normalized overhead of HA is much higher than that of LM due to the involvement of HA in the data transmission process. The overhead decreases with the increase of subnet residence times since higher T_r value implies less mobility rate, and this reduces various signaling traffic, such as, location updates, binding updates, etc.

Figure 3.5 shows the total cost of LM (SIGMA) and HA (HMIPv6) as a function of SMR for different session lengths. Total cost remains almost constant (actually decreases very slowly) with the increase of SMR. Increase of SMR (meaning slow mobility of mobile node) decreases the mobility signaling cost due to less number of subnet crossing. However, for all different session lengths, the total cost of HA is much higher than that of LM.

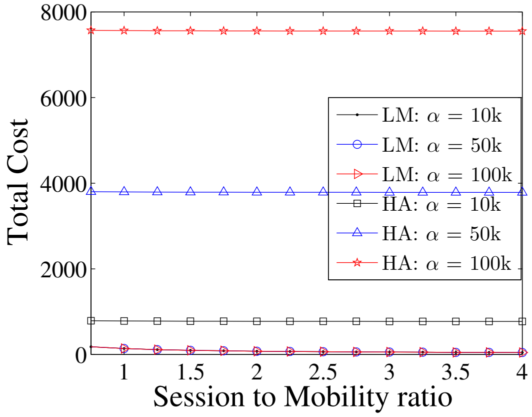


Figure 3.5: Impact of SMR on total cost of HA (HMIPv6) and LM (SIGMA).

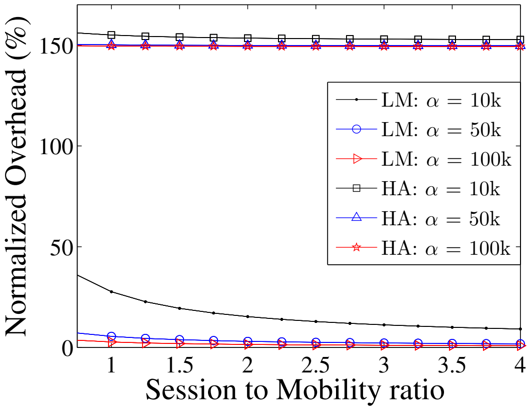


Figure 3.6: Normalized overhead of the HA (HMIPv6) and LM (SIGMA) vs. SMR.

Figure 3.6 shows the percentage normalized overhead of LM (SIGMA) and HA (HMIPv6) with respect to SMR for different session lengths. The normalized overhead of HA is much higher than that of LM. This is because the HA is directly involved in packet tunneling while the LM is not. The direct involvement in data delivery increases extra overhead on the HA, resulting in higher normalized overhead of HA.

3.9.1.2 Complete network

Fig 3.7 shows the total cost of SIGMA and HMIPv6 protocol as a function of number of MHs for different number of Internet hops. It is found the total cost incurred for SIGMA protocol is much less than HMIPv6 protocol. This is because of the sub-optimal routing of data packets and use of bidirectional tunnels in HMIPv6 protocol.

Figure 3.8 shows the efficiency of SIGMA and HMIPv6 protocol with respect to number of CNs for different session lengths. It is found that the efficiency of SIGMA is much higher than that of HMIPv6 protocol. This is because the total cost of SIGMA is much less than that of HMIPv6. SIGMA uses direct route between the MH and the CN whereas HMIPv6 uses sub-optimal routing through the HA, resulting in its higher total cost of the system, thereby reducing the overall efficiency.

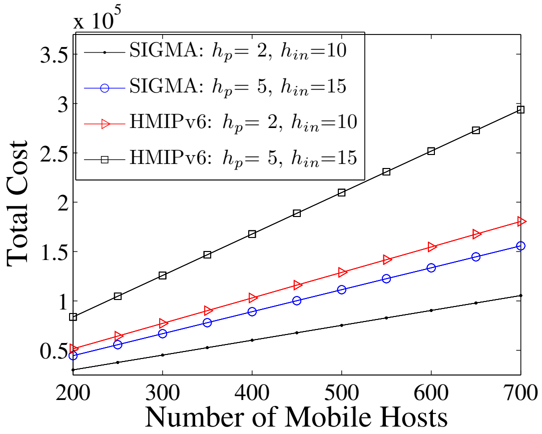


Figure 3.7: Impact of number of MHs on total cost of HMIPv6 and SIGMA.

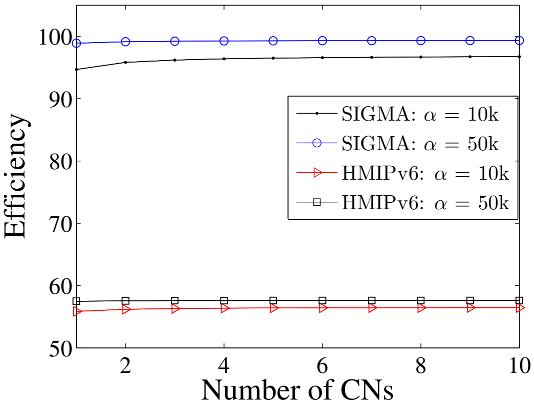


Figure 3.8: Efficiency of HMIPv6 and SIGMA vs. number of CNs.

3.9.1.3 Discussion

In this subsection, we find that the total costs and the normalized overhead of mobility management entities of SIGMA are significantly less than that of HMIPv6 whereas the efficiency of SIGMA is much higher than HMIPv6. SIGMA protocol benefits from exploitation of the direct route between the communication nodes which is not the case in HMIPv6. Moreover, the LM is not at all involved in data forwarding task in SIGMA; this relieves LM from extra overhead, unlike HMIPv6.

3.9.2 HMIPv6 vs. HiSIGMA

In this subsection, we compare the total cost, efficiency and normalized overhead of mobility management entities of HMIPv6 and HiSIGMA. The AZS and HZS of HiSIGMA correspond to the MAP and HA of HMIPv6, respectively.

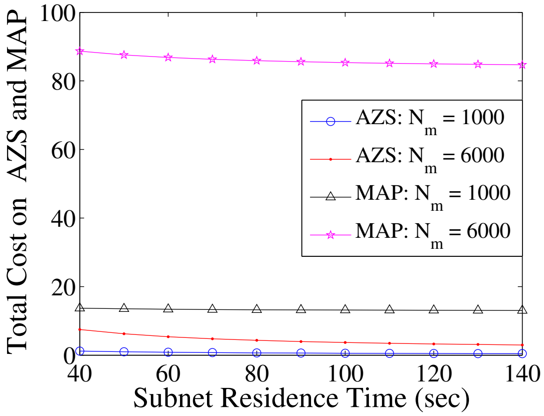


Figure 3.9: Total cost at each AZS and MAP as functions of subnet residence time.

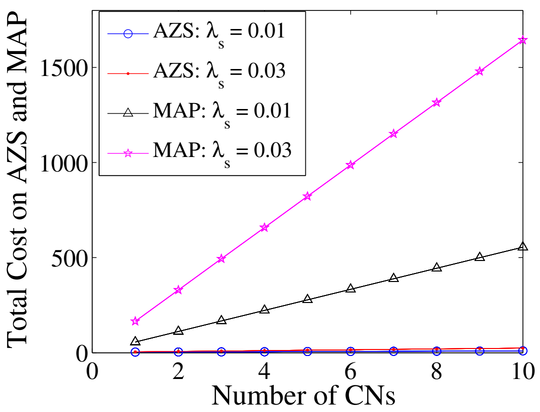


Figure 3.10: Total cost at each AZS and MAP as functions of number of CNs.

3.9.2.1 AZS vs. MAP

Figure 3.9 shows the impact of subnet residence time (T_r) on the total cost of each AZS (HiSIGMA) and MAP (HMIPv6). The total cost on each AZS is found to be much less than that for each MAP. This is because data transmission in HiSIGMA is decoupled from hierarchical location management and there is no tunneling or encapsulation of data packets through the AZS in HiSIGMA (unlike HMIPv6). From Fig. 3.9, we also find that total cost reduces for higher values of T_r which implies slower speed of MHs. Slow speed of MHs causes less number of handoffs, thereby reducing signaling costs, e.g., registration, BU, RR costs.

In Fig. 3.10, the total costs of each AZS and MAP are shown as functions of number of CNs. Higher number of CNs generates more traffic, thereby increasing the load on the MAP, unlike for AZS which does not deal with data traffic. Hence, the cost for AZS is very low compared to MAP.

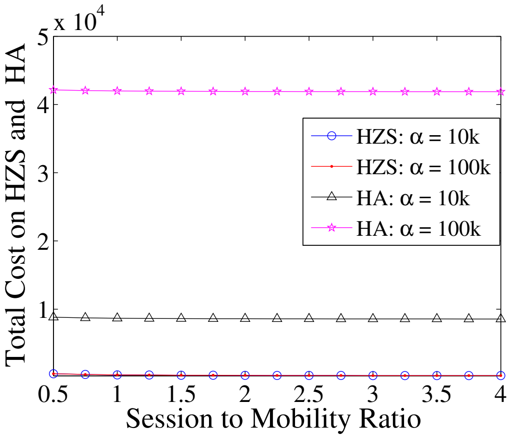


Figure 3.11: Impact of SMR on total cost of HZS (HiSIGMA) and HA (HMIPv6) for different session lengths.

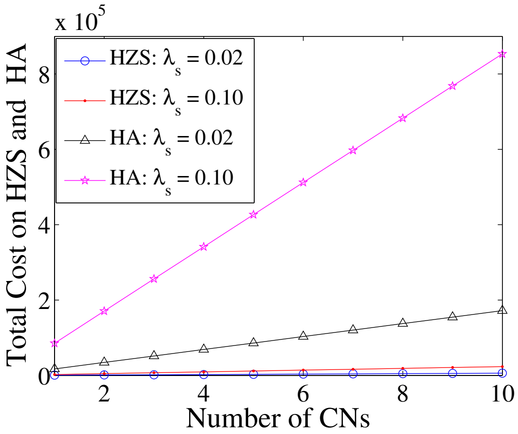


Figure 3.12: Impact of number of CNs on total cost of HZS (HiSIGMA) and HA (HMIPv6) for different session arrival rates.

3.9.2.2 HZS vs. HA

The impact of SMR on total cost of HZS (HiSIGMA) and HA (HMIPv6) is shown in Fig. 3.11 for different session lengths. We kept λ_s fixed at 0.01 while varying T_r between 50 and 400 sec, which yields a SMR in the range 0.50 - 4. Higher session lengths produces more data; hence, data delivery cost increases compared to the signaling traffic. This increases the total cost of HA. However, the total cost of HZS is not affected much by higher session length since HZS has nothing to do with data packets in HiSIGMA.

Figure 3.12 shows the impact of number of CNs on the total cost of HZS (HiSIGMA) and the HA (HMIPv6) for different session arrival rates. Again, the total cost of HA is much higher than that of HZS due to the tunneling/encapsulation of data packets between MH-CN pairs unlike in HiSIGMA. However, the total cost of HZS slowly increases for higher values of N_c due to higher signaling costs, e.g., query-lookup, registration and RR cost.

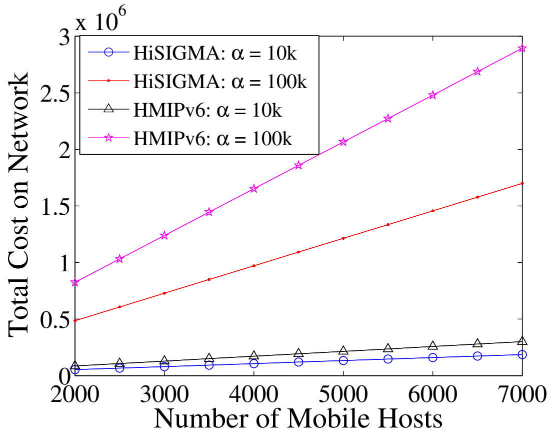


Figure 3.13: Total cost on the network as functions of number of MHs.

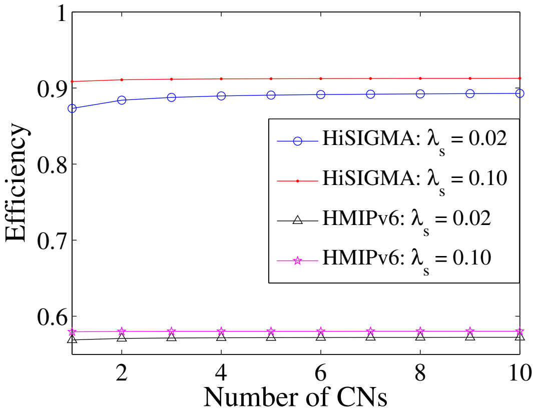


Figure 3.14: Efficiency of the whole system vs. number of CNs.

3.9.2.3 Complete network

Figure 3.13 shows the impact of number of MHs on the total cost of the complete system of HiSIGMA and HMIPv6 for varying session lengths. Higher number of MHs increases the total cost in all cases. However, the rate of increase is much higher for HMIPv6 protocol due to additional overhead of packet tunneling through the HA, and use of non-optimal route between MH-CN pairs.

Figure 3.14 shows the impact of number of CNs and session arrival rate on the efficiency of HiSIGMA and HMIPv6 protocols. It is found that HiSIGMA exhibits much higher efficiency than HMIPv6. Moreover, higher value of λ_s causes more data traffic into the system, thereby raising the efficiency values of both protocols.

Figure 3.15 shows the impact of number of MHs and subnet residence times on the efficiency of HiSIGMA and HMIPv6 protocols. Again, the efficiency of HiSIGMA is found to be much higher than HMIPv6 due to the fact that signaling overhead (including tunneling) is large in HMIPv6. In addition, efficiency for smaller subnet

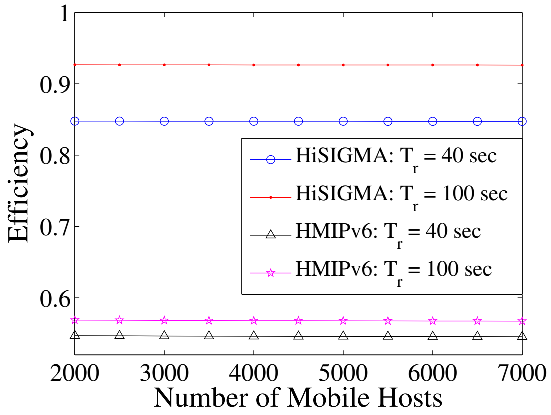


Figure 3.15: Efficiency of the whole system vs. number of MHs.

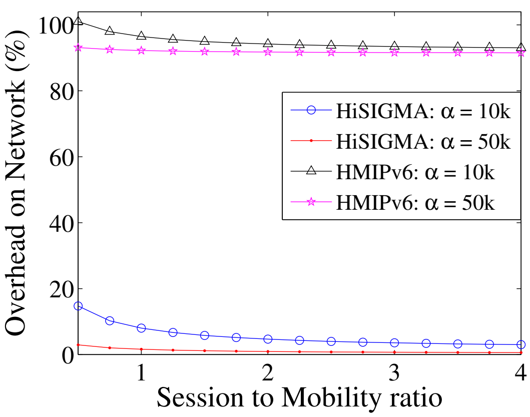


Figure 3.16: Percentage overhead of the whole system vs. SMR.

residence times are found to be less because smaller value of T_r (implies higher MH speed) produces more signaling traffic, such as, BU, RR and registration messages.

Finally, Fig. 3.16 shows the impact of SMR on the percentage overhead on the whole system per unit data traffic of HiSIGMA and HMIPv6 protocols. The overhead on HMIPv6 system is found to be much higher than HiSIGMA due to the use of suboptimal path and tunneling.

3.9.2.4 Discussion

In summary, we find that the total costs of mobility management entities of HiSIGMA are much less than that of HA and MAP (HMIPv6), respectively. This is because data transmission and location management are decoupled in HiSIGMA unlike HMIPv6. In addition, the efficiency of HiSIGMA is much higher than HMIPv6 since HiSIGMA lacks any encapsulation and tunneling of data packets and uses direct route between MH and CN.

3.9.3 NEMO vs. SINEMO

In this subsection, we compare SINEMO with NEMO entities with respect to their total cost, normalized overhead and efficiency of protocol operations.

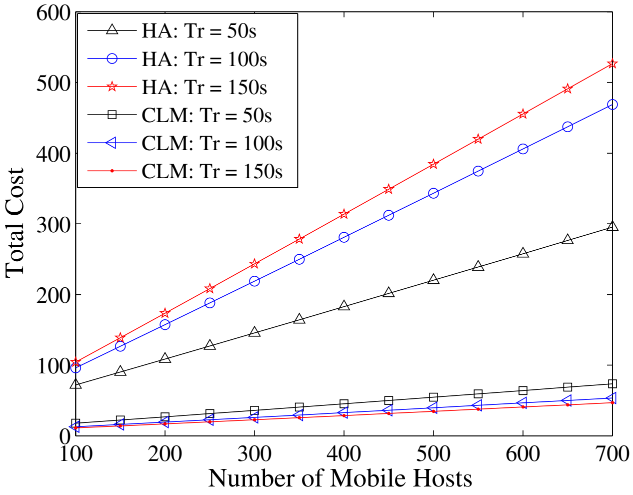


Figure 3.17: Impact of number of mobile hosts on the total cost of the HA and CLM for different subnet residence times.

3.9.3.1 CLM vs. HA

We compare NEMO's HA with SINEMO's CLM in Fig. 3.17 as their tasks are similar. The cost of NEMO's HA is found to be much higher than that of SINEMO's CLM as the first data packets of each session are routed through HA-MR. Total costs of HA-MR and CLM increase for higher number of MHs. However, in terms of T_r, total cost of CLM and HA-MR behave just opposite. For NEMO, when T_r increases, refreshing binding cost increases, although costs related to handoff reduces due to lower handoff frequency. Other costs (query and data delivery) remain unchanged. The net result is the increase of total cost. For SINEMO, the effect of refreshing BUs are much less than that related to handoff costs, thereby producing reduced total cost.

3.9.3.2 MR

In Fig. 3.18, the total cost of each MR is shown for varying number of mobile hosts and LFNs. Increase in LFNs results in constant shifting of the total cost graph due to the increase in query message cost and data delivery cost. Again, cost for SINEMO is found to be less than NEMO.

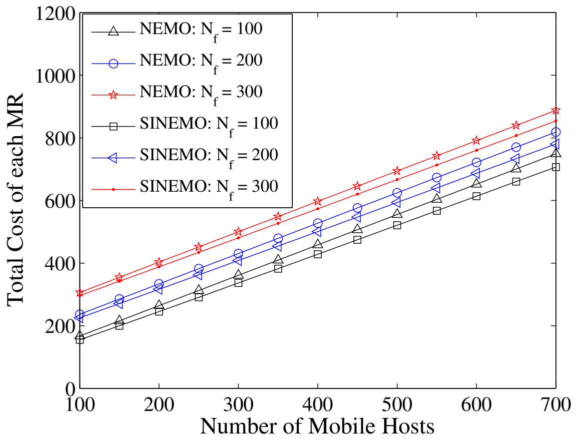


Figure 3.18: Impact of number of MHs on the total cost of each MR for different number of LFNs.

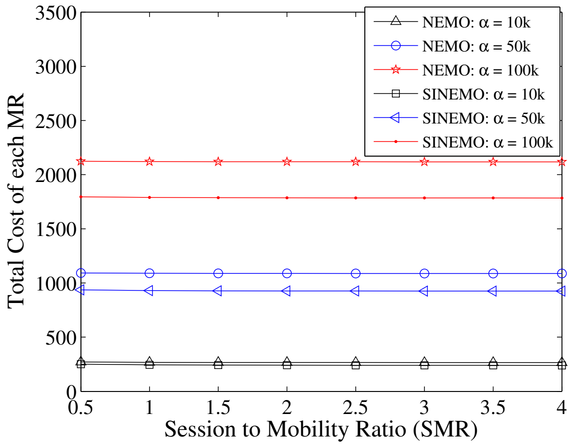


Figure 3.19: Impact of SMR on the total cost of each MR for different session lengths.

In Fig. 3.19, the total costs of each MR (for NEMO and SINEMO) are shown as a function of Session to Mobility Ratio (SMR) which is defined as $\lambda_s \times T_r$. We keep λ_s constant while varying the value of T_r between 50 to 400 sec.The session lengths are also varied. Higher session length causes more data packets to be routed through each MR, resulting in higher cost. The total cost is found to be invariant of SMR due to the dominance of data delivery cost while NEMO having higher cost than SINEMO in each case.

3.9.3.3 Complete network

The total cost of the complete network is shown as a function of number of mobile hosts in Fig. 3.20. Increased number of mobile hosts sends higher number of location and binding updates; in addition, query for the mobile hosts also increases for higher number of mobile hosts in the MN. The total cost is shown for different number of hops (h_{in}) in the Internet.The slope of the total cost rises for higher values of h_{in} since its value influences all the costs of the network.

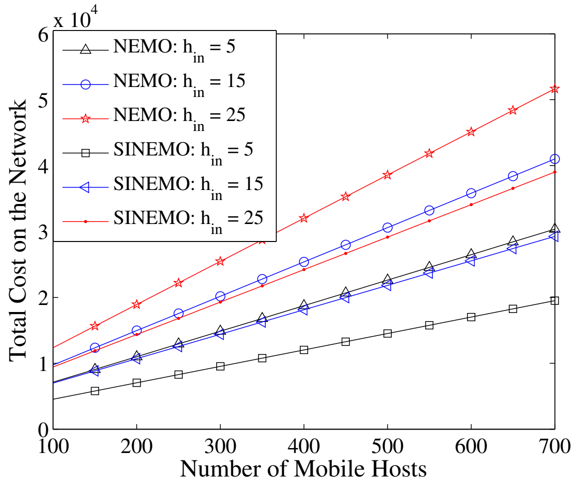


Figure 3.20: Impact of number of mobile hosts on the total cost of the network for different number of hops in Internet.

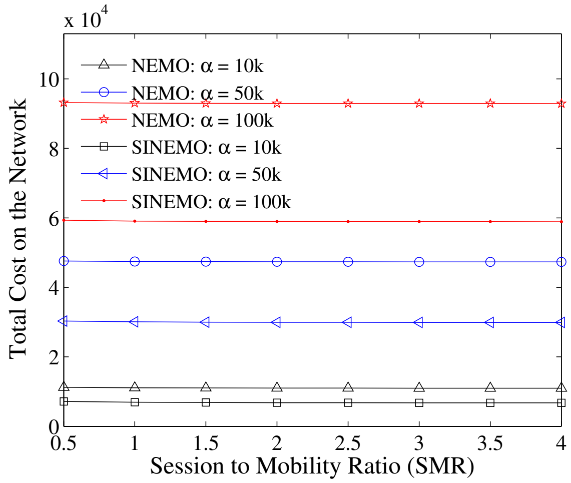


Figure 3.21: Impact of SMR on the total cost of the network for different session lengths.

In Fig. 3.21, total cost of the network is shown as a function of SMR for different session length. It is found that the total cost does not vary much (around 1%) with respect to SMR. This implies that data delivery cost (through optimized and unoptimized route) dominates the total cost.

3.9.3.4 Efficiency

In Fig. 3.22, efficiencies of NEMO and SINEMO protocols are shown for varying number of MHs and for different subnet residence times. Efficiency of SINEMO is found to be higher than that of NEMO, as SINEMO uses direct route to send/receive data packets between MNNs and CNs. Efficiency of each protocol increases for higher subnet residence times as the costs related to mobility signaling reduces due to fewer number of handoffs.

Figure 3.23 shows the efficiencies of NEMO and SINEMO for various SMRs and different session lengths. Efficiencies of both protocols increase for increased session lengths since the ratio of signaling traffic to data traffic becomes smaller. However, SINEMO shows a higher efficiency than NEMO irrespective of session lengths.

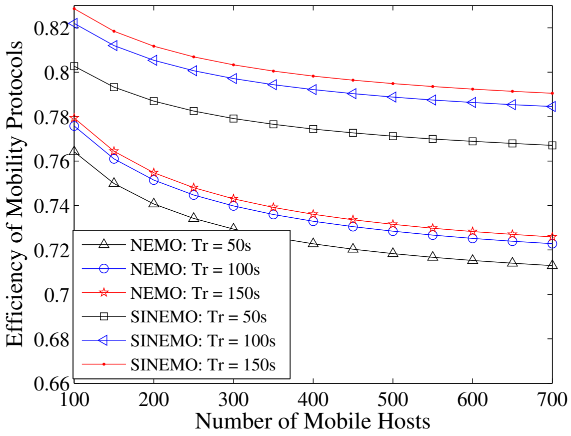


Figure 3.22: Efficiency of SINEMO and NEMO vs. number of MHs for various subnet residence times.

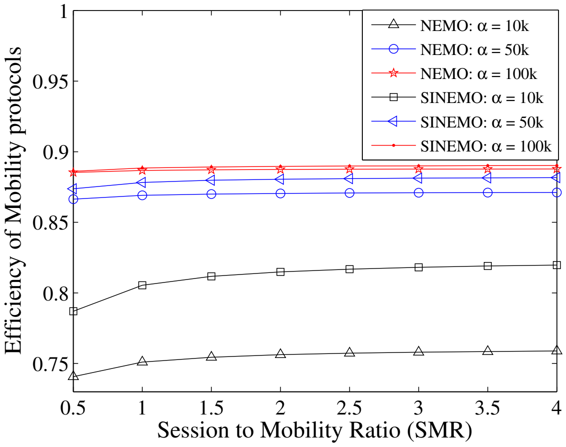


Figure 3.23: Efficiency of SINEMO and NEMO BSP vs. SMR for various session lengths.

3.9.3.5 Discussion

From Figs. 3.17, 3.18, 3.19, 3.20, and 3.21, we find that the total cost on various entities of SINEMO is much less than that of NEMO since SINEMO uses optimal route between mobile network and an arbitrary node in the Internet. Moreover, SINEMO has higher efficiency than NEMO (see Figs. 3.22 and 3.23) irrespective of session lengths, number of hosts and subnet residence time of the mobile network.

3.10 Summary

In this chapter, we have developed comprehensive cost analysis models to estimate total costs and efficiencies of different host and network mobility protocols and their key entities. Moreover, we have defined two novel metrics, namely normalized overhead and efficiency, to compare the performance of the mobility protocols and its entities. Our results show that the SIGMA (and SINEMO) incur much lower overhead on its key entities and yield higher efficiency than HMIPv6 (and NEMO) irrespective of session lengths, network size and mobility rate. Our comprehensive cost model can be used as a framework for estimating total cost of key mobility management entities of other handover protocols, and can aid in decision making to choose the most efficient protocol for future all-IP mobile and wireless networks. In the next chapter, we focus on the quantitative scalability analysis of the host and network mobility protocols based on the cost analysis presented in this chapter.

Chapter 4

Scalability Analysis of Mobility Protocols

With the rapid growth and popularity of mobile devices, increasingly larger number of IP-enabled mobile devices now require mobility support to gain ubiquitous Internet access, thereby putting enormous load on the mobility management entities. This may lead to serious consequences, specially for critical applications, such as in-flight communications for air traffic signaling in commercial aircrafts or in a battlefield where continuous connectivity with peers is very crucial. Therefore, it is essential to choose a mobility protocol that is scalable with the future growth of networks and user demands. In this chapter, we focus on the scalability analysis of different IP-mobility protocols. Based on the cost models presented in the previous Chapter 3, we perform scalability analysis of the mobility protocols (HMIPv6, SIGMA, NEMO and SINEMO) with respect to network size, mobility rate, and traffic rate.

4.1 Introduction

Increased number of mobile devices have significantly raised the amount of load on mobility management entities of a network. These mobility management entities were not initially designed to absorb such enormous load. Therefore, it is crucial to perform scalability analysis of mobility protocols to know whether these protocols can function properly with respect to the future growth of networks and user demands.

A number of researches on scalability analysis of networking protocols can be found in the literature. Santivanez et al. [19] present a novel framework to study the scalability of routing algorithms in ad hoc networks. This framework has been used in [20] and [21] for the scalability analysis of mobile ad hoc network and wireless sensor network, respectively. In our work, we use the notion of scalability factors from [19] since it is an excellent framework for quantitative scalability analysis.

A few simulation and testbed-based scalability analysis [22, 23] on IP-mobility protocols have also been performed. Gwon et al. [22] present an analysis on scalability and robustness of Mobile IPv6 [17] and related protocols using a large-scale simulation. Hautala et al. [23] present a study on the scalability of Mobile IPv6 in a wireless LAN laboratory testbed where multiple users hand over simultaneously. However, simulation-based works only focus on a particular scenario being simulated for a given set of system parameters. In contrast, analytical models represent general scenarios which provide better insights into the behavior of the system being analyzed.

The author is not aware of any research work that quantitatively analyzes the scalability of the mobility protocols which is required to visualize the effects of future network expansion on the performance of the protocols. This work is believed to be the first such work. This work will help in finding out the impact of network parameters on mobility management entities.

The objective of the work presented in this chapter is to perform a comprehensive scalability analysis of the host and network mobility management entities (of HMIPv6, SIGMA, NEMO and SINEMO) based on the mobility signaling overhead and to determine the influence of network size, mobility rate, traffic rate and data volume on them.

We have chosen HMIPv6 in our analysis (to compare it with our lab-proposed SIGMA protocol) as HMIPv6 is designed to reduce the signaling cost of the base MIPv6, and it has the lowest signaling cost in all versions of MIPv6 enhancements [5]. On the other hand, NEMO was chosen since it is the (IETF-proposed) base network mobility protocol and we want to show the scalability feature of SINEMO with respect to NEMO.

The contributions of the work presented in this chapter is to develop analytical models for the entity-wise scalability analysis of various mobility management entities of HMIPv6, SIGMA, NEMO and SINEMO with respect to network size, mobility rate and traffic rate.

Our results show that the host mobility protocols (HMIPv6 and SIGMA) and network mobility protocols (NEMO and SINEMO) have asymptotically identical scalability feature as far as the complete system is concerned though some of mobility entities exhibit differences in terms of scalability feature.

Our comprehensive scalability model can be used as a framework to perform quantitative scalability analysis of different mobility and handover protocols, and can aid in decision making to choose the most efficient and scalable protocol for future all-IP mobile and wireless networks. This can also help in visualizing the effects of future network expansion on the performance of the mobility protocols.

The rest of the chapter is organized as follows. In Section 4.2, we define scalability of mobility protocols along with the assumptions, notations and the scalability parameters of the model. The scalability models of HMIPv6, SIGMA, NEMO and SINEMO entities are presented in Section 4.3, followed by the results in Section 4.4. Section 4.5 has the concluding remarks.

4.2 Modeling Preparation

The definition of scalability for mobility protocols, scalability parameters and notations are explained in the following subsections.

4.2.1 Definition of Scalability

Santivanez et al. [19] presented a novel framework to study the scalability of routing algorithms in ad hoc networks. The same framework has been used in [20] and [21] to analyze the effect of several factors on the scalability of MANETs and wireless sensor network. We use this notion of scalability (from [19]) since it is an excellent framework for asymptotic scalability analysis which is also used in [20] and [21].

According to Santivanez et al. [19], scalability is the ability of a network to support the increase of its limiting parameters without degrading the network performance. Mobility protocol's scalability can thus be defined as the ability to support continuous increase of different (network) parameter values without degrading the performance of various entities that are responsible for mobility management. Examples of such limiting parameters are network size, mobility rate, traffic rate, etc.

Let $\Gamma^X(\lambda_1, \lambda_2, \dots)$ be the total overhead/cost induced by mobility protocol X, dependent on parameters $\lambda_1, \lambda_2, \dots$ (such as network size, mobility rate and traffic rate). Therefore, the protocol X's Mobility Scalability Factor (MSF) with respect to a parameter λ_i is defined as follows:

$$\rho^X_{\lambda_i} = \lim_{\lambda_i \to \infty} \frac{\log \Gamma^X(\lambda_1, \lambda_2, ...)}{\log \lambda_i}. \tag{4.1}$$

Protocol X is said to be more scalable than protocol Y with respect to parameter λ_i if $\rho^X_{\lambda_i} \leq \rho^Y_{\lambda_i}$.

4.2.2 Assumptions

Following are the assumptions of the scalability model:

- For mobility scalability factor computation, increase of a scalability parameter does not impact other system parameters.

- Similar mobility model and network topology are assumed for the two protocols.
- All mobile hosts are identical and uniform.
- Uniform distribution of mobile hosts over the region of the network.
- Session arrival rate for each mobile host is equal.
- The data (file) size in each session is equal.
- We consider costs on various entities considering only one mobile network (in case of network mobility protocols).

The above assumptions are made for all the mobility protocols so that they can be compared easily. Moreover, the scalability factor computation requires scalability parameters to be independent of one another.

4.2.3 Notations

For the scalability analysis, we use all the notations defined in Section 3.2.2. The scalability models are based on the cost analysis models presented in Chapter 3. We use the following notation to denote mobility scalability factor:

ρ_{λ}^{X} Scalability factor of Protocol X with respect to parameter λ.

4.2.4 Scalability parameters

For scalability analysis, we have chosen some parameters that represent network size, mobility rate, traffic rate and data volume. These parameters are mutually exclusive in a sense such that the value of one of the parameters can be increased independently while keeping others constant.

For the scalability analysis of host mobility protocols, we focus on the following network parameters:

- Network size which is represented by the number of mobile hosts (N_m) and the number of CNs per MH (N_c).
- Speed of the mobile host (V)
- Traffic rate which is represented by session arrival rate (λ_s)
- Data volume which is represented by session length (α)

For the scalability analysis of network mobility protocols, we focus on the following network parameters:

- Network size which is represented by the number of mobile hosts (N_m), number of LFNs (N_f) and and number of CNs per MNN (N_c)
- Speed of the mobile network (V) which is inversely proportional to subnet residence time (T_r)
- Traffic rate which is represented by packet arrival rate (λ_p)

4.3 Scalability Analysis

In this section, we perform the scalability analysis of SIGMA, HMIPv6, NEMO and SINEMO.

4.3.1 SIGMA

We derive the mobility scalability factors for different entities of SIGMA: Location Manager, Mobile Host and complete network based on their total cost.

4.3.1.1 Location Manager

In SIGMA, every subnet crossing (that happens every T_r seconds) by an MH triggers a location update message to be sent to LM. In addition, periodic binding refresh messages are sent by the MHs to the LM so that binding entries are not removed after their lifetime. Moreover, in every MH-CN association, LM has to deal with the location query message. Therefore, the total cost of the LM is as follows (see Section 3.4 for details):

$$\begin{aligned}\Gamma_{LM} &= \Gamma_{LM}^{LU} + \Gamma_{LM}^{RB} + \Gamma_{LM}^{QR} \\ &= N_m \frac{2\delta_L + \beta_{LM}}{T_r} + N_m(2\eta_r\delta_B + \beta_{LM}) + N_m N_c \lambda_s\big(2\delta_Q + \psi \log_2 N_m\big) \\ &= \frac{\Pi_1 N_m}{T_r} + \Pi_2 N_m + \Pi_3 N_m N_c \lambda_s + \Pi_4 N_m N_c \lambda_s \log_2 N_m.\end{aligned} \tag{4.2}$$

where $\Pi_1 = 2\delta_L + \beta_{LM}$, $\Pi_2 = (2\eta_r\delta_B + \beta_{LM})$, $\Pi_3 = 2\delta_Q$, $\Pi_4 = \psi$. Π_1, Π_2, Π_3 and Π_4 are constants as the quantities defined by them are unit transmission and processing cost which does not change with the increase of other (scalability) parameters.

Therefore, we get the asymptotic cost expression for LM using the Θ notation [1] as follows:

$$\Gamma_{LM} = \Theta(N_m(V + \lambda_s N_c \log_2 N_m)). \tag{4.3}$$

We consider $T_r \propto (1/V)$ for the above expression. Hence, SIGMA's MSF for the Location Manager with respect to N_m, N_c, V, λ_s and α can be computed as follows:

$$\rho_{N_m}^{LM-SIG} = \lim_{N_m \to \infty} \frac{\log(N_m(V + \lambda_s N_c \log_2 N_m))}{\log N_m} = 1 \tag{4.4}$$

$$\rho_{N_c}^{LM-SIG} = \lim_{N_c \to \infty} \frac{\log(N_m(V + \lambda_s N_c \log_2 N_m))}{\log N_c} = 1 \tag{4.5}$$

$$\rho_{V}^{LM-SIG} = \lim_{V \to \infty} \frac{\log(N_m(V + \lambda_s N_c \log_2 N_m))}{\log V} = 1 \tag{4.6}$$

$$\rho_{\lambda_s}^{LM-SIG} = \lim_{\lambda_s \to \infty} \frac{\log(N_m(V + \lambda_s N_c \log_2 N_m))}{\log \lambda_s} = 1 \tag{4.7}$$

$$\rho_{\alpha}^{LM-SIG} = \lim_{\alpha \to \infty} \frac{\log(N_m(V + \lambda_s N_c \log_2 N_m))}{\log \alpha} = 0. \tag{4.8}$$

[1] Standard asymptotic notation has been used. Denote by $f(n) \in \Theta(g(n))$ if there exist some positive constants c_1, c_2, and n_0 such that $c_1 g(n) \leq f(n) \leq c_2 g(n)$ for all $n \geq n_o$.

4.3.1.2 Mobile Host

In SIGMA, the total cost on the mobile host is due to the query messages exchanged with ARs, binding update messages exchanged with CNs, and location update and refreshing update messages sent to LM and packet delivery. The cost expression can be derived as follows (see Section 3.4.2 for detailed explanation):

$$\begin{aligned}\Gamma_{MH} &= \Gamma_{MH}^{DH} + \Gamma_{MH}^{BU} + \Gamma_{MH}^{LU} + \Gamma_{MH}^{RB} + \Gamma_{MH}^{DD} \\ &= \frac{2\sigma\delta_D}{T_r} + N_c \frac{6\sigma\delta_S + \beta_{AR}}{T_r} + \frac{2\sigma\delta_L}{T_r} + 2\eta_r\delta_B + N_c\left\lceil\frac{\alpha}{\kappa}\right\rceil\sigma\lambda_s(\delta_{DP} + \delta_{DA}) \\ &= \Theta(N_c(V + \alpha\lambda_s)).\end{aligned} \tag{4.9}$$

SIGMA's MSF for each MH with respect to N_m, N_c, V, λ_s and α can be computed as follows:

$$\rho_{N_m}^{MH-SIG} = \lim_{N_m \to \infty} \frac{\log(N_c(V + \alpha\lambda_s))}{\log N_m} = 0 \tag{4.10}$$

$$\rho_{N_c}^{MH-SIG} = \lim_{N_c \to \infty} \frac{\log(N_c(V + \alpha\lambda_s))}{\log N_c} = 1 \tag{4.11}$$

$$\rho_{V}^{MH-SIG} = \lim_{V \to \infty} \frac{\log(N_c(V + \alpha\lambda_s))}{\log V} = 1 \tag{4.12}$$

$$\rho_{\lambda_s}^{MH-SIG} = \lim_{\lambda_s \to \infty} \frac{\log(N_c(V + \alpha\lambda_s))}{\log \lambda_s} = 1 \tag{4.13}$$

$$\rho_{\alpha}^{MH-SIG} = \lim_{\alpha \to \infty} \frac{\log(N_c(V + \alpha\lambda_s))}{\log \alpha} = 1. \tag{4.14}$$

4.3.1.3 Complete network

The total cost on the network due to SIGMA protocol operation is as follows (see Section 3.4.3 for detailed explanation):

$$
\begin{aligned}
\Gamma_{Net} &= \Gamma_{Net}^{QR} + \Gamma_{Net}^{LU} + \Gamma_{Net}^{RB} + \Gamma_{Net}^{DH} + \Gamma_{Net}^{BU} + \Gamma_{Net}^{DD} \\
&= N_m N_c \lambda_s (2\delta_Q h_w + \psi \log N_m) + N_m \frac{2(h_w - 1 + \sigma)\delta_L + \beta_{LM}}{T_r} \\
&+ \eta_r N_m \big(2\delta_B (h_w - 1 + \sigma) + N_m \frac{2\sigma\delta_D + \beta_{AR}}{T_r} \\
&+ N_m N_c \frac{6(h_w - 1 + \sigma)\delta_s + \beta_{CN}}{T_r} + N_m N_c \lambda_s \left\lceil \frac{\alpha}{\kappa} \right\rceil (\delta_{DP} + \delta_{DA})(h_w - 1 + \sigma) \\
&= \Theta(N_m N_c \lambda_s (\alpha + V \log_2 N_m)).
\end{aligned}
\tag{4.15}
$$

Therefore, SIGMA's MSFs for the complete network with respect to N_m, N_c, V, λ_s and α are $\rho_{N_m}^{Net-SIG} = 1$, $\rho_{N_c}^{Net-SIG} = 1$, $\rho_{V}^{Net-SIG} = 1$, $\rho_{\lambda_s}^{Net-SIG} = 1$, $\rho_{\alpha}^{Net-SIG} = 1$, respectively.

4.3.2 HMIPv6

Next, we derive the mobility scalability factors of HMIPv6 entities: MAP, HA, MH and complete network.

4.3.2.1 Mobility Anchor Point

As explained in Section 3.3.1, the asymptotic cost expression for MAP can be derived from Eqn. (3.7) as follows:

$$
\begin{aligned}
\Upsilon_{MAP} &= \Upsilon_{MAP}^{LC} + \Upsilon_{MAP}^{RR} + \Upsilon_{MAP}^{RC} + \Upsilon_{MAP}^{DD} \\
&= \Theta\big(N_m N_c (V + \alpha\lambda_s \log_2 N_m)\big).
\end{aligned}
\tag{4.16}
$$

HMIPv6's MSFs for MAP with respect to N_m, N_c, V, λ_s and α are $\rho_{N_m}^{MAP-HMIP} = 1$, $\rho_{N_c}^{MAP-HMIP} = 1$, $\rho_{V}^{MAP-HMIP} = 1$, $\rho_{\lambda_s}^{MAP-HMIP} = 1$, $\rho_{\alpha}^{MAP-HMIP} = 1$, respectively.

4.3.2.2 Home Agent

As explained in Section 3.3.2, the asymptotic cost expression for MAP can be derived from Eqn. (3.13) as follows:

$$\begin{aligned}\Upsilon_{HA} &= \Upsilon_{HA}^{QR} + \Upsilon_{HA}^{RR} + \Upsilon_{HA}^{RC} + \Upsilon_{HA}^{RB} + \Upsilon_{HA}^{DD} \\ &= \Theta\big(N_m N_c (V + \alpha\lambda_s \log_2 N_m)\big).\end{aligned} \tag{4.17}$$

HMIPv6's MSFs for HA with respect to N_m, N_c, V, λ_s and α are $\rho_{N_m}^{HA-HMIP} = 1$, $\rho_{N_c}^{HA-HMIP} = 1$, $\rho_{V}^{HA-HMIP} = 1$, $\rho_{\lambda_s}^{HA-HMIP} = 1$, $\rho_{\alpha}^{HA-HMIP} = 1$, respectively.

4.3.2.3 Mobile Host

As explained in Section 3.3.3, the asymptotic cost expression for each MH of HMIPv6 can be derived from Eqn. (3.19) as follows:

$$\Upsilon_{MH} = \Upsilon_{MH}^{RC} + \Upsilon_{MH}^{LC} + \Upsilon_{MH}^{RR} + \Upsilon_{MH}^{RB} + \Upsilon_{MH}^{PD} = \Theta\Big(N_c(V + \alpha\lambda_s)\Big). \tag{4.18}$$

So HMIPv6's MSFs for each MH with respect to N_m, N_c, V, λ_s and α are $\rho_{N_m}^{MH-HMIP} = 0$, $\rho_{N_c}^{MH-HMIP} = 1$, $\rho_{V}^{MH-HMIP} = 1$, $\rho_{\lambda_s}^{MH-HMIP} = 1$, $\rho_{\alpha}^{MH-HMIP} = 1$, respectively.

4.3.2.4 Complete network

Finally, the asymptotic cost expression for HMIPv6 can be derived from Eqn. (3.26) as follows (details can be found in Section 3.3.4):

$$\begin{aligned}\Upsilon_{Net} &= \Upsilon_{Net}^{QR} + \Upsilon_{Net}^{LC} + \Upsilon_{Net}^{RR}\Upsilon_{Net}^{RB} + \Upsilon_{Net}^{RC} + \Upsilon_{Net}^{DD} \\ &= \Theta\big(N_m N_c (V + \alpha\lambda_s \log_2 N_m)\big).\end{aligned} \tag{4.19}$$

Therefore, HMIPv6's MSFs for the complete network with respect to N_m, V, and N_c λ_s and α are $\rho_{N_m}^{Net-HMIP} = 1$, $\rho_{N_c}^{Net-HMIP} = 1$, $\rho_{V}^{Net-HMIP} = 1$, $\rho_{\lambda_s}^{Net-HMIP} = 1$, $\rho_{\alpha}^{Net-HMIP} = 1$, respectively.

4.3.3 NEMO and SINEMO

We follow an approach similar to the one explained for host-mobility protocols (SIGMA and HMIPv6) to derive the asymptotical cost expressions and mobility scalability factors of NEMO and SINEMO based on the cost models presented in Chapter 3.

We perform an entity-wise scalability analysis of NEMO and SINEMO protocol. We have chosen CLM and MR of SINEMO for the entity-wise evaluation since CLM is involved in every session between a CN and a MNN, and all communications with the mobile network are carried out through the MR. On the other hand, we have chosen HA and MR of NEMO for entity-wise scalability analysis since these two entities are the key components for NEMO.

4.4 Results

We summarize our results in the following subsections. We first present results for host mobility protocols and then for network mobility protocols.

Table 4.1: Mobility scalability factors of SIGMA and HMIPv6.

Protocols	Entity	Cost Expression	$\rho^X_{N_m}$	ρ^X_V	$\rho^X_{N_c}$	$\rho^X_{\lambda_s}$	ρ^X_α
SIGMA	LM	$\Theta(N_m(V + \lambda_s N_c \log_2 N_m))$	1	1	1	1	0
	MH	$\Theta(N_c(V + \alpha\lambda_s))$	0	1	1	1	1
	Network	$\Theta(N_m N_c \lambda_s(\alpha + V \log_2 N_m))$	1	1	1	1	1
HMIPv6	MAP	$\Theta\big(N_m N_c(V + \alpha\lambda_s \log_2 N_m)\big)$	1	1	1	1	1
	HA	$\Theta\big(N_m N_c(V + \alpha\lambda_s \log_2 N_m)\big)$	1	1	1	1	1
	MH	$\Theta(V + N_c)$	0	1	1	1	1
	Network	$\Theta\big(N_m N_c(V + \alpha\lambda_s \log_2 N_m)\big)$	1	1	1	1	1

4.4.1 HMIPv6 vs. SIGMA

Table 4.1 lists the asymptotic cost expressions and mobility scalability factors for SIGMA and HMIPv6 entities. It is found that the MSFs of the overall network for these two host-mobility protocols exhibit identical values. However, there is a difference in the MSFs between the LM and HA. The MSF of LM (SIGMA protocol) with respect to α is 0 (denoted by $\rho_{\alpha}^{LM-SIG} = 0$), whereas the MSF of HA (HMIPv6 protocol) is 1, which is denoted by $\rho_{\alpha}^{Net-HMIP} = 1$ (see the rightmost column in Table 4.1). This implies that LM is more scalable than HA with respect to data rate. The significance of this result is that in HMIPv6, data packets are routed through the HA whereas in SIGMA this is not the case (data packets are not routed through the LM). Thus, the LM is not flooded with data traffic. This can improve the performance of SIGMA protocol and the LM can respond quickly to the location query, and perform other signaling jobs more efficiently than HA. On the other hand, the HA for HMIPv6 protocol can become overloaded and may not function as quickly as required to reply to location query of a CN at the beginning of each session (compared to SIGMA protocol). Moreover, the acknowledgement of the binding updates to the MH may also get delayed, resulting in increased handover latency that will affect handover process and may subsequently lead to connection termination.

4.4.2 NEMO vs. SINEMO

Table 4.2 summarizes the asymptotic cost expressions of the mobility management entities of NEMO and SINEMO. This was derived from the cost expressions of NEMO and SINEMO in Chapter 3. In Table 4.3, SINEMO and NEMO's mobility scalability factors are listed with respect to N_m, N_f, λ_p, V and N_c. It is to be noted that the HA for NEMO corresponds to SINEMO's CLM.

Results show that the mobility scalability factors of the overall network for these two network mobility protocols are identical. However, the scalability factors of CLM and HA are different. SINEMO's CLM is found to scale better than NEMO's HA with respect to packet arrival rate (λ_p). The reason for this difference is that like SIGMA, SINEMO uses optimal route for data traffic between any MNN and

Table 4.2: Asymptotic cost expressions of NEMO and SINEMO.

Protocols	Entity	Cost Expressions
NEMO	HA	$\Theta((V+\lambda_p)N_c(N_m+N_f)\log_2(N_m+N_f))$
	MR	$\Theta(N_c(N_m+N_f)(V+\lambda_p))$
	Network	$\Theta(N_c(N_m+N_f)(V+\lambda_p\log_2 N_c(N_m+N_f)))$
SINEMO	CLM	$\Theta(N_c(N_m+N_f)(V+N_c\log(N_m+N_f)))$
	MR	$\Theta(V(N_m+N_f)+\lambda_p\log_2(N_m+N_f)+VN_c\log_2 N_c))$
	Network	$\Theta((\lambda_pN_c+N_m+N_f)\log_2(N_m+N_f)+VN_c\log_2 N_c)$

Table 4.3: Mobility scalability factors of SINEMO and NEMO.

Protocols	$\rho^X_{N_m}$	$\rho^X_{N_f}$	$\rho^X_{\lambda_p}$	ρ^X_V	$\rho^X_{N_c}$	Entity
NEMO	1	1	1	1	1	HA
	1	1	1	1	1	MR
	1	1	1	1	1	Complete network
SINEMO	1	1	0	1	1	CLM
	1	1	1	1	1	MR
	1	1	1	1	1	Complete network

the CN, whereas NEMO routes data packets through the HA. Therefore, the HA does not scale with respect to increased data rate into the mobile network. As the amount of multimedia data (e.g., audio-video streaming and high resolution images) access over the mobile network increases very rapidly, the HA serving the MR of a mobile network is overloaded very quickly, resulting in its performance degradation (with higher response time and data forwarding time) that subsequently affects the quality of service of Internet access inside the mobile network.

4.5 Summary

In this chapter, we have developed a mathematical model for the quantitative scalability analysis of host and network mobility protocols with respect to network size, mobility rate, traffic rate and data volume. We have compared the scalability features of SIGMA (SINEMO) with its IETF counterpart HMIPv6 (NEMO). Our results show that the IP-mobility protocols exhibit asymptotically identical scalability pattern as far as the complete network is concerned. However, the key mobility management entities, (such as HA and LM) exhibit differences in their scalability factors. We found that decoupling the location management from data transmission process has relieved the LM (of SIGMA/SINEMO protocol) from its responsibilities, thereby making it more scalable than HMIPv6/NEMO protocols. On the other hand, the HA of HMIPv6 and NEMO protocol can become overloaded very quickly with increased data rate, resulting in its performance degradation (with higher response time and data forwarding time) that subsequently affects the quality of service of Internet access inside the mobile network. The scalability analysis framework presented in this chapter can help in quantifying and visualizing the effects of future network expansion on the performance of mobility protocols.

In the next chapter, we focus on protecting the crucial mobility signaling against losses due to the high volume of real-time and non-real data in the network.

Chapter 5

Multi-class Traffic Analysis of Mobility Protocols

The traffic characteristics in recent years have changed considerably having more video and audio traffic with critical real-time constraints. This situation is compounded by the increased number of mobile nodes communicating over Internet. Due to such high data traffic, the access routers are often overloaded with packets with different priority. The access routers must act quickly enough to avoid loss of connection of mobile nodes' ongoing communication. In this chapter, we focus on the multi-class traffic analysis for mobility protocols. Our objective is to protect the all-important signaling traffic against losses due to the high volume of real-time and non-real time data in the network. We propose a scheduling algorithm that gives the highest priority to signaling traffic, thereby ensuring its minimum loss while dropping some real-time traffic as they are loss-tolerant.

5.1 Introduction

In each IP-mobility protocol, mobile nodes communicate with the access router through the wireless channel for sending data packets and signaling packets, such as binding updates. Due to the proliferation of mobile computing and extensive use of wireless devices (accessing the Internet), the number of mobile nodes under an access router is rising rapidly, resulting in the increase of total packet arrival rates.

Mobile nodes send signaling packets that need to be processed very quickly by the access routers to notify the mobility agent, such as location manager or home agent to keep the location database up-to-date. On the other hand, for the traffic class having time critical deadlines (e.g., real-time audio-video data), packets must be delivered to the destination before the deadlines. Otherwise, they are useless and can be dropped by the router.

There have been several research works on multi-class traffic that considered mobility. Nam et al. [24] presented a two-layer downlink queuing model and a scheduling mechanism for providing lossless handoff and Quality of Service (QoS) in mobile networks. Iftikhar et al. [25] presented an analytical model for multiple queue systems by considering two different classes of traffic that exhibit long-range dependence and self-similarity. Iftikhar et al. [26] used a G/M/1 queuing system and developed analytical models based on non-preemptive priority queuing, low-latency queuing and custom queuing systems. However, these works lack the quantitative analysis of average queuing delay, queue occupancy and packet drop probability of data and signaling packets at the access router while managing IP-mobility. None of the existing work focused on the dynamic scheduling of signaling packets to protect them from dropping due to the high volume of real-time and non-real time data in the wireless networks.

In this work, we have considered two classes of data traffic: Real Time (RT) and Non-Real Time (NRT) traffic. Another class of traffic is the signaling traffic (binding update) sent by the mobile nodes to update mobility agents about their current locations. We have proposed a scheduling algorithm that selects a packet based on its priority and the current location of the mobile node within a cell so that it produces the least loss to the signaling traffic and handoff traffic, while allowing packet loss to real-time traffic since RT traffic is loss-tolerant. Based on the scheduling algorithm, we have derived expressions for average queuing delay, queue occupancy, and packet drop probability of each class of traffic. We have presented results that reflect the impact of node density, service rate and traffic pattern on these measures.

Our goal of the work presented in this chapter is to analyze the performance of access router in processing real-time, non-real-time and signaling traffic while managing IP-mobility simultaneously.

The contributions of the work presented in this chapter are: (i) developing a mathematical model to estimate the average queuing delay, average queue occupancy and the packet drop probability at the AR of IP-mobility protocol and (ii) analyzing the impact of node density, service rate and traffic distribution on them.

The analytical model developed in this chapter can be used to better manage the access routers and other network components for improved performance of future wireless and mobile networks.

The rest of the chapter is organized as follows. In Section 5.2, the cell topology of the access routers are explained along with the computation of the overlapped area of a cell. The analytical model is presented in Section 5.3. In Section 5.4, we present our simulation results. Section 5.5 has the concluding remarks.

5.2 Cell Topology and Overlapping Area

The cell topology covered by access routers is shown in Fig. 5.1, where each cell has overlapping areas with four other neighboring cells. In this Section, we compute the overlapping area of a cell that will be used in queuing analysis (Section 5.3.5).

Let the radio coverage area of each cell be a circular region of radius r and let two adjacent cells overlap at a maximum length of l along its diameter. In Fig. 5.1, let $AC = a = BC$, and $DE = l$. Since the length of AB is $2r$, we find that $a = \sqrt{2}r = r + r - l$. Therefore,

$$l = (2 - \sqrt{2})r. \tag{5.1}$$

In order to find out the overlapping area, let us consider Fig. 5.2, where the center

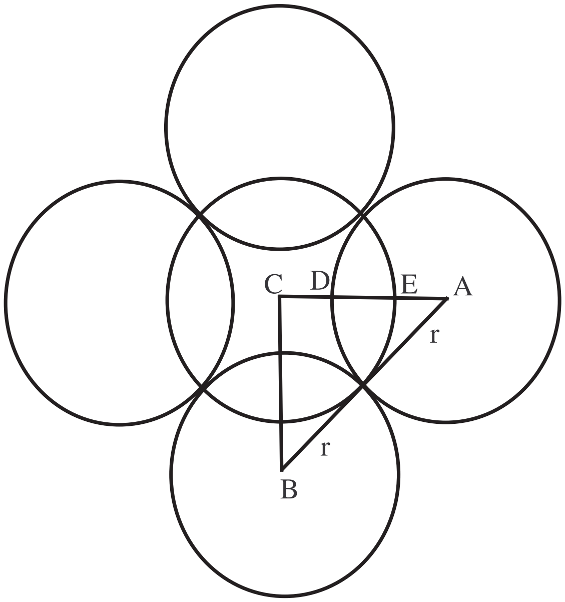


Figure 5.1: Cell topology.

(C) of a cell is situated at the origin. The cord MN bisects the line segment DE at point F. Therefore, $DF = EF = l/2$. Since point C is the origin, $CE = r$ and $CF = CE - EF = r - l/2 = r/\sqrt{2}$. As point F is on the x-axis, the coordinate of point F is $(\frac{r}{\sqrt{2}}, 0)$. Therefore, the area of the region CFMP (denoted by φ) is the area under the circle $x^2 + y^2 = r^2$ from the origin to point F and can be obtained by integrating between 0 to $\frac{r}{\sqrt{2}}$ as follows:

$$\varphi = \int_0^{\frac{r}{\sqrt{2}}} y dx = \int_0^{\frac{r}{\sqrt{2}}} \sqrt{(r^2 - x^2)} dx. \tag{5.2}$$

Now, putting $x = r\sin\theta$, the limit becomes 0 to $\pi/4$. Thus, we get

$$\varphi = r^2 \int_0^{\pi/4} \cos^2\theta d\theta = \frac{r^2}{4}\Big(\frac{\pi}{2} + 1\Big). \tag{5.3}$$

As the area of the cell is πr^2, the portion of the cell in the first quadrant (i.e., the region CPME) is $\pi r^2/4$. So the area of the shaded region (MFE) can be obtained by subtracting Eqn. (5.3) from $\pi r^2/4$:

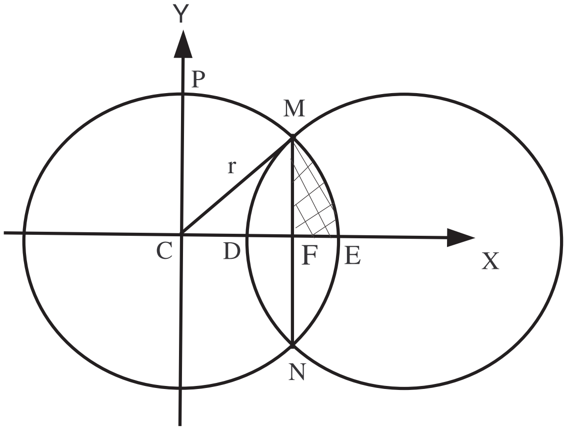


Figure 5.2: Overlapping area among the cells.

$$\text{Area of MFE} = \frac{\pi r^2}{4} - \frac{r^2}{4}\Big(\frac{\pi}{2} + 1\Big) = \frac{r^2}{8}\big(\pi - 2\big). \tag{5.4}$$

The overlapping area between two adjacent cells (MDNE), which is four times the shaded area (of MFE), equals $\frac{r^2}{2}\big(\pi - 2\big)$. For each cell, there are four neighboring cells with which it has overlapping areas (see Fig. 5.1). Hence, the total overlapping area of a cell (ϕ) is given by,

$$\phi = 4 \times \frac{r^2}{2}\big(\pi - 2\big) = 2r^2(\pi - 2). \tag{5.5}$$

The ratio of the overlapping area (Eqn. (5.5)) to the total area (πr^2) is given by

$$\gamma = \frac{2r^2(\pi - 2)}{\pi r^2} = \frac{2(\pi - 2)}{\pi}. \tag{5.6}$$

From Eqn. (5.6), we find that $\gamma = 0.7268$, which means that 72.68% area of the cell is in overlapping area whereas 27.32% area is non-overlapping.

5.3 Analytical model

First, the assumptions and notations of the model are listed in Sections 5.3.1 and 5.3.2. Section 5.3.3 explains the node architecture of the AR, followed by the scheduling algorithm in Section 5.3.4. Section 5.3.5 presents the queuing analysis.

5.3.1 Assumptions

To make the model analytically tractable, the following assumptions are made.

- Uniform density of the Mobile Nodes (MN) in the network is assumed.
- The number of MNs leaving the coverage area of an AR is equal to the MNs entering the area. Therefore, the net change in number of MNs under an AR is zero.
- The ratio of the overlapping area to the total coverage area is used as a measure of probability for being in handoff status.
- Packet arrival process is a Poisson process.
- Type of queue discipline used in the analysis is FIFO with non-preemptive priority among various traffic classes.
- One-way (download) traffic is considered for simplicity.
- As Binding Update (BU) messages exchanged between MN and Home Agent are essential to track a mobile node, BU packets are assigned the highest priority.

5.3.2 Notations

The notations used in the analysis are listed below.

N_m Number of MNs in a cell (AR),

λ_{BU} Binding update packet rate at each MN,

λ_{RBU} Refreshing binding update packet rate at each MN,

λ_{RT} Real time packet arrival rate at each MN,

λ_{NRT} Non-real time packet arrival rate at each MN,

λ_i Packet arrival rate at class-i queue of AR,

λ Total packet arrival rate to the system,

μ Service rate of the system,

ρ System utilization factor,

r Radius of each cell,

γ Ratio of overlapping area to the total area of the cell,

$E(T_i)$ Average queuing delay of a class-i packet,

$E(n_i)$ Average queue occupancy of class-i packets,

$P_d(i)$ Packet drop probability of class-i packets.

5.3.3 Node architecture of the AR

We have considered three classes of traffic: RT and NRT data traffic, and signaling traffic (BU and refreshing BU packets) exchanged between mobile nodes and mobility agents for location management. Depending on the location of the MNs, there can be two statuses of each packet: handoff (H) and non-handoff. *Non-handoff status* is when the MN is inside the coverage area of the AR only, whereas *handoff status* when it is moving towards some neighboring cell and about to handoff.

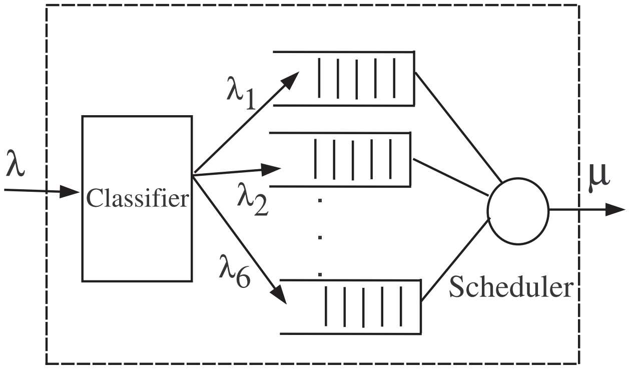


Figure 5.3: Node architecture of the AR.

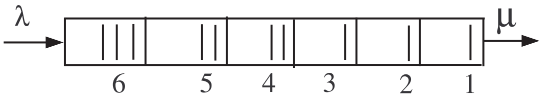


Figure 5.4: Multi-class packets in the queue of the AR.

While sending data packet in the overlapped coverage region, the MN may set the flag handoff flag to indicate that it is about to handoff to some new AR. Other (data) traffic and RBU packets are assumed to have a non-handoff status.

Figure 5.3 shows the queuing architecture at the access router. Packets of all types arrive at the AR at a rate λ. The classifier categorizes the packets into six classes according to the traffic and handoff status. The packets are then queued appropriately. The scheduler follows the scheduling algorithm (Section 5.3.4) to select the next packet to transmit. Thus, packets of higher priority can be considered to have been queued ahead of other lower priority packet as shown in Fig. 5.4.

5.3.4 Scheduling algorithm

The steps followed in the scheduling algorithm are listed below:

- If the arrived packet is a BU packet, it is scheduled immediately after the packets in the BU queues if there is any.

- When the data packets (RT or NRT) with handoff flag = 1 arrive at the AR, they are selected with a higher priority than data traffic with non-handoff status. In addition, RT traffic gets higher priority than NRT traffic.

- While serving RT traffic, check the queue length of NRT traffic. If the NRT queue grows beyond minThreshold (usually three-fourth of queue size), allow dropping of RT traffic of different flows (if dropping is allowed).

- While serving RT traffic, if the NRT queue grows beyond maxThreshold (equals queue size), the transmission of RT traffic is suspended, and NRT traffic is transmitted until the NRT queue length comes below minThreshold.

As BU/RBU traffic has the highest priority, the queue associated with them will not grow much and with a reasonable size of the queue (in compliance with the node density) will ensure the least loss of binding updates. As RT traffic (usually UDP packets) is loss-tolerant and useless if delivered after a long delay, the queue length of the RT traffic should be kept small. In case of overflow in RT queue, the (UDP) traffic will be lost which does not harm the RT communication. As NRT traffic (usually TCP traffic) is not loss-tolerant, NRT queue size is kept larger than RT queue. As NRT traffic is assigned the lowest priority, NRT packets (usually TCP) may be lost, resulting in the retransmission of the same packets.

5.3.5 Queueing analysis

Based on the proposed scheduling algorithm, we have computed the average queuing delay, queue occupancy and packet drop probability for BU packets, RT and NRT traffic through wireless media. We have used non-preemptive queuing, which means lower priority packets in service will not be preempted by higher priority packets.

Let N_m be the number of mobile nodes under the radio coverage area of the AR. Among them, the fraction of MNs that are in the overlapped region is γN_m and these MNs will send data to the AR setting the handoff flag. Without loss of generality, we can assume that γ fraction of the data packets will have handoff (H) status and the rest will have non-handoff (NH) status. The probability that an MN will be in the handoff mode is given by γ (see Eqn. (5.6) in Section 5.2).

5.3.5.1 Packet arrival rates

In the overlapping area, binding updates are sent by MNs. As the BU packet rate at each MN is λ_{BU}, BU packet arrival rate at the AR from the $N_m\gamma$ MNs will be given by

$$\lambda_1 = \lambda_{BU} \times N_m\gamma. \tag{5.7}$$

Similarly, as the RBU packet rate at each MN is λ_{RBU}, RBU packet arrival rate at the AR from the $N_m(1-\gamma)$ MNs will be given by

$$\lambda_2 = \lambda_{RBU} \times N_m(1-\gamma). \tag{5.8}$$

For RT traffic, as the packet arrival rate to each MN is λ_{RT}, the packet arrival rate at AR with handoff status is

$$\lambda_3 = \lambda_{RT} \times N_m\gamma. \tag{5.9}$$

Similarly, the RT packet arrival rate at AR with non-handoff status is given by

$$\lambda_4 = \lambda_{RT} \times N_m(1-\gamma). \tag{5.10}$$

For NRT traffic, as the packet arrival rate to each MN is λ_{NRT}, the packet arrival rate at AR with handoff status is

$$\lambda_5 = \lambda_{NRT} \times N_m\gamma. \tag{5.11}$$

Finally, the NRT packet arrival rate at AR with non-handoff status is given by

$$\lambda_6 = \lambda_{NRT} \times N_m(1-\gamma). \tag{5.12}$$

Since the packet arrival rate to each queue is a Poisson process, the total arrival rate of all classes of packets to the system will collectively be a Poisson process with rate $\lambda = \lambda_1 + \lambda_2 + \ldots + \lambda_6$. Let the service time of the system be exponentially distributed with the mean $1/\mu$. Let $\rho_i = \lambda_i/\mu$. Then, the system utilization factor can be computed as

$$\rho = \frac{\lambda}{\mu} = \rho_1 + \rho_2 + \ldots + \rho_6 = \sum_{i=1}^{6} \rho_i. \tag{5.13}$$

5.3.5.2 Mean packet delay and queue length

When a class-1 packet arrives, the system may be in the process of serving a packet of any other class. The probability that a class-1 packet finds a class-2 packet in service equals the fraction of time the server spends on class-2 packets which is $\lambda_2/\mu = \rho_2$. Thus, class-1 packet finds any of the class-i packets in service is ρ_i, where $2 \leq i \leq 6$.

According to Little's theorem [27], the average number of packets waiting in the system is equal to the average delay times the average arrival rate of the system. This theorem can be used in some part of the system. Let us apply this theorem to the queue of class-i. As the average packet delay for class-1 packets is $E(T_1)$ and the average queue occupancy is $E(n_1)$ with arrival rate λ_1, we have

$$E(n_1) = \lambda_1 E(T_1). \tag{5.14}$$

Again, when a packet of class-1 arrives at the system, there are, on the average, $E(n_1)$ packets on class-1 queue. Since class-1 packets have the highest priority, the mean time delay of the packet that has just arrived depends on $E(n_1)$ packets which are already buffered in the class-1 queue plus the packet in service. As $E(n_1)$ includes the class-1 packet in service (if any), the mean time delay of class-1 packet in the system is the queuing delay and the service time $(1/\mu)$. Therefore,

$$\begin{aligned} E(T_1) &= \frac{E(n_1)}{\mu} + \frac{1}{\mu} + \frac{1}{\mu}(\rho_2 + \rho_3 + ... + \rho_6) \\ &= \frac{E(n_1)}{\mu} + \frac{1}{\mu} + \frac{1}{\mu}\sum_{i=2}^{6} \rho_i. \end{aligned} \tag{5.15}$$

Now substituting Eqn. (5.14) into the Eqn. (5.15) and after simplification, we get

$$E(T_1) = \frac{\left(1 + \sum_{i=2}^{6} \rho_i\right)}{(1 - \rho_1)\mu} = \frac{1}{\mu} + \frac{\rho/\mu}{1 - \rho_1}. \tag{5.16}$$

Hence, Eqn. (5.14) can be rewritten as follows:

$$E(n_1) = \frac{\left(1 + \sum_{i=2}^{6} \rho_i\right)\rho_1}{1 - \rho_1} = \rho_1 + \frac{\rho\rho_1}{1 - \rho_1}. \tag{5.17}$$

While considering the delay of class-2 packets, we have to consider the average number of class-1 packets ($E(n_1)$) in the system as they have the higher priority of service. So the class-2 packet which arrives in the system must wait for all the class-1 packets buffered in the class-1 queue and the class-2 queue. Thus, the average queuing delay of class-2 packets is given by

$$\begin{aligned} E(T_2) &= \frac{E(n_1)}{\mu} + \frac{E(n_2)}{\mu} + \frac{1}{\mu} + \frac{1}{\mu}(\rho_3 + \rho_4 + ... + \rho_6) \\ &= \frac{E(n_1)}{\mu} + \frac{E(n_2)}{\mu} + \frac{1}{\mu} + \frac{1}{\mu}(\rho - \rho_1 - \rho_2). \end{aligned} \tag{5.18}$$

Using Little's theorem for class-2 queue, that is, $E(n_2) = \lambda_2 E(T_2)$ in Eqn. (5.18) and after simplification, we get

$$E(T_2) = \frac{1}{(1 - \rho_2)\mu}\left(1 + \frac{\rho}{1 - \rho_1} - \rho_2\right). \tag{5.19}$$

Hence, the average queue length of class-2 queue is given by,

$$E(n_2) = \frac{\rho_2}{(1 - \rho_2)}\left(1 + \frac{\rho}{1 - \rho_1} - \rho_2\right). \tag{5.20}$$

Therefore, the general expression for packets of class-i can be derived as,

$$\begin{aligned} E(T_i) &= \frac{E(n_1)}{\mu} + \frac{E(n_2)}{\mu} + \ldots + \frac{E(n_i)}{\mu} + \frac{1}{\mu} + \frac{1}{\mu}(\rho_{i+1} + \rho_{i+2} + \ldots + \rho_k) \\ &= \frac{1}{\mu}\left(\sum_{j=1}^{i} E(n_j) + \sum_{j=i+1}^{k} \rho_j + 1\right). \end{aligned} \tag{5.21}$$

Again, using Little's theorem for class-i queue and simplifying, we get

$$E(T_i) = \frac{1}{\mu(1-\rho_i)}\left(\sum_{j=1}^{i-1} E(n_j) + 1 + \rho - \sum_{j=1}^{i} \rho_j\right). \tag{5.22}$$

Therefore, the average queue length of class-i queue is given by,

$$E(n_i) = \frac{\rho_i}{(1-\rho_i)}\left(\sum_{j=1}^{i-1} E(n_j) + 1 + \rho - \sum_{j=1}^{i} \rho_j\right). \tag{5.23}$$

5.3.5.3 Packet drop probability

Let us assume that N_i denotes the size of the ith queue. Therefore, the packet drop probability at each queue is the probability of the queue being full. Thus the packet drop probability at the ith queue can be obtained as follows:

$$P_d(i) = \frac{\rho_i^{N_i}(1-\rho_i)}{1-\rho_i^{N_i+1}}. \tag{5.24}$$

5.4 Results

In this section, simulation results are presented to show the effect of node density, service rate and traffic distribution on the average queuing delay, queue occupancy and packet drop probability at the AR. We have used the ns-2 network simulator [28] and have implemented the multi-class scheduling algorithm for the ARs.

5.4.1 Simulation Environment

Figure 5.5 shows the topology used for the simulation. The simulation area is a 10 km × 10 km area which is covered by 196 access routers (arranged in 14 rows and 14 columns) with overlapping area similar to that shown in Fig. 5.1. The hierarchical addresses of all the nodes are listed in Table 5.1. The CNs act as the FTP (over TCP) or CBR (over UDP) sources whereas the mobile nodes are the corresponding sinks, thereby simulating non-real-time and real-time communications, respectively.

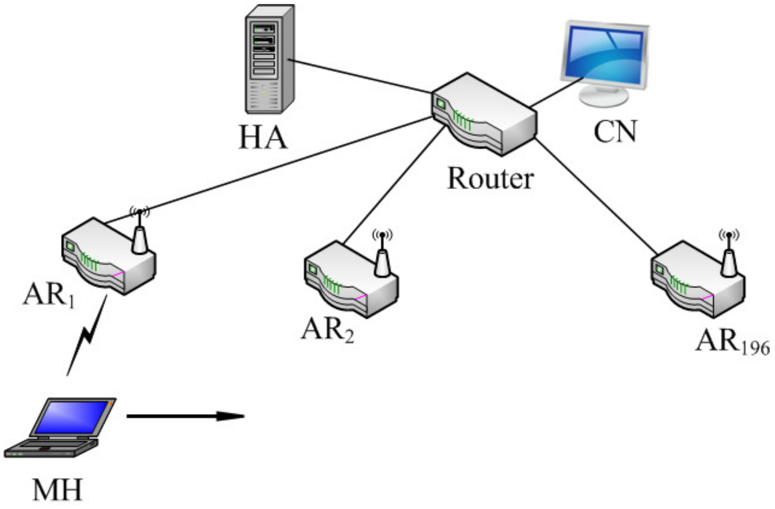


Figure 5.5: Simulation topology.

Table 5.1: Hierarchical addresses used for entities in the simulation.

Node Type	**Hierarchical addresses**
CN$_i$	0.0.i
HA	1.1.0
Router	1.0.0
AR$_1$ - AR$_{196}$	1.1.0 - 1.196.0
MH$_i$	1.1.i

Each mobile node moves according to the mobility model explained in Chapter 10 starting from a random location.

The default values of parameters used for simulations are listed in Table 5.2. Each AR's transmission range is 500 m. IEEE 802.11g standard is used for wireless communications. The wireless link bandwidth is 54 Mbps and the wired link bandwidth is 100 Mbps. The RT and NRT packet size was set to 512 bytes whereas the size of BU/RBU packets is 64 bytes. The percentage of users accessing RT and NRT traffic was set to be equal, that is, 50% for both types.

The results obtained from the analysis of the simulation traces are presented in the following subsections. We have measured the average queuing delay, queue

Table 5.2: Values of parameters used in the simulation.

Simulation Parameters	**Values**
Simulation area	10 km × 10 km
Wireless range	500 m
Wired link BW	100 Mbps
Wireless link BW	54 Mbps
Binding lifetime	30 sec
Number of Access Routers	196
Data packet size	512 bytes
Size of BU packets	64 bytes
Service rate of each AR	30000 pkts/sec

occupancy and packet drop probability at the AR. We have denoted the class-1 through class-6 as BU, RBU, RT-H, RT-NH, NRT-H and NRT-NH (in the graphs of this section) for clear understanding.

5.4.2 Impact of node density

In Fig. 5.6, the average queuing delays at the six queues are shown for varying numbers of MNs in a cell (covered by an AR). Our results show that the queuing delay of NRT traffic rises for higher node density whereas the queuing delay for signaling traffic remains unchanged ensuring faster delivery of signaling traffic. This shows that the increased number of MNs in the cell area does not affect the average delay of the signaling traffic. This implies that the signaling traffic is served by the AR quickly, thereby preventing disruption of connectivity for the mobile devices which is the main objective of the scheduling algorithm.

In Fig. 5.7, the average queue occupancy at the six queues is shown for different node density at the cell region. It is found that the average queue occupancy is almost zero for signaling traffic as they are served very quickly (due to having highest priority) by the AR. The NRT traffic in handoff status has the highest queue

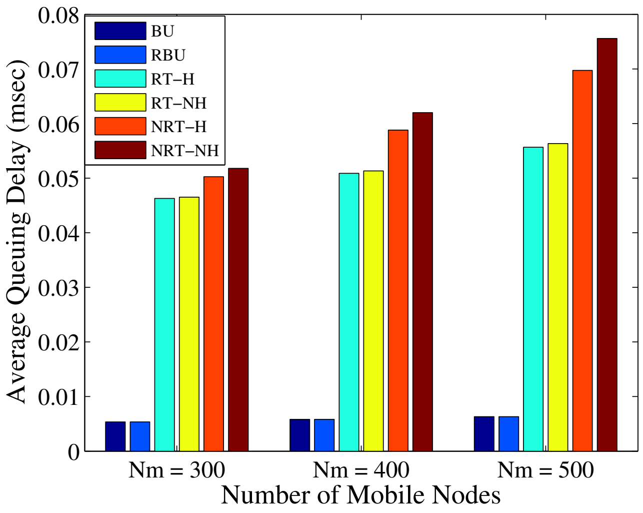


Figure 5.6: Average queuing delay at the six queues for different number of MNs.

occupancy as the arrival rate is the highest unlike the RT traffic which is served with a higher priority to meet the real-time deadlines.

In Fig. 5.8, the packet drop probabilities for the six classes of packets are shown for different node density at the cell region. The probability increases for higher node density for RT and NRT traffic. However, the packet drop probability for signaling traffic is zero, meaning that the scheduling algorithm protects all signaling packets from getting lost. The RT class has the smaller queue size resulting in more packet drop. In addition, data traffic (RT and NRT) with handoff status has the higher arrival rates, causing more loss of such traffic. Our scheduling algorithm protects signaling traffic (BU) in lieu of sacrificing some RT and NRT traffic (with handoff status). The lost RT traffic may not be required to be retransmitted. However, the lost NRT traffic triggers the adjustment of congestion window between the sender and the receiver of the traffic and can be recovered through retransmissions.

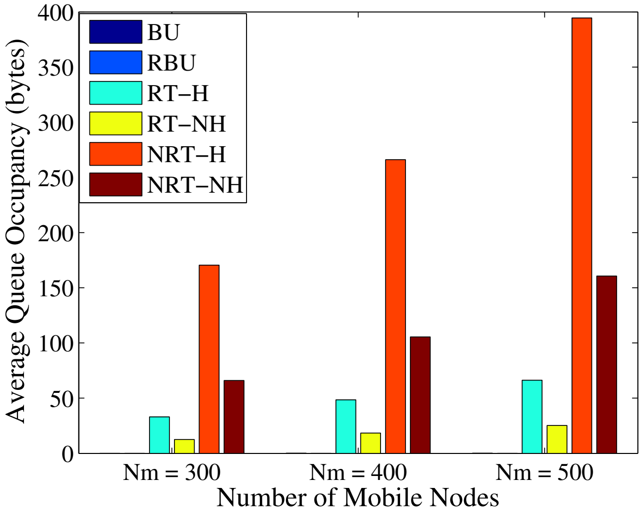


Figure 5.7: Average queue occupancy at the six queues for different number of MNs.

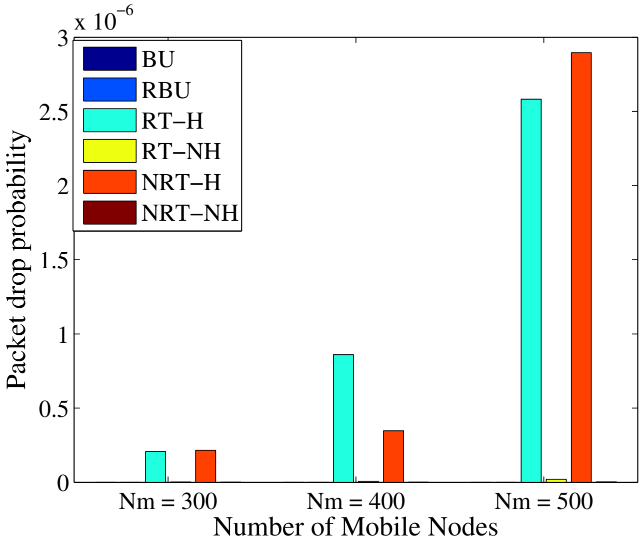


Figure 5.8: Packet drop probability at the six queues for different number of MNs.

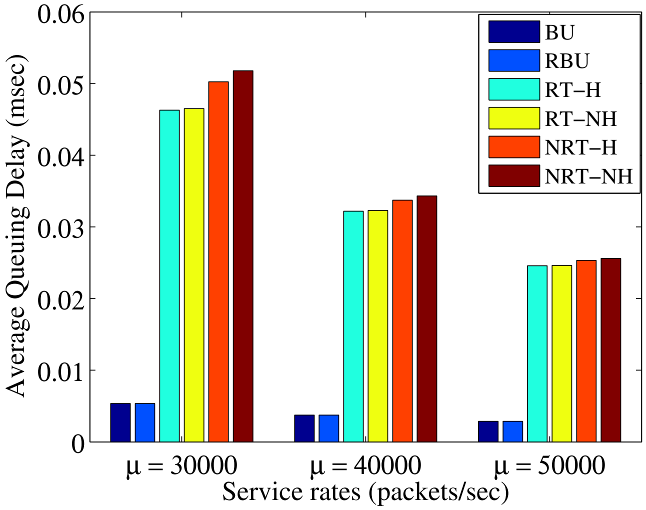


Figure 5.9: Average queuing delay at the six queues for different service rates.

5.4.3 Impact of service rate

In Fig. 5.9, the average queuing delays are shown for different service rates of the AR. The queuing delay decreases for higher service rates. Again, the delay for signaling traffic is the least whereas that of NRT traffic is the highest. Thus, in order to reduce the delays of the all traffic classes, increase of service rates is important and today improvement of wireless access technology is facilitating such higher service rates of the AR.

In Fig. 5.10, the average queue occupancy for the six classes of traffic is shown for different service rates of the AR. Results show that the average occupancy at the queue decreases for higher service rates. For RT traffic, it is much less than the NRT traffic due to the use of higher priority and the smaller queue size in our proposed scheduling system.

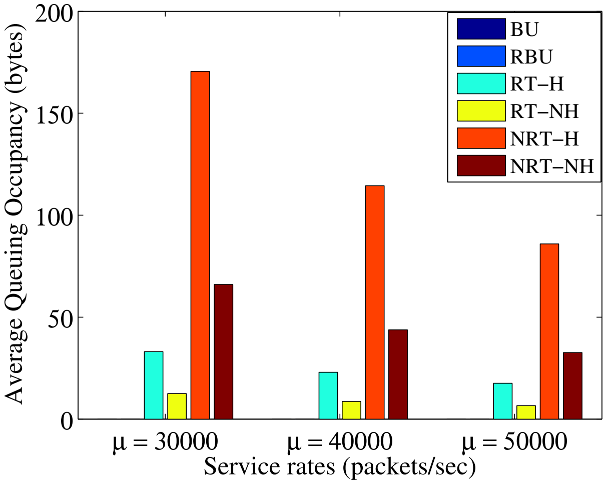


Figure 5.10: Average queuing occupancy at the six queues for different service rates.

5.4.4 Impact of traffic distribution

Figure 5.11 shows the queuing delay for different classes of traffic varying the percentage of users accessing RT and NRT traffic. In case 1, MNs accessing RT and NRT traffic are assumed to be 5% and 95%, respectively. For cases 2 and 3, the distributions are 25%-75% and 50%-50%, respectively. It is found that as the RT traffic increases, the queuing delay for data traffic increases. This is because the increased RT arrival forces NRT to be delayed/dropped. However, the delay of the signaling traffic remains almost unchanged for all the cases since they are served with higher priority, ignoring the high RT and NRT traffic arrival.

5.5 Summary

In this chapter, we have proposed a scheduling algorithm that selects packets based on the priority and current location of the mobile nodes, ensuring the least loss to signaling traffic, while dropping some loss-tolerant real-time traffic. Based on the scheduling algorithm, we have derived closed-form expressions for average queuing

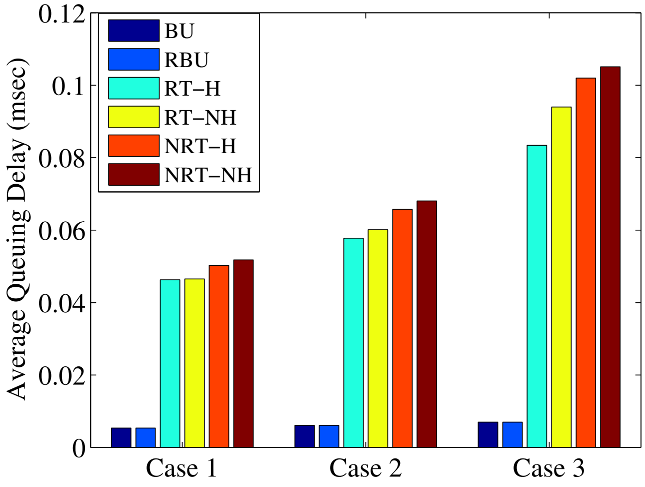


Figure 5.11: Average queuing delay at the queues for different traffic.

delay, queue occupancy and packet drop probability of each class of traffic. We have presented simulation results showing the impact of node density, service rate and traffic distribution on those measures. The scheduling algorithm proposed in this chapter will be helpful to better manage access routers with different classes of data and signaling traffic, causing the least queuing delay and packet loss for the all-important signaling traffic while ensuring seamless Internet connectivity for mobile devices. In the next chapter, we focus on maximizing utilization of available bandwidth for mobility protocols exploiting the band-sharing approach among different traffic classes.

Chapter 6

Multi-band Architecture for Mobility Protocols

Recent trend in Internet usage has seen large amounts of multimedia traffic accessed through large numbers of mobile devices. To facilitate higher bandwidth, modern mobile routers are now capable of supporting simultaneous multi-band frequencies, incurring less interference (and loss) while raising the total capacity and reliability of wireless communications. However, there exists no previous work that attempts to maximize utilization of available bandwidth through the sharing of traffic classes among different frequency bands of the mobile router. In the previous chapter, we have focused on the multi-class traffic analysis in the single band and tried to protect the all-important signaling traffic against losses due to the high volume of real-time and non-real-time data in the network. In this chapter, we propose a novel scheduling algorithm for multi-band mobile routers which transmit different classes of traffic through different frequency bands to achieve an improved performance while ensuring maximum possible utilization. The objective in this chapter is to maximize the utilization of multi-band mobile routers through band sharing.

6.1 Introduction

In recent years, there has been an explosive growth of mobile users accessing large multimedia files (such as high definition audio, video, images, etc.) over the Internet.

Therefore, the bandwidth demand for mobile Internet access is increasing exponentially [29]. To satisfy such a higher bandwidth requirement, today wireless routers are available commercially with simultaneous multi-band support of 2.4 and 5 GHz. Future IEEE 802.11ad (WiGig) tri-band-enabled devices, operating in the 2.4, 5 and 60 GHz bands, are expected to deliver data transfer rates up to 7 Gbps [30]. The benefit of using a multi-band router is to have less interference, higher capacity and better reliability. Exploitation of rarely-used frequency bands in wireless networks reduces interference in heavily-used frequency band, e.g., 2.4 GHz, thereby increasing the total capacity of the wireless network.

Current simultaneous multi-band routers make use of two different bands (2.4 GHz and 5GHz) for different types of devices in a home network. However, they do not attempt to exploit the under-utilized frequency band when one of them is overloaded with data. Moreover, there are different classes of traffic, and some of the traffic types (such as, real-time) have strict delay constraints associated with it while the signaling traffic (required for mobility management) is crucial for maintaining Internet connectivity for the nodes in motion. Therefore, it is essential to propose an appropriate scheduling and queue management scheme for the multi-class traffic to ensure the maximum possible utilization of the system resources in multi-band mobile routers [31]. The aim of the work presented in this chapter is to propose methods for such maximal utilization of multi-band mobile routers through band sharing.

There have been several research works reported in the literature [29–35] that attempt to extend current single band technology through the use of multiple frequency bands, leading to increased bandwidth while reducing interference. Even though multi-band usage has been widely investigated in cell networks [33, 34], it is a relatively new concept in wireless networks. Verma and Lee [35] explained a possible Wi-Fi architecture with multiple physical and link layers to support multiple frequency bands simultaneously. Singh et al. [29] proposed a method to assign to end-devices different bands based on their distances from the access router. In [30, 32], the authors proposed the use of 60 GHz frequency band (having low range) to attain faster data transfer rate in wireless networks. However, none of these works [29–35] proposed any scheduling algorithm for multi-band system considering multi-class traffic, neither do they perform any queueing analysis to measure various performance metrics of the multi-band system.

To the best our knowledge, there has been no earlier work on scheduling and queue management for multi-band mobile routers that attempts to maximize utilization of available bands. Moreover, no previous work exists that proposed the sharing of multiple bands to transmit different classes of traffic. This is a novel work that aims at attaining maximum possible band utilization and proposes a scheduling algorithm which exploits band sharing.

The objective of the work presented in this chapter is to analyze the performance of multi-band mobile routers while ensuring maximum possible utilization through sharing of bands among different classes of traffic.

The contributions of the work presented in this chapter are: (i) proposing a band-sharing router architecture and a novel scheduling algorithm to ensure the maximum possible utilization of the system, (ii) developing an analytical model to evaluate the performance (utilization of bands, average class occupancy, packet drop rate, average delay, and throughput) of the proposed multi-band system, and (iii) validating our analytical model through extensive simulations.

Our proposed algorithm considers the multi-class Internet traffic and schedules them through alternate under-utilized frequency bands, thereby reducing packet loss and delay of the system.

Results show that the packet drop rate and throughput are significantly improved in the proposed band-sharing architecture of the mobile router. Moreover, the simulation results validate our analytical model.

Our proposed scheme and related analysis will help network engineers build next generation mobile routers with higher throughput and utilization, ensuring the least packet loss of different classes of traffic.

The rest of the chapter is organized as follows. In Section 6.2, we explain the typical architecture of multi-band mobile routers, followed by the proposed architecture in Section 6.3. Section 6.4 presents the analytical model to derive different performance metrics of the proposed architecture. In Section 6.5, we present the simulation results that validate our analytical model; we also compare the performance of the proposed and typical architectures. Section 6.6 has the concluding remarks.

6.2 Typical Multi-band Router Architecture

Commercial (simultaneous) multi-band mobile routers available today make use of two different bands (2.4 GHz and 5GHz) for different types of devices in a home network. Laptops may connect to 2.4 GHz network while WiFi-enabled TV, gaming devices may connect to 5 GHz network. This reduces interference with the heavily-used 2.4 GHz network (as cordless phones, microwave oven use similar band). In addition, video streaming can be done through the high frequency band. Future IEEE 802.11ad (WiGig) tri-band-enabled devices, operating in 2.4, 5 and 60 GHz bands, are expected to deliver data at a much higher rate (up to 7 Gbps) [30].

The main principle of today's simultaneous multi-band MR is the non-sharing of bands among different flows of traffic. Moreover, some of the devices available today (such as IPTV) mostly deals with real-time traffic. Based on this fact, we have assumed that each of the band of typical simultaneous multi-band MR only deals with one type of traffic. This might be a slight deviation from the real MR

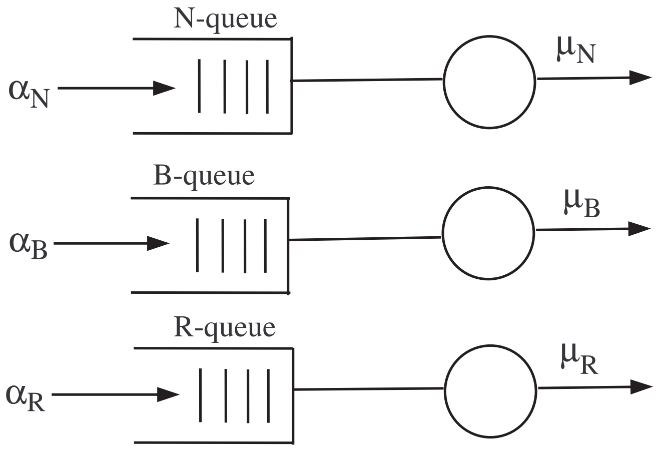


Figure 6.1: Architecture of a typical (simultaneous) multi-band mobile router.

used today. However, we have assumed this to compare our proposed architecture with typical simultaneous multi-band MR.

Figure 6.1 shows the typical architecture of a simultaneous tri-band MR. Here, three bands are assumed to be used for three different classes of traffic: signaling traffic or Binding Update (BU), real-time (RT) and non-real-time (NRT) traffic. Each class of traffic is solely assigned to each designated frequency band as shown in Fig. 6.1, and we name the corresponding queues as B-queue, R-queue and N-queue. There will be absolutely no sharing of traffic among different bands even if a band is under-utilized due to low traffic arrival to that queue while the others are full.

6.3 Proposed Multi-band Router Architecture

In this section, we explain the proposed architecture of multi-band MRs that promotes sharing of bands in order to maximize the overall system utilization. We have considered three different queues (shown in Fig. 6.2), each of which corresponds to a frequency band of a simultaneous tri-band Mobile Router.

As shown in Fig. 6.1, we consider three classes of traffic and each queue is designated for each class of traffic. However, unlike the typical architecture, in this

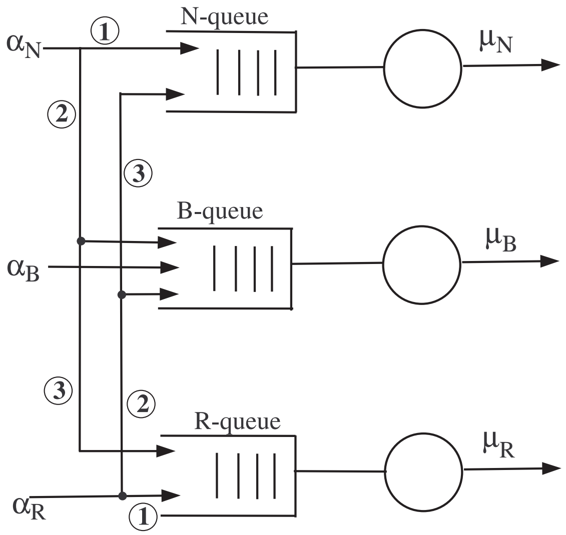


Figure 6.2: Proposed architecture of a simultaneous multi-band mobile router.

proposed architecture (see Fig. 6.2) the traffic of one class can flow through other queues, provided the other queues have empty slots, thereby ensuring better utilization of buffer spaces available. For example, if the B-queue has some empty spaces available and a bursty RT traffic comes in, the overflowed RT traffic can be queued in the B-queue and subsequently served (or sent) through the B-server (transmitter).

6.3.1 Time and space priority

The time and space priority for the three queues of the proposed architecture are explained in Figs. 6.3, 6.4 and 6.5. For B-queue, BU packets have the highest priority; RT and NRT packets have dynamic priority based on arrival rates (see Eqns.(6.4) and (6.5)). Regarding space priority, BU packets are queued in front of B-queue and if there are empty spaces available, other types (RT and NRT) can be accommodated as shown in Fig. 6.3.

R-queue can have only RT and NRT packets as shown in Fig. 6.4. RT traffic has higher priority over NRT traffic. Therefore, R-queue can have NRT packets

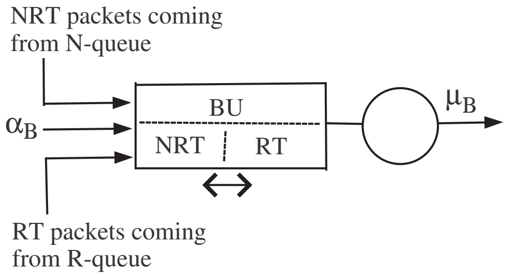


Figure 6.3: Queue corresponding to BU band.

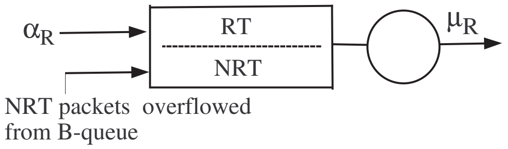


Figure 6.4: Queue corresponding to RT band.

only if RT packets cannot fill the R-queue at any time and there are NRT packets overflowed from the B-queue.

Finally, Fig. 6.5 shows the N-queue which is designated for NRT traffic. However, if there are empty spaces available in this queue, overflowed RT traffic out of B-queue can be enqueued in N-queue (see Fig. 6.2).

6.3.2 Scheduling algorithm

We have considered the following two crucial factors to ensure improved performance of the multi-band MR:

- The unused buffer space of one queue (or band) can be used for other traffic types, thereby reducing the system idle time.

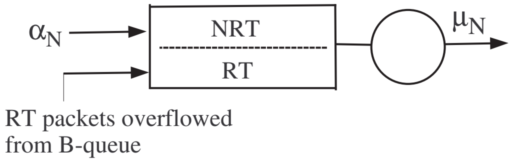


Figure 6.5: Queue corresponding to NRT band.

- Priorities of different traffic classes are considered while selecting a particular type of packet over others. Priority has an inverse relationship with the arrival rate of the corresponding traffic class.

The proposed scheduling policies are explained as follows:

- Attempts are first made to queue different class of traffic in their corresponding buffer.
- If N-queue (or R-queue) overflows, the corresponding traffic is forwarded to the B-queue.
- If B-queue does not have enough empty slots to handle the overflowed NRT and RT packets, they race for slots in B-queue based on priority (see Eqns. (6.4) and (6.5).
- If overflowed RT packets cannot be accommodated in B-queue, they are queued in N-queue, if space is available in the queue.
- If the R-type packets cannot even be accommodated in N-queue, they are dropped from the system.
- Similar policy is enforced when dealing with NRT packets in the B-queue and then in the R-queue.

6.4 Analytical Model

In this section, we present our analytical model to derive various performance metrics of the proposed multi-band mobile router architecture. First, the assumptions and the notations of the model are listed in Sections 6.4.1 and 6.4.2, followed by the analytical models in the subsequent subsections.

6.4.1 Assumptions

To make the model analytically tractable, the following assumptions have been made.

- Packet arrival follows a Poisson distribution [10, 12].
- Queue discipline is FIFO with non-preemptive priority among various traffic classes.

6.4.2 Notations

The notations used in the analysis are listed below. To simplify our notation, we use T as the common notation for different traffic class types, and $T \in$ { B, N, R }.

N_T Queue size of T-queue in the MR,

α_T Total packet arrival rate at T-queue of i-th MN,

μ_T Service rate at T-queue of i-th MN,

$\sigma_{T_{BQ}}$ Priority of class-T traffic in B-queue,

$P_{dT_{XQ}}$ T-type packet drop probability in X-queue, where $X \in$ {B, N, R}

$E(D_T)$ Average queuing delay of class T packets,

$E(n_T)$ Average queue occupancy of class T packets,

P_{dT}^{sys} Final packet drop probability of class T packets.

6.4.3 Total arrival rates in each queue

For queuing analysis of the proposed system, we need to determine the total arrival rate of all classes of traffic in each queue. In general, overflow in a queue can happen when the arrival rate is larger than the buffer space or when the service rate is smaller than the arrival rate and there is no buffer space left, as specified by the following conditions:

$$\alpha_T > N_T \quad or \quad \mu_T < \alpha_T. \tag{6.1}$$

6.4.3.1 B-queue

Packets overflowed from N-queue and R-queue go to B-queue. Thus, the arrival rates of N-type and R-type packets to the B-queue (denoted by $\alpha_N^{'}$ and $\alpha_R^{'}$) can be obtained as follows:

$$\alpha_N^{'} = \alpha_N P_{dN_{NQ}}. \tag{6.2}$$

$$\alpha_R^{'} = \alpha_R P_{dR_{RQ}}. \tag{6.3}$$

where $P_{dN_{NQ}}$ and $P_{dR_{RQ}}$ are packet drop probabilities of N-type packets in N-queue and R-type packets in R-queue.

Priorities of different classes are taken into account while allowing traffic into B-queue. Priority of B packets in B-queue is $\sigma_{B_{BQ}} = 1$. Priorities of other classes of traffic in B-queue are measured as follows:

$$\sigma_{N_{BQ}} = \frac{\alpha_R}{\alpha_B + \alpha_R + \alpha_N}. \tag{6.4}$$

$$\sigma_{R_{BQ}} = \frac{\alpha_N}{\alpha_B + \alpha_R + \alpha_N}. \tag{6.5}$$

Now, the total (effective) arrival rate of all classes of traffic in B-queue can be obtained as follows:

$$\alpha_{B(Total)} = \alpha_B + \alpha_N^{'} + \alpha_R^{'}. \tag{6.6}$$

6.4.3.2 N-Queue

The N-queue is designated for NRT traffic. However, if there are empty buffer spaces available in the N-queue (due to low NRT arrival rate), this queue can be used to transmit the RT traffic that is overflowed from the B-queue. Let $\alpha_R^{''}$ denote the arrival rate of RT packets in N-queue. Then, the total arrival rate (of both N and R-type packets) in the N-queue is:

$$\alpha_{N(Total)} = \alpha_N + \alpha_R^{''}. \tag{6.7}$$

6.4.3.3 R-Queue

The R-queue is designated for RT traffic. However, if there is an empty buffer space available in the R-queue (due to low RT arrival rate), this queue can be used to transmit NRT traffic which has been overflowed from the B-queue. Let $\alpha_N^{''}$ denote the arrival rate of N-type packet going to R-queue. Then, the total arrival rate (of both R and N-type packets) in the R-queue is:

$$\alpha_{R(Total)} = \alpha_R + \alpha_N^{''}. \tag{6.8}$$

6.4.4 Packet drop probability

The packet drop probability of R-type packets in R-queue can be obtained using standard M/M/1/N formula as follows [36]:

$$P_{dR_{RQ}} = \frac{\rho_R^{N_R}(1-\rho_R)}{1-\rho_R^{N_R+1}}. \tag{6.9}$$

where $\rho_R = \frac{\alpha_R}{\mu_R}$. Similarly, the packet drop probability of N-type packets in N-queue can be obtained as follows:

$$P_{dN_{NQ}} = \frac{\rho_N^{N_N}(1-\rho_N)}{1-\rho_N^{N_N+1}}. \tag{6.10}$$

where $\rho_N = \frac{\alpha_N}{\mu_N}$. Let us assume that the priority of R-type packets is higher than that of N-type packets in B-queue. Therefore, while computing the R-type packet drop probability in B-queue, we can safely consider only B-type and R-type packets in B-queue. Let us define utilization in B-queue as $\rho_{BR} = \frac{\alpha_B + \alpha_R^{'}}{\mu_B}$ while considering only B-type and R-type packets in the B-queue. Thus, the packet drop probability of B-type packet in B-queue, denoted by $P_{dB_{BQ}}$, can be obtained as follows [37]:

$$P_{dB_{BQ}} = \frac{\rho_{BR}\rho_1^{N_B}(1-\rho_1)(1-\rho_{BR}^{N_B+1})}{(1-\rho_1^{N_B+1})(1-\rho_{BR}^{N_B+2})}. \tag{6.11}$$

where $\rho_1 = \alpha_B/\mu_B$ and $\rho_2 = \alpha_R^{'}/\mu_B$. Using Eqn. (6.11), the packet drop probability of R-packets in B-queue can obtained as follows [37]:

$$P_{dR_{BQ}} = \frac{(1-\rho_{BR})}{\left(1-\rho_{BR}^{N_B+2}\right)}\rho_{BR}^{N_B+1} + \frac{\alpha_B}{\alpha_R^{'}}\left(\frac{(1-\rho_{BR})}{\left(1-\rho_{BR}^{N_B+2}\right)}\rho_{BR}^{N_B+1} - P_{dB_{BQ}}\right). \tag{6.12}$$

Hence, the RT packet arrival in N-queue can be obtained as follows:

$$\alpha_R^{''} = \alpha_R^{'} P_{dR_{BQ}}. \tag{6.13}$$

The total arrival of N-queue is the sum of two arrival rates α_N and $\alpha_R^{''}$ (see Eqn. (6.7)); the former has the higher priority than the latter. Therefore, following a similar approach as in Eqn. (6.11), we can compute $P_{dN_{NQ}}$. Then we can follow a similar approach in Eqn. (6.12) to compute $P_{dR_{NQ}}$ which is the final drop of R-type packets from the system. That is, $P_{dR}^{sys} = P_{dR_{NQ}}$. The computation of N-type packet drop probability follows similar steps as followed for R-type packets. Therefore, $P_{dN}^{sys} = P_{dN_{RQ}}$.

6.4.5 Average queue length

Each queue behaves as an M/M/1/N queue. Therefore, the estimated queue length can be obtained as follows:

$$E(n_T) = \begin{cases} \frac{\rho_T - (N_T+1)\rho_T^{N_T+1} + N_T\rho_T^{(N_T+2)}}{\left(1-\rho_T\right)\left(1-\rho_T^{N_T+1}\right)} & \text{if } \rho_T \neq 1 \\ \frac{N_T}{2} & \text{if } \rho_T = 1. \end{cases} \tag{6.14}$$

The average queue occupancy of R-type packets depends on the queue occupancy of R-packets in R-queue, B-queue and N-queue. This is computed as follows:

$$
\begin{aligned}
E(n_R^{sys}) &= E(n_R^{RQ}) + E(n_R^{BQ}) + E(n_R^{NQ}) \\
&= E(n_R^{RQ}) + \Big(E(n_{B+R}^{BQ}) - E(n_B^{BQ})\Big) + \Big(E(n_{N+R}^{NQ}) - E(n_N^{NQ})\Big). \qquad (6.15)
\end{aligned}
$$

To compute $E(n_R^{RQ})$, we need to put $N_T = N_R$, $\rho_T = \rho_R = \alpha_R/\mu_R$ in Eqn. (6.14). A similar approach can be used for the rest of the terms in Eqn. (6.15).
Since B-type packets are only queued in B-queue, the average queue occupancy of the B-type packets in the system can be obtained as follows:

$$E(n_B^{sys}) = E(n_B^{BQ}). \qquad (6.16)$$

To compute the average queue occupancy of N-type packets in the system, a similar approach as in Eqn. (6.15) can be used:

$$E(n_N^{sys}) = E(n_N^{NQ}) + E(n_N^{BQ}) + E(n_N^{RQ}). \qquad (6.17)$$

6.4.6 Throughput

The throughput of the T class of traffic can be obtained as follows:

$$\gamma_T^{sys} = \big(1 - P_{dT}^{sys}\big)\alpha_T. \qquad (6.18)$$

6.4.7 Average packet delay

The average packet delay of T class is given by,

$$E(D_T^{sys}) = \frac{E\big(n_T^{sys}\big)}{\big(1 - P_{dT}^{sys}\big)\alpha_T}. \qquad (6.19)$$

6.5 Results

We have used a discrete event simulation in MATLAB by taking into account the assumptions and scheduling policies mentioned in Sections 6.2 and 6.3. We have followed M/M/3/N procedures [36] for the implementation of simulation programs. We have kept equal buffer length (50 packets) for each queue. Buffer lengths are kept small [38] similar to real routers to decrease packet delay. RT and NRT packets are assumed to be 512 bytes [30,39] whereas the BU packets are assumed to be 64 bytes. The service rates of the B, N and R-queues are kept 27, 75 and 132 packets/sec which is proportional to service rates of multi-band routers [30]. We ran each simulation for 20 trials having different traffic class arrival rates as follows:

$\lambda_B(i) = \{ \text{ i } \}$, $\lambda_N(i) = \{ \text{ 3i } \}$, $\lambda_R(i) = \{ \text{ 18i } \}$

where i = 1, 2, 3, ..., 20. We have run simulations with increased arrival rates of all types of traffic to observe the impact of heavy traffic on the multi-band system. The arrival rates of B-queue and N-queue are increased slowly in each trial whereas the RT traffic arrival rate is increased at a much higher rate. This eventually saturates the R-queue and we explain the impact of this overflow on different performance parameters of our proposed system and typical existing system.

6.5.1 Validation of analytical model

In this subsection, we show the simulation results for proposed architecture and compare them with those produced by the analytical expressions derived in Section 6.4 to validate our analytical model.

6.5.1.1 Average class occupancy

Figure 6.6 shows the average class occupancy of the proposed multi-band system obtained through simulations and analytical model. The simulation and analytical results are very close to each other. The class occupancy of N-class and B-class are very low as their service rate is higher than arrival rates. However, this is not the case for R-class where the arrival rate is higher than R-queue service rate; hence, the excessive RT packets are enqueued in other two queues, thereby increasing the average occupancy of R-class.

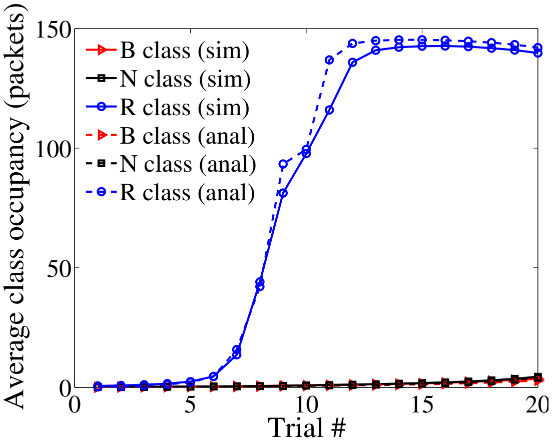


Figure 6.6: Average class occupancy of proposed architecture obtained through simulations and analytical model.

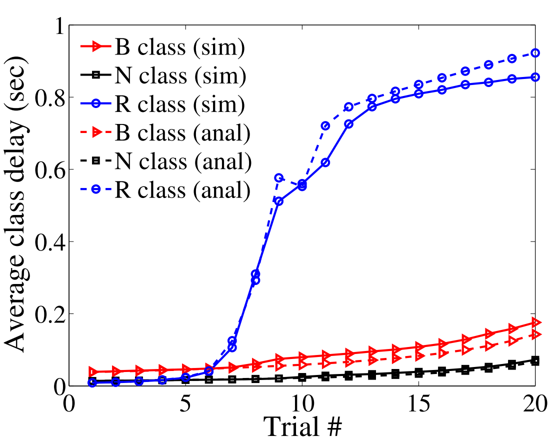


Figure 6.7: Average class delay of proposed architecture obtained through simulations and analytical model.

6.5.1.2 Average class delay

Figure 6.7 shows the average class delay for the proposed multi-band system obtained through simulations and analytical model. As we can see the simulation results are close to the analytical one (for all three traffic types) though these exists little differences between them because of discrete time simulation environment. Delays for B, N and R classes depend on occupancy. Therefore, Fig. 6.7 follows the similar patterns as in Fig. 6.6 with different rates.

6.5.1.3 Class drop rate

Figure 6.8 shows the class drop rate for the proposed multi-band system. Again the simulation results closely match the analytical one. Drop rates of B-class and N-class are low due to low arrival rate. However, the R-class drop rate rises as we use very high arrival rates in subsequent trials.

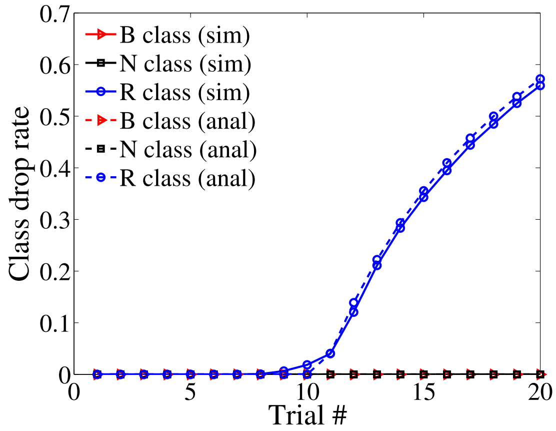


Figure 6.8: Class drop rate of proposed architecture obtained through simulations and analytical model.

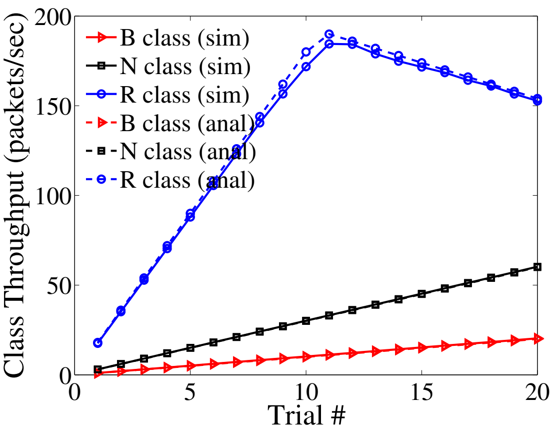


Figure 6.9: Class throughput of proposed architecture obtained through simulations and analytical model.

6.5.1.4 Class throughput

Figure 6.9 shows the class throughput for the proposed multi-band system. Again the simulation result closely matches with the analytical one.

Thus, it is evident from Figs. 6.6, 6.7 6.8, and 6.9 that the analytical and simulation results are very close to each other, thereby validating the analytical model.

6.5.2 Proposed architecture vs. typical architecture

In this subsection, we present results comparing the performance results of our proposed (band-sharing) architecture and typical (non-sharing) architecture.

6.5.2.1 Utilization

Utilization is a performance measure that indicates how efficiently bands are used and whether there is any unused capacity of the system. Figure 6.10 shows band utilizations for proposed (shared) and typical (non-shared) architectures of multi-band routers. When packet arrival rates are low (in trials #1 through 7) compared to the capacity of each queue, all the queues have somewhat similar and low utilization for both the architectures. However, for trials #8 through 20, the utilizations of B-queue and N-queue are much higher for proposed architecture than for typical one. This is because increased number of RT packets are dropped in typical architecture (see Fig. 6.13) whereas in proposed one, they are accommodated in B-queue and N-queue, thereby improving their utilizations and maximizing system performance.

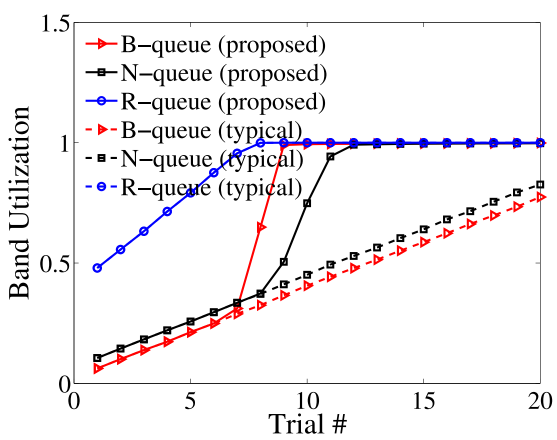


Figure 6.10: Band utilizations for proposed and typical architectures for different simulation trials.

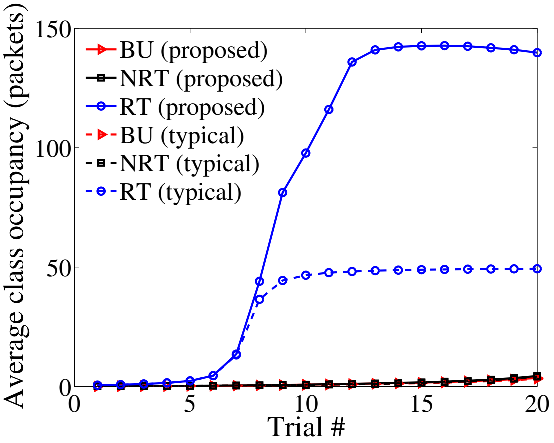


Figure 6.11: Average class occupancy for proposed and typical architectures for different simulation trials.

6.5.2.2 Average class occupancy

Figure 6.11 shows the average queue occupancies of each class of traffic (in the system) for proposed and typical architecture. Average class occupancy for B-queue and N-queue are similar for both architectures as we have used low arrival rates (for BU and NRT) compared to the service rates. However, the RT arrival rates are increased significantly (in trials # 7 through 20) compared to R-queue service rate. As a result, the class occupancy for RT class is much higher for the proposed architecture than the typical one. This is because a large amount of RT traffic is transmitted though the other frequency bands as they are under-utilized having empty buffer space.

6.5.2.3 Average packet delay

Figure 6.12 shows the average packet delay of each class of traffic for proposed and typical architecture. The delay for RT traffic (for trails # 8 through 20) in proposed architecture is higher than the typical one. This is because excessive RT packets are immediately dropped from the system in typical architecture and these lost packets do not come into account in delay calculations. On the contrary, in proposed architecture overflowed RT packets get chances to be enqueued in B and N-queue before being dropped. RT packets are second priority packets in N-queue and B-queue and they have to wait for NRT and BU packets, respectively before being scheduled for service. Hence, it increases the delay of RT packets.

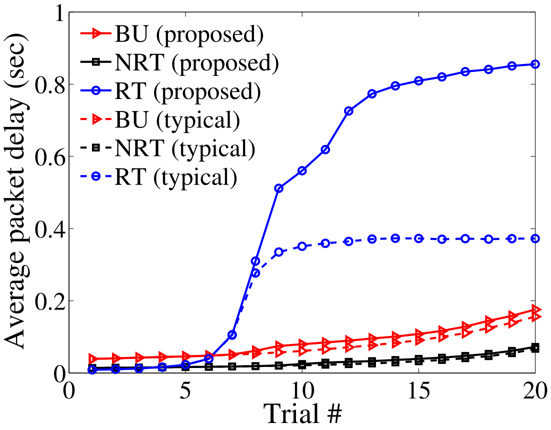


Figure 6.12: Average packet delay for proposed and typical architectures for different simulation trials.

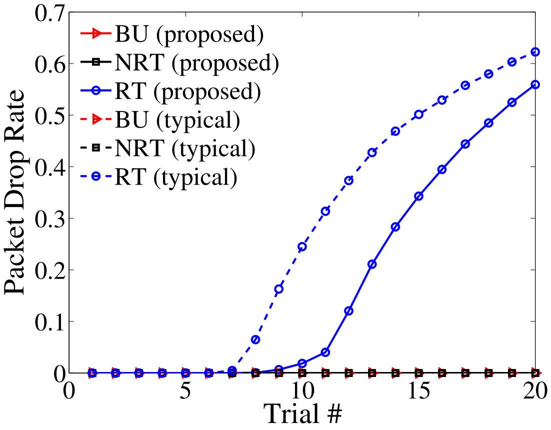


Figure 6.13: Average packet delay for proposed and typical architectures for different simulation trials.

6.5.2.4 Packet drop probability

Figure 6.13 shows packet drop probability of each class for proposed and typical architecture. For both the approaches, the drop rate of BU and NRT are low and similar since the arrival rates are lower than the service rates. When the arrival rates for RT increases (in trials # 7 through 20), RT packet drop rate gradually goes up for typical (non-shared) architecture. However, the proposed architecture does not allow RT traffic to drop as long as the excessive RT packets can be queued in the empty buffer spaces at B-queue and N-queue. Therefore, the RT packet drop rate is much lower for proposed architecture than the typical one.

6.5.2.5 Throughput

Figure 6.14 shows the throughput of each class for proposed and typical architectures. The throughput of NRT and BU class is increased with the increase of their arrival rates for both architectures. However, in case of RT class and for the typical architecture, the throughput is saturated at μ_R (= 132 pkts/sec) when the RT arrival rate reaches this value. However, the RT class throughput for the proposed architecture goes much higher (due to sharing of other under-utilized bands) and reaches its peak value in trial #11. After that it starts to decrease slowly due to the impact of increased arrival rates of other queues (B and N-queue) that result in a less available space for overflowed RT packets.

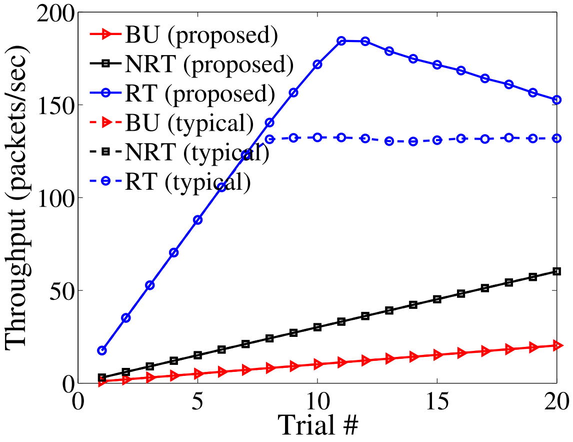


Figure 6.14: Throughput of proposed and typical architectures for different simulation trials.

6.5.3 Discussion on results

Our simulation results closely match the analytical ones (see Figs. 6.6, 6.7, 6.8, and 6.9), thereby validating our analytical model. Moreover, we find that the proposed architecture attempts to maximize the system utilization (Fig. 6.10) through band sharing. This affects the average queue occupancy and delay of RT traffic (Fig. 6.11 and 6.12). However, the packet drop and throughput (Figs. 6.13 and 6.14) are significantly improved in the proposed band-sharing architecture.

6.6 Summary

In this chapter, we have proposed a novel scheduling algorithm for multi-band mobile routers that exploits a band sharing approach to attain the maximum possible utilization. We have developed an analytical model to a perform queuing analysis of the proposed multi-band system and derived various performance metrics that have been validated through extensive simulations. Our results show that the proposed architecture can ensure the maximum possible utilization through the sharing of capacities among the bands. Our proposed scheduling algorithm and the related analytical model can help network engineers build next generation mobile routers with higher throughput and utilization ensuring the least packet loss for different classes of traffic.

In the next chapter, we focus on the quantitative survivability evaluation of mobile networks.

Chapter 7

Survivability Analysis of NEMO

Survivability is a crucial aspect for any kind communication. A survivable network has the ability to withstand malicious attacks and continue to work properly even in the presence of natural or man-made disturbances. Wireless and mobile networks have the challenge of survivability, since users are mobile and the communication channels are accessible to anyone. The mobile router acts as the gateway for all the nodes inside the mobile network and it is the key entity in NEMO. Since a single MR can be single point of failure, increase in number of MRs can improve the reliability of NEMO. However, there has been no survivability analysis of multiple MR-based NEMO. In this chapter, we have performed quantitative survivability analysis of NEMO with multiple MRs, taking into consideration possible node and link failures along with Denial of Service attacks.

7.1 Introduction

Network survivability is a crucial aspect for any kind communication. A survivable network has the ability to withstand malicious attacks and continue to work properly even in the presence of natural or man-made disturbances. It focuses on delivery of essential services and rapid recovery of full services when situation improves [40, 41]. Wireless and mobile networks have the challenge of survivability, since users are mobile and the communication channels are accessible to anyone. Hence, it is essential to analyze the survivability of NEMO.

The mobile network can have one or more mobile routers that act as the gateways for all the nodes inside the mobile network known as Mobile Network Nodes (MNN). These MRs connect the MNNs to the global Internet, forwarding signaling traffic required for mobility management as well as data traffic to the desired Internet hosts. As the MR is the key entity in NEMO, the load on the MR can be very high as the MR has to deal with every ongoing session between any MNN and any Internet node. In addition, the MR sends signaling messages to the Home Agent whenever the mobile network changes its point of attachment. Therefore, the MR can become the performance bottleneck for the mobile network. Hence, increase of number of MRs can improve the performance and enhance reliability of the network as a single MR in NEMO can be single point of failure. Therefore, we are interested in the survivability analysis of NEMO.

Earlier attempts [41–46] focused on the redundancy and load balancing of the mobility agents, such as, home agent to improve the survivability mobility protocol. These works did not focus on the survivability of mobile router. However, Kuntz et al. [47] proposed that the cooperation of multiple mobile routers in NEMO can improve the bandwidth, network coverage and reliability as well as incorporate dynamic load sharing among the MRs. The solution was based on Neighbor Discovery and was validated by a real testbed. Our work aims at quantitative survivability of mobile routers to improve the performance of NEMO.

There have been a few works on network survivability evaluation. Chen et al. [40] used a Continuous Time Markov Chain (CTMC) model to evaluate the end-to-end availability of wireless ad hoc networks. Heegaard et al. [48] developed an analytical model to assess the survivability of a network with virtual connections exposed to link or node failures; the model has been validated by simulations. Fu et al. [49] performed the survivability analysis of SIGMA and Mobile IP which are based on multiple location managers for mobility management. However, the author is not aware of any survivability evaluation of NEMO that considers various failure types and denial of service attacks which can drastically degrade the performance of the mobile network. This work differs from previous work in this respect and is believed to be the first such work.

The objective of the work presented in this chapter is to perform quantitative survivability evaluation of NEMO with multiple MRs, taking into consideration different failure types as well as malicious attack traffic. We have used the CTMC as the analytical tool to develop the survivability model.

The contributions of the work presented in this chapter are: (i) developing a survivability model for NEMO to compute the packet drop probability and mean packet delay through CTMC modeling, and (ii) presenting numerical results showing performance and delay metrics while under attacks or failures.

Our results reveal interesting relationship among network performance, failure rates and attack strengths. Our developed model can be used by the network engineers to evaluate survivability and robustness of their networks.

The rest of the chapter is organized as follows. In Section 7.2 the survivability model is presented. Section 7.3 presents the numerical results. Section 7.4 has the concluding remarks.

7.2 Survivability Analysis

Figure 7.1 shows the architecture of NEMO with multiple MRs that has been considered for the survivability evaluation. The survivability evaluation of the multihomed NEMO aims at providing improved availability of the nodes inside the mobile network and allowing load-sharing among the routers.

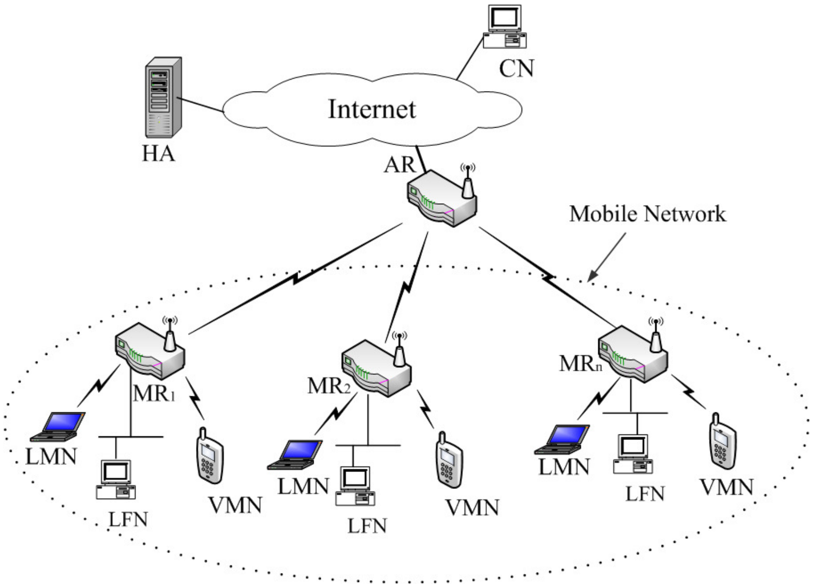


Figure 7.1: NEMO with multiple mobile routers.

7.2.1 Assumptions

Following assumptions have been made to make our model tractable:

- Total packet arrival to the NEMO is assumed to be λ which includes data packets, signaling packets and Distributed Denial of Service (DDoS) attack packets. Arrival of all such packets are Poisson processes.

- There is no way to distinguish between legitimate data (or signaling) packets and the DDoS attack packets.

- Processing time for data, signaling and DDoS packets are exponentially distributed, having the same mean value (μ). This assumption is made to limit the complexity of the CTMC model.

- We are assuming that MRs act in such a way that load on each MR is shared among them similar to the approach specified in [47].

- Only one MR (with highest load) may fail at a time .

- The failure of one MR does not halt the whole mobile network.
- Packets in the buffer of a failed MR are assumed to have been dropped. If at any time there are R routers and B packets in the system, then due to a router failure $L_1 = \lceil \frac{B}{R} \rceil$ packets in the failed router's buffer will be lost.
- Link failure may happen for any router in the mobile network. If at any time there are R routers and B packets in the system, the link failure will only reduce the number of packets in the system by $L_2 = \lfloor \frac{B}{R} \rfloor$ and those L_2 packets are considered to remain in the router buffer and will add to the total count when this link is up again at a recovery rate of δ_2.

7.2.2 Notations

The notations used in this chapter are listed as follows:

N_r Total number of MRs in the mobile network,

R_i Number of available MRs in state i,

B_i Number of packets in state i,

S Queue size (in number of packets) of each MR,

λ_d arrival rate of data packets,

λ_s arrival rate of signaling packets,

λ_a arrival rate of DDoS attack packets,

λ Total arrival rate, i.e., $\lambda = \lambda_d + \lambda_s + \lambda_a$,

μ Processing rate,

γ_1 Node failure rate,

γ_2 Link failure rate,

δ_1 Repair rate from node failure,

δ_2 Repair rate from link failure,

ψ_1 Irreversible node failure rate (when node failure cannot be repaired),

ψ_2 Irreversible link failure rate (when a link failure cannot be repaired),

β Rate of introducing a error-free MR to the system.

7.2.3 Survivability model with multiple MRs

We use a CTMC model for the survivability analysis of NEMO with multiple MRs. There are N_r MRs in the mobile network each of which has a buffer of size S. Errors can happen in two ways: node failure or link failure. Node failure can happen due to hardware failure of the MR or battery power failure. Link failure can happen when a MR loses connection with the Access Router of the home/foreign network.

Figure 7.2 shows the state transition diagram of the CTMC model. Each state of the model is labeled as (I, J, K) where I represents number of active (or available) routers in the mobile network, J represents total number of packets in the system, and K represents type of failure, where $I \in \{0,\ 1,\ 2,\ ..,\ N_r\}$, $J \in \{0, 1, 2, .., SN_r$ $\}$, and $K \in \{0,\ 1,\ 2\}$ (0 for no failure, 1 for node failure, 2 for link failure). The states with $K = 1$, and $K = 2$ are intermediate states and these (node and link) failure may or may be resolved, thereby leading to the states with $K = 0$ (i.e., no error states). In the case of recovery (involving transition rates of δ_1 or δ_2), the next state will have the same number of MRs. In case of no recovery scenarios (involving transition rates of ψ_1 or ψ_2), the next state will have one less MRs than the current state.

Figure 7.2 shows the transition diagram using a representative state of $(R, B, 0)$ which means that there are R active MRs with a total number of B packets in the

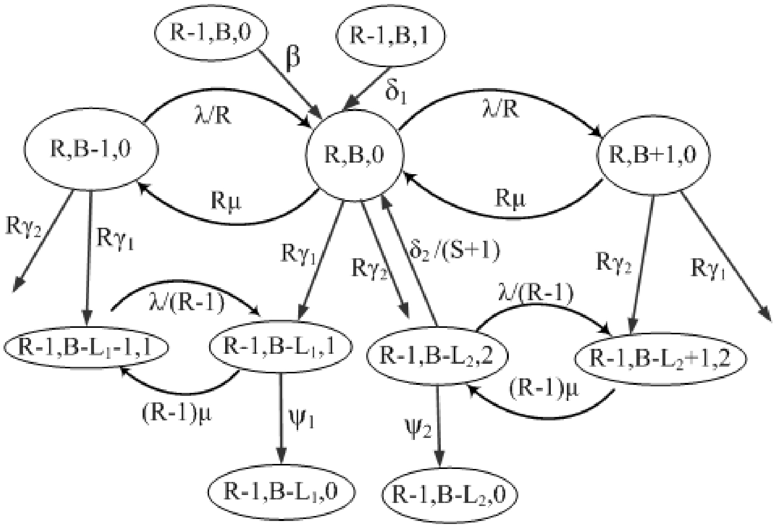


Figure 7.2: State transition diagram of NEMO with multiple mobile routers.

system in a ”no failure” state. The possible states that can be reached from this state are:

- Arrival of packets: This event will lead to the state $(R, B+1, 0)$ with a transition rate of λ/R.
- Departure of a packet: This will lead to the state $(R, B-1, 0)$ with a transition rate of $R\mu$.
- Node failure: This event happens for the MR with heaviest load, having $L_1 = \lceil \frac{B}{R} \rceil$ packets which will be lost as a result of the failure. This leads to the state $(R-1, B-L_1, 1)$ with a rate of $R\gamma_1$.
- Link failure: This can happen for any of the active MRs having $L_2 = \lfloor \frac{B}{X} \rfloor$ packets in its queue. The event leads to the state $(R-1, B-L_2, 2)$ with a rate of $R\gamma_2$.

The following events from the following states can lead to the state $(R, B, 0)$:

- Departure of a packet from the state $(R, B+1, 0)$.

- Arrival of a packet from the state $(R, B-1, 0)$.
- Recovery of node failure from the state $(R-1, B, 1)$.
- Recovery of link failure from the state $(R-1, B-L_2, 2)$.
- From the state $(R-1, B, 0)$ and due to the introduction an error-free MR.

When $R_i < N_r$, number of states is $3(R_iS+1)$; for $R_i = N_r$, it is SN_r+1. Thus, the number of states with R_i MRs can be expressed as follows:

$$f_s(R_i) = \begin{cases} 3(R_iS+1), & \text{when} \quad R_i < N_r \\ \\ SR_i + 1, & \text{when} \quad R_i = N_r \end{cases} \tag{7.1}$$

Therefore, the size n of the generator matrix Q, can be obtained as follows:

$$\begin{aligned} n &= |K| \sum_{i=0}^{N_r-1} (iS+1) + (SN_r+1) \\ &= 3\Big(S \sum_{i=0}^{N_r-1} i + \sum_{i=0}^{N_r-1} 1\Big) + (SN_r+1) \\ &= \frac{SN_r}{2}(3N_r - 1) + (3N_r + 1). \end{aligned} \tag{7.2}$$

We number the states $(0,0,0)$, $(0,0,1)$, and $(0,0,2)$ as states 1, 2 and 3, respectively. Hence, the states $(1,0,0)$, $(1,0,1)$, $(1,0,2)$, and so on are numbered as states 4, 5, 6, and so on.

For the CTMC shown in Fig. 7.2, we can determine each element of the generator matrix $Q = [q_{i,j}]$ $(0 \leq i, j \leq n)$ as follows:

$$q_{i,j} = \begin{cases} \lambda/R_i, \quad j = i+3, B_i \leq SR_i \quad \text{(arrival)} \\ \\ R_i\mu, \quad j = i-3, B_i \geq 1 \quad \text{(departure)} \\ \\ R_i\gamma_1, \ j = i - f_s(R_i - 1) - (3\lceil \frac{B_i}{R_i} \rceil - 1) \ \text{(node failure)} \\ \\ R_i\gamma_2, \ j = i - f_s(R_i - 1) - (3\lfloor \frac{B_i}{R_i} \rfloor - 1) \ \text{(link failure)} \\ \\ \delta_1, \quad j = i + f_s(R_i) - 1 \qquad \text{(node repair)} \\ \\ \delta_2/(S+1), \quad j = i + f_s(R_i) - 2 + 3X \quad \text{(link repair)} \\ \\ \beta, \quad j = i + f_s(R_i) \quad \text{(new MR introduction)} \\ \\ 0, \quad \text{other } j \neq i \\ \\ -\sum_{k=1}^{n} q_{i,k}, \quad j = i, k \neq i. \end{cases}$$

Let us explain each of the transition rates of $q_{i,j}$ in details.

- Arrival: The arrival of packets in any state increases the number of packets in the system by 1 as long as there is buffer space available in the system, i.e., $B_i \leq SR_i$.

- Departure: The transmission (departure) rate of the system is proportional to number of available MRs at any time.

- Node failure: The node failure happens for the MR with highest load, and $R_i > 0$. This will cause the packets in the failed MR to be lost.

- Link failure: The link failure can happen for any link involving an active MR and the AR when $R_i > 0$.

- Node repair: This increases number of available MR(s) by 1 with no addition in the number of packets.

- Link repair: The link repair can bring a MR into the system with X number of packets, where $X \in \{0, 1, .., S\}$.

- Introduction of new MR: This increments the number of active MRs by 1 without increasing the number of packets in the system.

Once the infinitesimal generator matrix Q have been determined, the steady state probability distribution (π) of the CTMC can be obtained as follows:

$$\pi Q = \mathbf{0}. \tag{7.3}$$

When a packet arrives, if the system is in state $(0, 0, K)$ or state (R_i, SR_i, K) (where $K \in \{0, 1, 2\}$), the packet is dropped due to lack of buffer space in the system. Therefore, the dropping probability can be calculated by:

$$\begin{aligned} P_d &= \pi D^T \\ \text{where } D &= [D_0, D_1, \cdots D_j \cdots D_{N_r}], \\ \text{and } D_j &= [0, \cdots 0, 1, 1, 1]_{3(jS+1)}, j = 0, \cdots, N_r - 1 \\ \text{and } D_{N_r} &= [0, \cdots 0, 1]_{SN_r+1}. \end{aligned} \tag{7.4}$$

The average number of packets ($E(m)$) in the whole system can be determined as follows:

$$\begin{aligned} E[m] &= \pi v^T \\ \text{where } v &= [v_0, v_1, \cdots v_j \cdots v_{N_r}] \\ \text{and } v_j &= [0, 0, 0, 1, 0, 0, \cdots, Sj + 1, 0, 0], j = 0, \cdots, N_r - 1 \\ \text{and } v_{N_r} &= [0, 1, 2, \cdots, SN_r]. \end{aligned} \tag{7.5}$$

Hence, we can obtain the average packet delay using Little's law as follows:

$$E[T] = \frac{E[m]}{\lambda_{accepted}} = \frac{E[m]}{\lambda(1-P_d)}. \tag{7.6}$$

7.2.4 Survivability model with single MR

The transition diagram for NEMO with single MR is a simplified version of Fig. 7.2 which is shown in Fig. 7.3. Here, the total number of states is $S + 4$.

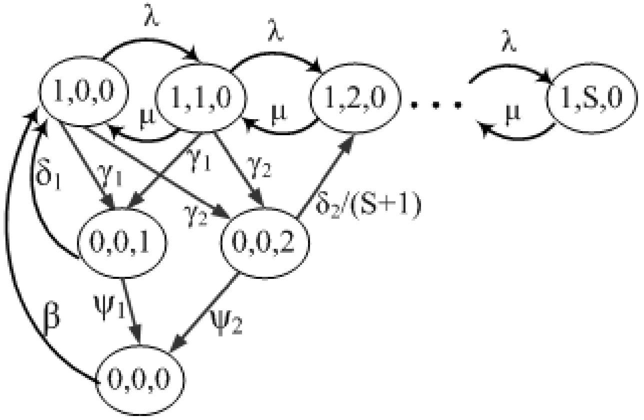


Figure 7.3: State transition diagram for NEMO with one MR.

We number the states $(1,0,0)$, $(1,1,0)$, $(1,2,0)$, ...,$(1,S,0)$ as states $1,2,3,..,(S+1)$. The states $(0,0,1)$, $(0,0,2)$, and $(0,0,0)$ are numbered as $(S+2)$ through $(S+4)$. The generator matrix Q can be expressed as follows:

$$q_{i,j} = \begin{cases} \lambda, \quad j = i+1, B_i \leq S, R_i = 1 \quad \text{(arrival)} \\ \mu, \quad j = i-1, B_i \geq 1, R_i = 1 \quad \text{(departure)} \\ \gamma_1, \; i \leq S+1, j = S+2 \text{ (node failure)} \\ \gamma_2, \; i \leq S+1, j = S+3 \text{ (link failure)} \\ \delta_1, \quad i = S+2, j = 1 \qquad \text{(node repair)} \\ \delta_2/(S+1), \quad i = S+3, j \leq S+1 \;\; \text{(link repair)} \\ \beta, \quad i = S+2, j = 1 \;\; \text{(new MR introduction)} \\ 0, \quad \text{other } j \neq i \\ -\sum_{k=1}^{n} q_{i,k}, \quad j = i, k \neq i. \end{cases}$$

Steps similar to Eqns. (7.3)-(7.6) can be followed to compute the steady state probabilities, packet dropping probabilities, the average number of packets in the system and the average packet delay.

7.3 Results

In this section, we evaluate the survivability of NEMO using the analytical model developed in Secs. 7.2.3 and 7.2.4. The values of the system parameters are listed in Table 10.2 which are similar to [40]. We explain the logic behind setting these values to the system parameters. The data packet arrival rate is kept 100 packets/sec and the signaling traffic rate is used as one-tenth of data traffic. Node failures happen once in every 500 sec whereas link failures happen twice in every 500 sec. Link recovery requires twice the node recovery time as it might be difficult to detect in the first place. Some failure may be fatal or irreversible with very small rates.

Table 7.1: Values of system parameters used in numerical analysis.

Parameters	**Meaning**	**Value**
λ_d	Data packet arrival rate	100 per sec
λ_s	Signaling packet arrival rate	10 per sec
μ	packet transmission rate	300 per sec
γ_1	Node failure rate	0.002 per sec
γ_2	Link failure rate	0.004 per sec
$1/\delta_1$	Node repair time	10 sec
$1/\delta_2$	Link repair time	20 sec
ψ_1	Irreversible node failure rate	0.00001 per sec
ψ_2	Irreversible link failure rate	0.00002 per sec
β	Rate of introducing error-free MR	0.001 per sec
S	Queue size	15 packets

Figure 7.4 shows the impact of DDoS attack strength on the packet drop probability at MR for different queue size. We have used $N_r = 3$ for this graph. It is found that with increasing DDoS attack strength, the packet drop probability increases drastically compared to its normal values as there is no way to distinguish between legitimate data (or signaling) packets and the DDoS attack packets. In addition, more packets are dropped for smaller queue size as the buffer slots are filled up quickly by the attack packets.

Figure 7.5 shows the impact of number of MRs and failure rates on the mean packet delay. It is found that increase in number of MRs reduces the mean delay since long queues cannot develop in a MR; rather packets are distributed among the MRs and get served faster. Moreover, the mean delay is higher for higher failure rates since higher failure rates may cause more packets to be queued in the active MRs, thereby increasing the delay.

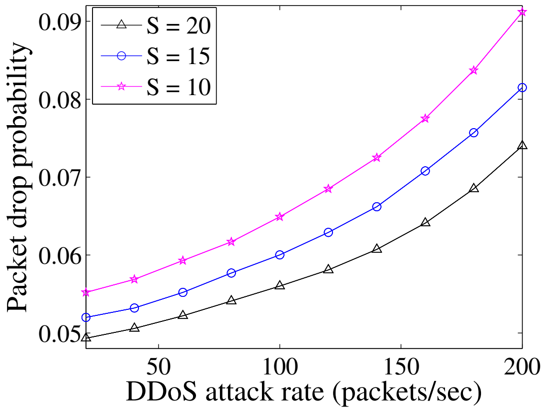


Figure 7.4: Effect of DDoS attack strength on the packet drop probabilities for different queue size.

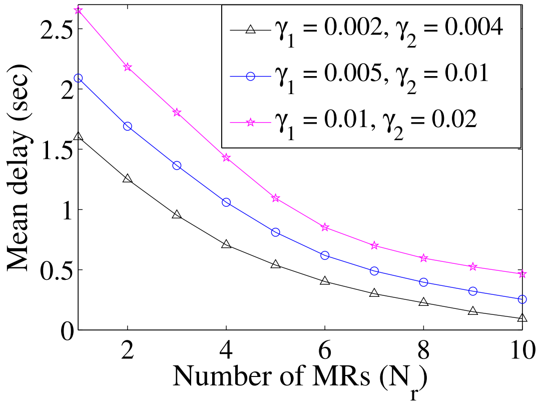


Figure 7.5: Impact of number of MRs on the packet delay for different failure rates.

Figure 7.6 shows the impact of mean time to recover (MTTR) on the average packet drop probability for different data packet arrival rates. In this case, we have assumed the failure recovery times for node and link failures to be equal. It is found that higher MTTR causes more packets to be dropped due to the lack of sufficient number of active MRs. In addition, higher data traffic rate raises the packet drop rate since more data packets are fed into the system, causing more drops.

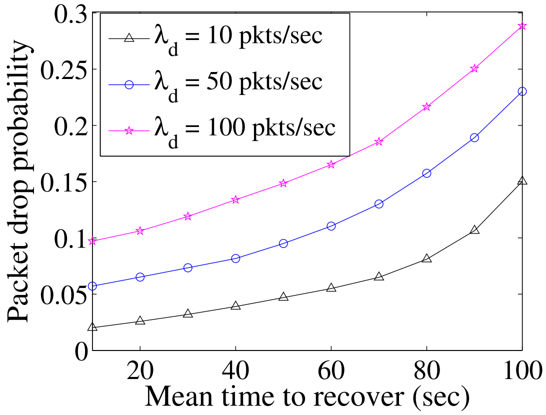


Figure 7.6: Impact of MTTR on the packet drop probability.

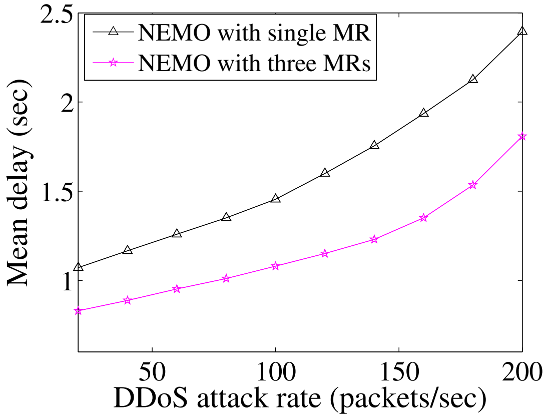


Figure 7.7: Impact of DDoS attack rate on mean packet delay.

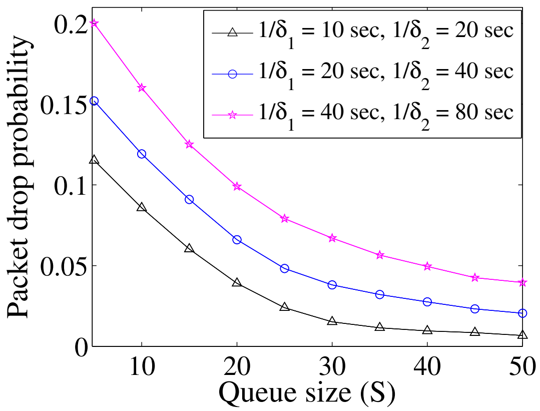


Figure 7.8: Effect of queue size on the packet drop probabilities for different mean recovery times.

In Fig. 7.7, the impact of DDoS attack strength on the mean packet delay is shown for two NEMO scenarios. We have used $S = 10$ for this graph. Mean delay increases for higher values of λ_a, as more attack traffic causes queues to be filled up with these traffic, raising the delay for all the packets. Moreover, the delay is higher for single MR case, as this means lower queue size and less processing speed for the incoming packets in the system.

Figure 7.8 shows the impact of queue size on the packet drop probability at MR for different recovery times. Larger queue size can accommodate more packets, thereby reducing the drop rate. On the other hand, lower recovery time causes nodes or links to be up again, resulting in better performance, i.e., lower drop probability.

7.4 Summary

In this chapter, we have developed an analytical model for quantitative survivability evaluation of NEMO with multiple mobile routers taking into consideration possible natural or man-made failures as well as DDoS attacks. Our results reveal interesting relationships among network performance, failure rates and attack strengths. It is found that increase in the number of mobile routers reduces the mean delay and drop probability of data packets while withstanding attack packets. Our survivability model can be used as a framework for quantitative survivability evaluation of other host or network-based mobility protocols.

Next, we perform experimental evaluation of multihomed NEMO which aims at seamless handover and increased availability.

Chapter 8

Experimental Evaluation of Multihomed NEMO

In the previous chapter, we performed quantitative survivability analysis of NEMO having multiple routers. In this chapter, we propose a seamless handover scheme for multi-homed NEMO architecture by exploiting the multihoming feature (having multiple physical network interfaces) of the mobile router. We perform experimental evaluation for the handoff performance of the proposed multihomed NEMO and compare it with NEMO basic architecture. Results demonstrate that the proposed multihomed NEMO outperforms the basic NEMO while achieving seamless handover.

8.1 Introduction

Mobile router plays the key role in a mobile network and acts as the gateway for all its nodes, connects them to the global Internet, forwarding signaling traffic as well as data traffic to the desired remote hosts. Mobile router usually has higher transmission capability and is usually powered by the vehicle. In basic NEMO, the MR uses single interface to connect to the access network. During the handover of NEMO, the MR has to break the connection with the old access network before establishing a connection ('break-before-make') with new access network, resulting in high handover latency and packet loss. Applications that are highly sensitive to delay and packet loss are affected badly due to NEMO protocol operation. Therefore, it is essential to ensure seamless handover of the mobile network so that the performance of real-time applications (e.g., voice-over-IP) is not compromised.

Original NEMO supported only one (primary) care-of-address to be registered with its home agent (or correspondent node). However, multiple physical interfaces in a mobile node can benefit from increased availability, fault tolerance, ubiquitous access, load balancing, and flow distribution through simultaneous wireless access, thereby reducing the delay and packet loss during handoff. Recently, IETF has proposed extension to NEMO allowing Multiple Care-of-Addresses registration (MCoA) [50] although it has not specified the way to exploit it to ensure seamless handover between access networks. The MCoA registration policy can be used to establish a new connection before breaking the old one, known as make-before-break strategy.

There have been several works related to NEMO with MCoA. Pan et al. [51] proposed a capacity-aware MCoA framework for mobile nodes to choose the preferred link with highest available bandwidth, without having any experimental validation. Romain [52] demonstrated the fault-tolerance and load-balancing of NEMO MCoA with an experimental testbed although the handover was not seamless. Chen et al. [53] proposed another handover algorithm for NEMO in a heterogeneous environment and analyzed the performance through experimentation. Sazzad et al. [54] compared the performance of NEMO with a transport layer mobility protocol using experimental testbed. Petander et al. [55] measured the handoff performance and routing overheads of multihomed NEMO using an experimental testbed. However, they [54, 55] did not use MCoA registration feature of NEMO. The author is not aware of any thorough experimental evaluation of NEMO MCoA that exploits the make-before-break strategy. Our work differs from the previous works in a way that we have used a cross layer approach to exploit the extension of multihomed NEMO and presented a thorough analysis of experimental results demonstrating seamless handover of NEMO.

Our objective of the work presented in this chapter is to exploit the multihoming feature of NEMO to achieve seamless handover and to demonstrate it through experimentation. Our proposed scheme is a cross layer approach which senses link layer signal strength in the overlapping area and makes the soft handover decision, thereby reducing delay and packet loss during handoff.

Our contributions in this chapter are (i) proposing the system framework for NEMO MCoA that exploits 'make-before-break' strategy to achieve seamless handover through multihoming, and (ii) evaluating the handover performance of the proposed scheme and comparing it with NEMO through real experimental testbed.

Our experimental results validate that our proposed scheme outperforms basic NEMO in terms of handoff delay, round trip time and throughput– three major performance metrics for any mobility management scheme.

The rest of the chapter is organized as follows. Our proposed seamless NEMO architecture is explained in Section 8.2. In Section 8.3, the experimental setup for basic NEMO and NEMO MCoA is described, followed by the experimental results in Section 8.4. Section 8.5 concludes the chapter.

8.2 Seamless Handover Scheme for NEMO

Original NEMO basic support protocol allowed only one care-of-address registration per home address of a mobile router. Wireless devices available nowadays have multiple network interfaces that aim at constant connectivity with the Internet through different access technologies, such as Wi-Fi, GPS, 3G networks. Recently IETF has proposed extension to NEMO allowing MCoA registration [50] of a MR's home address in the HA. However, the IETF RFC 5648 [50] has not specified the way to exploit MCoA feature to ensure seamless handover between wireless access networks.

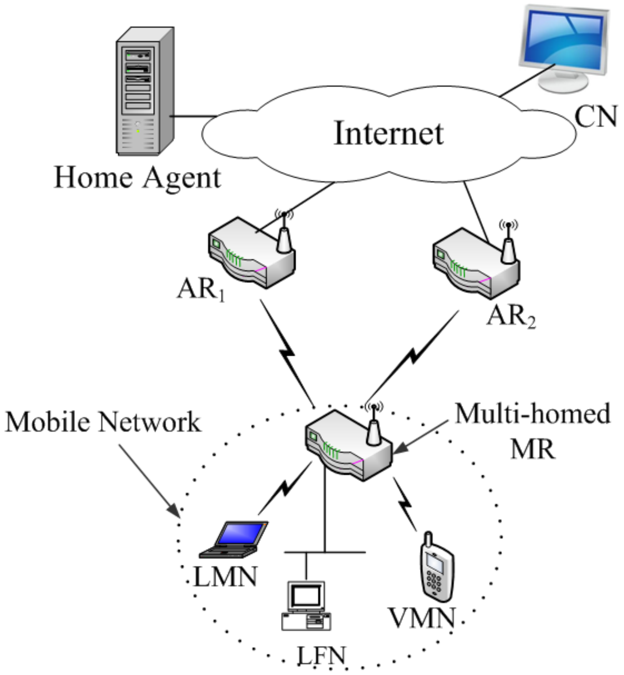


Figure 8.1: Architecture of multi-homed NEMO.

We propose a cross layer approach that works in combination with MCoA registration policy to ensure seamless handover for multihomed NEMO (see Fig. 8.1) in which the MR has multiple network interfaces that can acquire IP prefixes from ARs while residing in the overlapping radio coverage area. The MR then sends binding update to the HA to register the acquired CoAs in HA's binding cache (facilitated by IETF's MCoA registration policy [50]). This ensures establishing a new connection before breaking the old one (called make-before-break strategy). The new CoA is sent (through BU) to the CN so that traffic is sent through the new AR to avoid packet loss during handover. The MR also scans the link layer signal strength to make decision of handoff to the stronger access network. We name our proposed scheme as *M-NEMO* since it exploits the multihoming feature.

The proposed soft handover process of M-NEMO is explained briefly using a flow diagram in Fig. 8.2. The MR obtains CoAs from ARs through MR's multiple

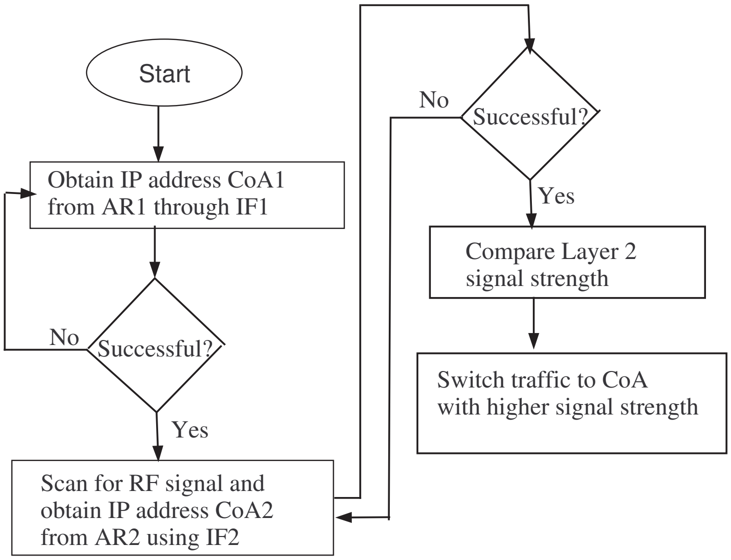

Start
Obtain IP address CoA1 from AR1 through IF1
Successful?
No
Yes
Scan for RF signal and obtain IP address CoA2 from AR2 using IF2
Successful?
No
Yes
Compare Layer 2 signal strength
Switch traffic to CoA with higher signal strength

Figure 8.2: Soft handover algorithm using M-NEMO.

interfaces and compares the link layer signal strength to make decision to handoff to the higher valued signal. This ensures least packet loss and delay during handoff.

8.3 Experimental Setup

For the performance evaluation of basic NEMO and M-NEMO, we have used Linux-based experimental testbeds. The testbed setup for basic NEMO and M-NEMO are described in the following subsections.

8.3.1 NEMO testbed setup

There exists several open source implementation for Basic NEMO, e.g., NEPL [56], SHISHA [57], etc. For our NEMO testbed setup, we used the NEPL implementation since we are using a Linux-based testbed for M-NEMO and NEPL is based on

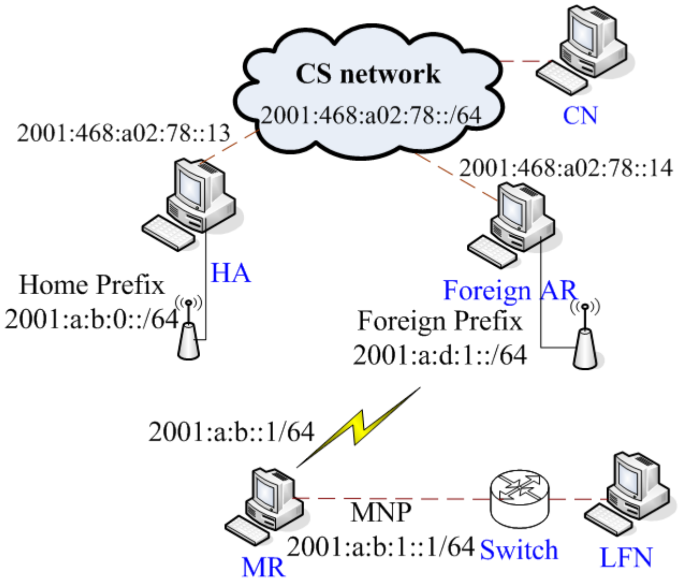


Figure 8.3: Testbed architecture for basic NEMO.

Linux platform unlike SHISHA which uses BSD platform. This will ensure a fair comparison with M-NEMO testbed.

Figure 8.3 shows the experimental testbed for basic NEMO where the mobile network has single level of nesting. Table 8.1 summarizes the hardware and software configurations of the devices used in the NEMO testbed. To capture the real network phenomena, the testbed is connected to the University of Oklahoma (OU) operational network that carries production traffic.

The testbed architecture of NEMO consists of home network (which advertises the home prefix 2001:a:b:0::/64), foreign network (which advertises the foreign network prefix 2001:a:d:1::/64), CN, MR, and LFN. The access networks (home or foreign) are connected to CS network. The global IPv6 prefix of the CS network is 2001:468:a02:78::/64. All the devices of the mobile network were placed on a trolley that was moved between the home and the foreign network, and handover data was captured using wireshark network protocol analyzer.

Table 8.1: Configuration of devices for basic NEMO testbed.

No	Device Type	Software Configuration	Hardware Configuration
1	MR	Debian 2.6.22 Kernel + NEPL	CPU: Intel Pentium 4, 2.20 GHz, 512 MB RAM, NIC: 802.11 based Netgear MA111
2	LFN	FC5 + FTP Client	CPU: Intel Pentium 4, 1.73 GHz, 1 GB RAM
3	HA	Debian 2.6.22 Kernel + NEPL	CPU: Intel Pentium 4, 1.50 GHz, 512 MB RAM
4	AR_1	FC6 2.6.18-1 kernel + radvd-1.0	CPU: Intel P4, 1.50 GHz, 512 MB RAM
5	AR_2	FC6 2.6.18-1 kernel + radvd-1.0	CPU: Intel P4, 1.73 GHz, 512 MB RAM
6	APs	Channel 6 and Channel 11	DLink WBR-1319
7	CN	FC5 + FTP Server	CPU: Intel Celeron, 2.8 GHz, 512 MB RAM

8.3.2 M-NEMO testbed setup

Figure 8.4 shows the experimental testbed for M-NEMO and Table 8.2 summarizes the hardware and software configuration of the devices used in the testbed. The testbed configurations are almost similar to each other. Some of the devices, such as AR_1 and AR_2 are exactly the same while other devices have very little differences between them and it does not have much impact on the output pattern. As shown in Fig. 8.4, there are two access routers, AR_1 and AR_2 that advertises two foreign prefixes (2001:a:d:1::/64 and 2001:a:c:1::/64, respectively). The MR is now equipped with two wireless NIC cards that can connect to both the foreign links simultaneously whenever the mobile network is in the radio coverage area of both ARs. In that case, the MCoA registration is done in the HA and we can verify the addresses through *bc* command in the HA as shown in Fig. 8.5.

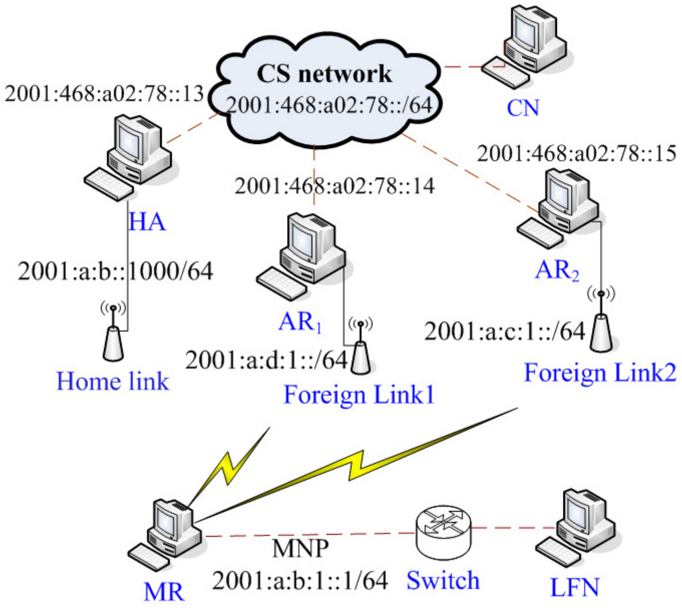
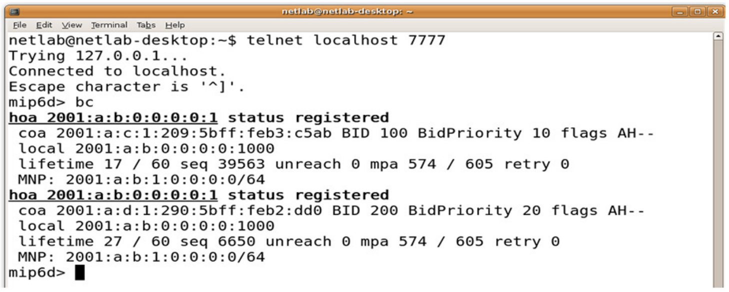


Figure 8.4: Testbed architecture for M-NEMO.

Table 8.2: Configuration of devices for M-NEMO testbed.

No	Device Type	Software Configuration	Hardware Configuration
1	MR	Ubuntu 8.04 Kernel 2.6.23 + NEPL	CPU: Intel Core 2 Duo, 2.20 GHz, 2 GB RAM, NIC: 802.11 based two Netgear MA111
2	LFN	Windows XP + FTP Client	CPU: Intel Celeron, 2.19 GHz, 256 MB RAM
3	HA	Ubuntu 8.04 Kernel 2.6.23 + NEPL	CPU: Intel Core 2 Duo, 2.20 GHz, 2 GB RAM
4	AR_1	FC6 2.6.18-1 kernel + radvd-1.0	CPU: Intel P4, 1.50 GHz, 512 MB RAM
5	AR_2	FC6 2.6.18-1 kernel + radvd-1.0	CPU: Intel P4, 1.73 GHz, 512 MB RAM
6	APs	Channel 6 and Channel 11	DLink WBR-1310
7	CN	Windows Vista + FTP Server	CPU: Intel Core 2 Duo, 2.2 GHz, 2 GB RAM



```
netlab@netlab-desktop:~$ telnet localhost 7777
Trying 127.0.0.1...
Connected to localhost.
Escape character is '^]'.
mip6d> bc
hoa 2001:a:b:0:0:0:0:1 status registered
 coa 2001:a:c:1:209:5bff:feb3:c5ab BID 100 BidPriority 10 flags AH--
 local 2001:a:b:0:0:0:0:1000
 lifetime 17 / 60 seq 39563 unreach 0 mpa 574 / 605 retry 0
 MNP: 2001:a:b:1:0:0:0:0/64
hoa 2001:a:b:0:0:0:0:1 status registered
 coa 2001:a:d:1:290:5bff:feb2:dd0 BID 200 BidPriority 20 flags AH--
 local 2001:a:b:0:0:0:0:1000
 lifetime 27 / 60 seq 6650 unreach 0 mpa 574 / 605 retry 0
 MNP: 2001:a:b:1:0:0:0:0/64
mip6d> █
```

Figure 8.5: Binding cache entry in Home Agent for M-NEMO.

8.4 Results

The experimental results are presented in this section. We measure the throughput, Round Trip Time (RTT) and handoff latency by analyzing the packet flows through Wireshark network protocol analyzer at the CN, MR and LFN.

8.4.1 Throughput

The rate at which payload data are received at any node is termed as its throughput. In our experiment, LFN receives data traffic from CN. We measure throughput at the LFN by analyzing the wireshark capture data.

Figure 8.6 shows the throughput at LFN for NEMO BSP tested during handoff between home network and foreign network. The variations in throughput graph within a network is caused by the network congestion resulting from cross traffic in CS operational network. However, while the mobile network performs handover from the home to foreign network between t = 23 sec and t = 36 sec , the throughput (at LFN) becomes zero for about 13 sec. The LFN does not receive any data from the CN during this period which is explained further in Section 8.4.2.

Figure 8.7 shows the throughput at LFN for M-NEMO testbed. Unlike NEMO BSP, the throughput in M-NEMO does not drop to zero during the handoff period t = 20.073 sec to t = 20.148 sec. This verifies that M-NEMO throughput is not

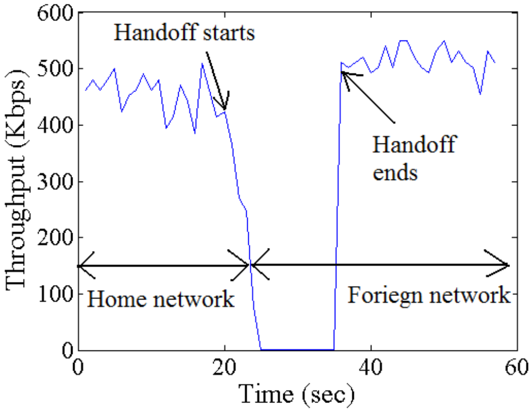


Figure 8.6: Throughput at LFN for NEMO BSP.

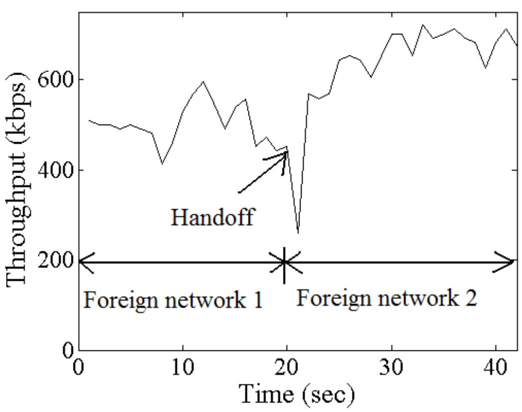


Figure 8.7: Throughput at LFN for M-NEMO.

affected much by the handoff as the communication can continue through the MR's other network interface. Thus, M-NEMO is benefitted by the multihoming feature of MR unlike NEMO BSP.

8.4.2 Handover latency

Handover latency is defined as the time interval between the last data segment received through the old access network and the first data segment received through the new access network. In order to measure the handover latency of NEMO BSP, we analyzed the wireshark capture data at the LFN (see Fig. 8.8). As we moved from the home network to foreign network, the MR acquired CoA from the foreign network, sent binding update to the HA (at t = 29.528 sec) and received the binding acknowledgement from the HA (at t = 31.135 sec). We also found that the last data segment (with sequence number 999) received by the LFN through the old AP is at t = 23.2397 sec, whereas the first data segment (with sequence number 1000) received through the new AP is at t = 36.1593 sec. Therefore, the handover latency is 12.9196 sec (36.1593 - 23.2397).

For M-NEMO, the TCP sequence numbers received at the LFN are plotted in Fig. 8.9. Initially, the MN was in the AR_1's radio coverage area, so the HA had only

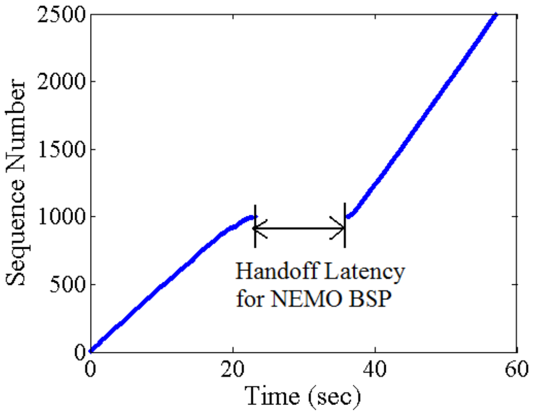


Figure 8.8: TCP sequence numbers of data received at LFN in NEMO BSP.

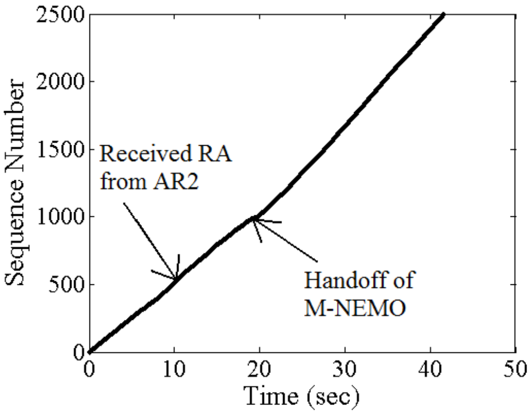


Figure 8.9: TCP sequence numbers of data received at LFN in M-NEMO.

one binding entry in its binding cache. Next, we moved towards the AR_2. We found that the MR received the router advertisement from AR_2 at t = 11.5487 sec. This means that the MN entered the radio coverage area of AR_2 and acquired second CoA from AR_2 and notifies HA to register both CoAs. Thus, the make-before-break was achieved by the MR. At t = 20.0735 sec, LFN received the last packet (with sequence number 1021) through AR_1, and at t = 20.1484 sec, the first packet with sequence number 1022) through AR_2 arrived at the LFN. Therefore, the handover latency for M-NEMO is 75 msec (20.1484 - 20.0735). As in Fig. 8.9, there is almost no gap between sequence numbers during M-NEMO handover which verifies that LFN did not face in any disruption in receiving data segment during handoff. This essentially demonstrates the seamless handoff capability of M-NEMO through soft handover by exploiting MR's multihoming facility.

We ran experiments for 10 different trials and the handoff latencies for each run of NEMO BSP and M-NEMO are shown in Figs. 8.10 and 8.11, respectively. The average values of the handoff latencies are shown in (blue and red) straight lines which are 12.96 sec and 75 ms for NEMO and M-NEMO, respectively. Using two-sample right-tail t-test, we found that the t-test rejects the null hypothesis with 5% significance level. The 95% confidence interval on the mean of the differences

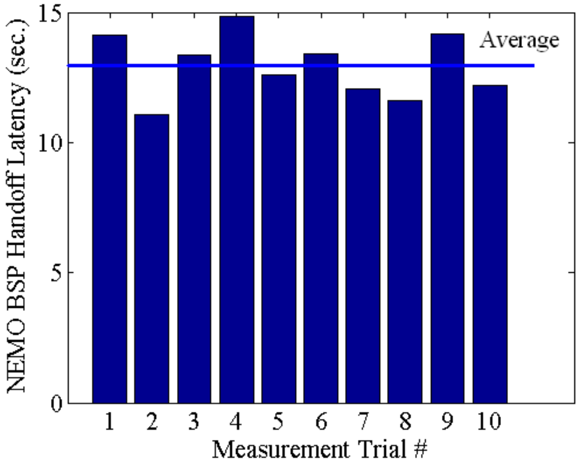


Figure 8.10: Handoff latency bar chart for NEMO BSP.

between NEMO and M-NEMO handoff latencies is at least 12.91 sec. Thus the hypothesis testing verifies that mean handoff latency of M-NEMO is much less than NEMO BSP.

8.4.3 Retransmissions during handoff

From our analysis, we found that a large number of packets were retransmitted by the CN due to the large handoff latency (connection disruption). This is shown in Fig. 8.12 where we find that data segments with sequence number 1000, 1001, and 1002 were retransmitted thrice, twice and twice, respectively. The following segments were retransmitted once. This forces CN to back off, and the timeout value for the TCP sender at CN is increased, thereby producing poor throughput at the LFN during handoff (see Fig. 8.6).

However, this is not the case for M-NEMO testbed. As shown in Fig. 8.13, we can see that there were fewer number of retransmissions (10) than NEMO BSP testbed (28). Hence, the throughput of M-NEMO did not drop drastically during handoff.

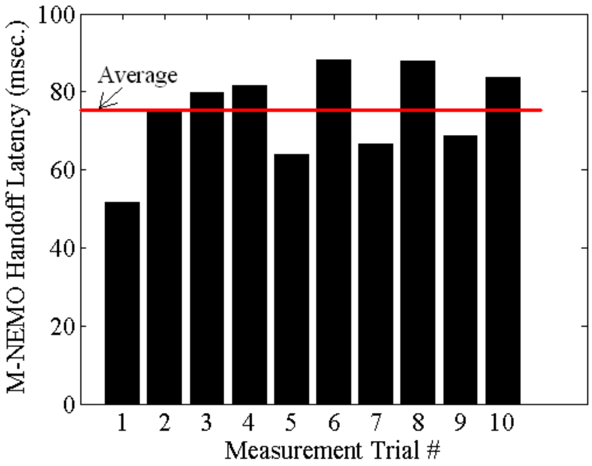


Figure 8.11: Handoff latency bar chart for M-NEMO.

8.4.4 Round trip time

RTT is measured by the difference in time between the CN sending a packet and receiving the corresponding acknowledgement. In case of lost (data or acknowledgement) packets, the RTT takes time difference between successful reception time of Acknowledgement at CN and first time sent time of the data packet.

Figure 8.14 shows the RTT between CN and LFN measured at CN for NEMO BSP. The gap in the RTT graph (between t = 23 sec and t = 36 sec) represents the handoff when the connection was interrupted. This disruption caused retransmission timeout at CN due to lost data or acknowledgement packets. So the CN started to retransmit those packets. Hence, we observe the sudden spike (of around 15 sec) in the RTT graph of NEMO. Packets transmitted during handover period suffer from large RTT which is also explained earlier in Fig. 8.12.

Figure 8.15 shows the RTT between CN and LFN measured at CN for M-NEMO. The values of the RTT remains fairly stable during the handoff period (t = 20 sec) of M-NEMO. There are few small spikes due to the cross traffic from CS production

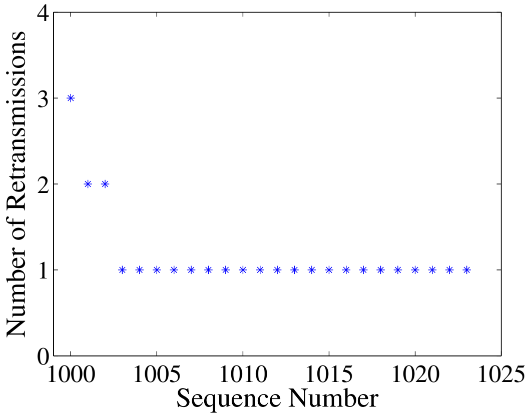


Figure 8.12: Number of retransmissions during NEMO BSP handoff.

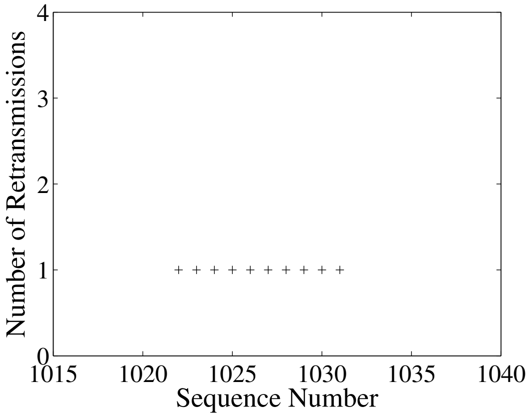


Figure 8.13: Number of retransmissions during M-NEMO handoff.

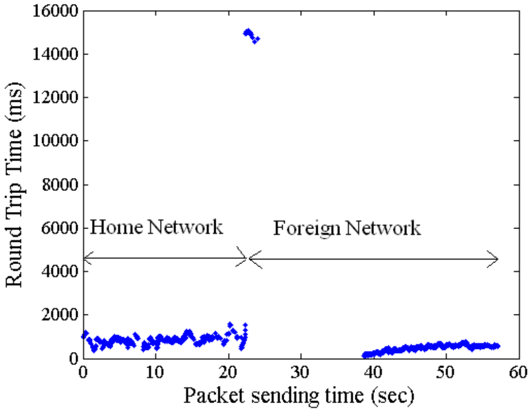


Figure 8.14: RTT observed at the CN in NEMO BSP.

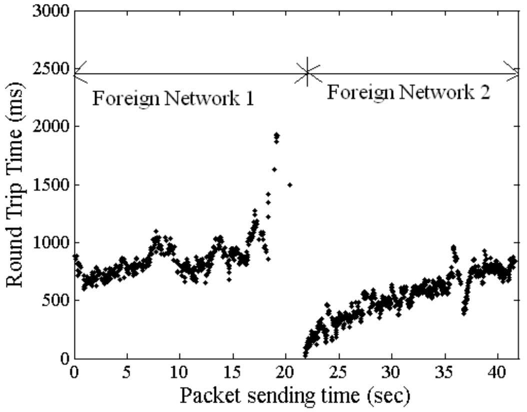


Figure 8.15: RTT observed at the CN in M-NEMO.

network. The stability of RTT implies that packet loss during handoff is minimum, thereby confirming M-NEMO handover to be seamless.

8.5 Summary

In this chapter, we have proposed a seamless handover scheme for NEMO exploiting the multihoming feature of the mobile router. We have used experimental testbeds to measure the handoff performance (throughput, round trip time, and handoff latency) of multihomed NEMO and compared it with basic NEMO. Results show that basic NEMO and multihomed NEMO have handover delay of 13 sec and 75 msec, respectively. In addition, the throughput remains fairly unaffected during handoff. Thus, our experimental results validates that our proposed scheme outperforms basic NEMO in terms of handoff delay, round trip time and throughput– three major performance metrics for any mobility management scheme.

In the next chapter, we focus on the security issues and corresponding defense mechanisms of the mobility management protocols.

Chapter 9

Security Issues of Mobility Protocols

Mobility protocols can be vulnerable to security threats. This is because the communication media is wireless (therefore, easily accessible by attackers) and the mobility protocols require data to be delivered to a node which are moving and can have multiple point-of-attachments. These special features in mobile computing have introduced several security issues. Any malicious agent can try to fool mobility agents by sending spoofed control messages and redirecting traffic away from the victim nodes, hijacking ongoing sessions, or even modifying the contents if proper protection mechanisms are not enforced. In this chapter, we explain with illustrative examples major security threats on various components of the network due to IP-mobility protocols. We have analyzed the existing defense mechanisms along with their capabilities and limitations to prevent or mitigate the security threats. We also propose an effective defense mechanism for SIGMA to protect against possible threats from malicious agents.

9.1 Introduction

IP-mobility protocols require signaling among the mobility agents to keep track of the mobile node's current location and maintain its reachability. Moreover, mobility protocols that employ route optimization (to ensure direct data path between end hosts) are required to send binding updates to its correspondent nodes and it can significantly improve the performance of mobility protocols. However, these binding updates are vulnerable to various attacks since malicious agents may send fabricated binding updates to deceive mobile host, home agent or its correspondent node. Thus, the requirement of seamless connectivity in mobile environment and use of route optimization techniques between the communicating nodes have introduced several security issues in IPv6 networks.

Some of the major threats for mobility protocols are traffic redirection attack, man-in-the-middle attack, replay attack, bombing attack, denial-of-service attack, home agent poisoning, etc. These are serious threats for the integrity and confidentiality of data packets, leading to session hijacking and resource exhaustion as well as degrading performance of key network entities.

To prevent or mitigate security attacks, the existing defense mechanisms aim at choosing solutions that are simple enough to be implemented in mobile devices, computationally less expensive and low latency solutions so that the main objective (seamless connectivity) of mobility protocol is not affected. Several defense mechanisms have been proposed to protect against the vulnerabilities of mobility protocols, such as return routability protocols, IP Security (IPsec) protocols, IKE-based schemes, etc.

There have been several attempts to identify potential threats to the public Internet due to IP-mobility protocols' operation. Kempf et al. [58] outlines the security threats to Mobile IPv6 and explain how the security features of Mobile IPv6 protocol mitigate them. Hu et al. [59] discusses and outlines the security threats for network mobility architecture and propose a public Key Infrastructure (PKI) and secret key based protection approach for it. Elgoarany et al. [60] present a survey on the Mobile IPv6 security through the classification of threats and possible scenarios. IETF RFC for Mobile IPv6 [17] also lists possible security threats and discusses some techniques that can protect against such threats. However, there is lack of research work that outlines all the possible security vulnerabilities caused by mobility protocols along with detailed analysis of existing defense mechanisms.

The objective in this chapter is to point out major security vulnerabilities of mobility protocols, and analyze existing defense mechanisms to compare their effectiveness to protect mobility infrastructure.

The contributions in this chapter are (i) explaining with illustrative examples all the major security threats on various components of the IPv6 network due to the introduction of the IP-mobility protocol, (ii) critically analyze the existing defense mechanisms while pointing out their capabilities and limitations to guard against major security threats, and (iii) proposing effective defense mechanisms for SIGMA protocol.

Our comprehensive analysis on the security threats for mobility protocols and corresponding defense mechanisms presented in this chapter can help mobile users identify possible vulnerabilities in their networks and choose a suitable solution to protect against threats.

The rest of the chapter is organized as follows. In section 9.2, a brief explanation is given for binding updates in IP-mobility protocols. In Section 9.3, major security threats relating to IP-mobility protocols are discussed with illustrative examples. In Section 9.4, existing defense mechanisms are explained in brief. In section 9.5, we critically analyze the defense mechanisms along with their capabilities and limitations to prevent or mitigate threats. In section 9.6, we propose security measures to prevent or mitigate security threats for SIGMA protocol. Section 9.7 has the concluding remarks.

9.2 Binding Update in IP-mobility protocols

In IP-mobility protocols, while away from its home, an MH is also associated with a Care-of Address (CoA), which provides its current location information. Whenever

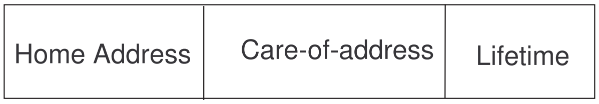


Figure 9.1: Major fields in the binding update.

a MH acquires a new CoA from a foreign network, it must inform its HA about this new CoA through a Binding Update (BU).

Figure 9.1 shows the main information contained in a BU message: HoA, CoA and the binding entry lifetime The HA accepts the BU and updates it binding cache (a table maintained by the HA for all its MHs), and sends a binding acknowledgement.

Data packets from the Correspondent Node (CN) follows an un-optimized route to MH (CN –> HA –> MH) (as shown in Fig. 2.1) instead of direct route (CN –> MH). This degrades the performance of Mobile IP introducing larger delay.

To alleviate the performance penalty, Mobile IPv6 or other protocols another mode of operation known as route optimization (RO). Figure 9.2 shows the MIPv6 route optimization where MH sends BU to the CN informing the newly acquired CoA along with its home address. The CN, an IPv6 node, caches the binding of the MH's home address with the CoA, and send any packets destined for the MH directly to it at this CoA. Thus, using Mobile IPv6, an MH may change their point-of-attachment to the Internet without changing its home address, allowing them to maintain transport and higher-layer connections while roaming.

9.3 Threats for Mobility Protocols

In this section, we explain major security threats for the mobility protocols with illustrative scenarios. In summary, most of the potential threats exploits false bindings, usually resulting in Denial-of-Service (DoS) attacks. Unauthenticated binding

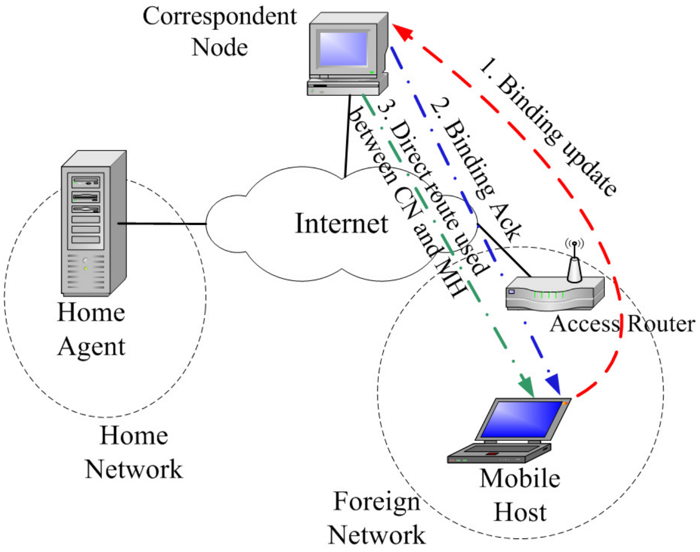


Figure 9.2: Mobile IPv6 Route Optimization.

updates can create serious security vulnerabilities. The attacker can use fabricated BU, thereby deceiving CN about the MH's current location. This may lead to traffic redirect attack as well as Man-In-The-Middle (MITM) attacks, compromising the secrecy and integrity of data packets.

There are some attacks that are due to manipulation in the IPv6 routing headers or in the home address options. Some attack tries to use all the resources (CPU, memory) to degrade the performance of the mobile host or the mobility agent. Some attack blocks MH sending legitimate binding update by flooding packets in the radio access network while some forces to use sub-optimal route for delay sensitive packets. These security threats for mobility protocols are due to the fact that mobility is transparent to upper layer protocols and also due to the effort of making things simpler for the low-power mobile devices.

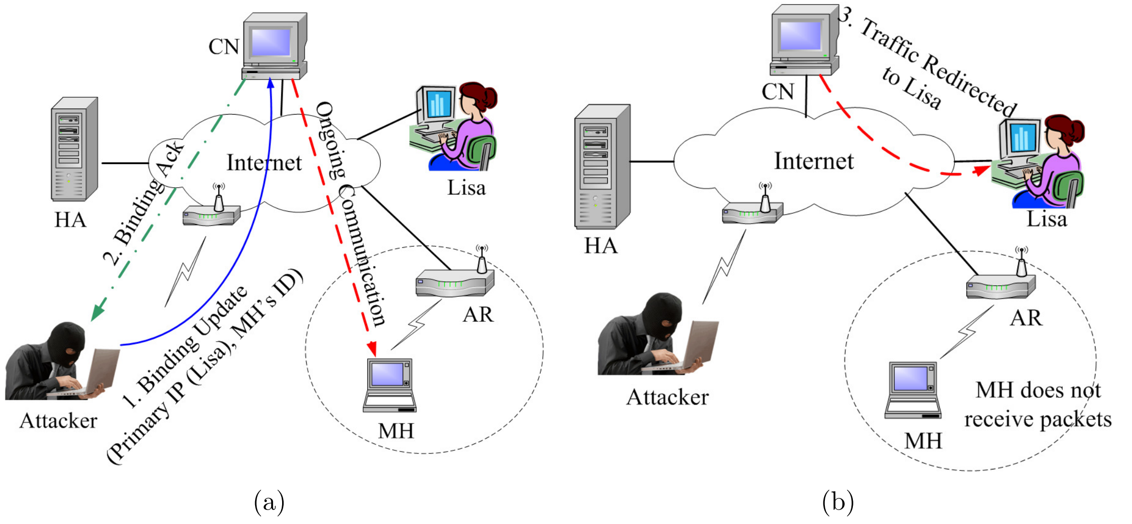
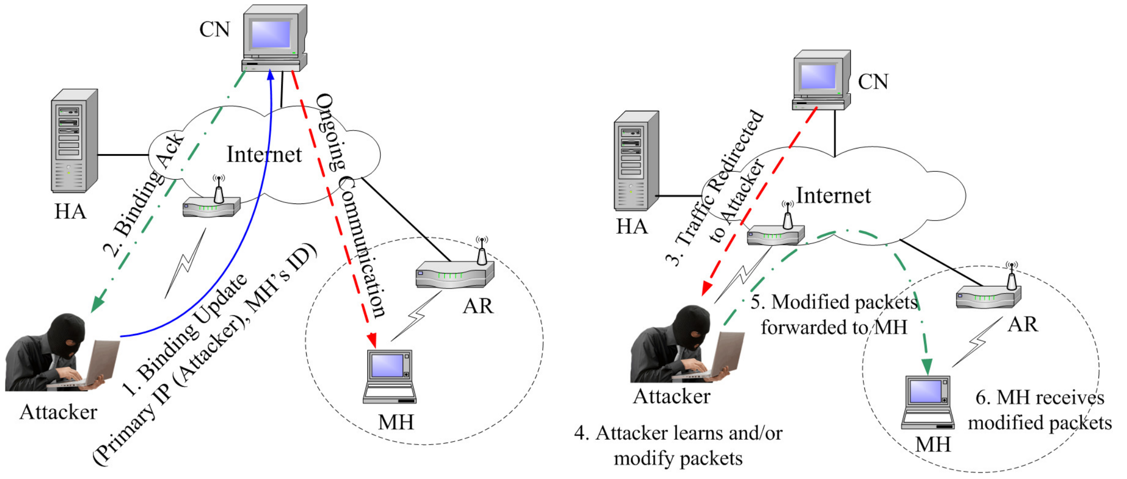


Figure 9.3: Traffic redirection attack (a) The attacker sends fabricated BU to the CN to modify the binding cache for the MH to some fictitious (Lisa) IP address and CN accepts the BU (b) Traffic is redirected away from the MH to Lisa's IP address.

9.3.1 Traffic redirection attack

The attacker may send a fake binding update message claiming that a node (victim) has changed its care-of address due to its movement to a new location. This may happen if the BU is not authenticated. If such BU is accepted by the the CN, it will start sending packets to the new CoA and the victim node will not get any traffic. As shown in Fig. 9.3(a), the attacker sends fabricated BU to the CN to replace MH's IP with some fictitious IP address (say Lisa's IP address) and CN accepts the BU. As the result, the ongoing session of CN with the MH has been redirected towards Lisa's location as shown in Fig. 9.3(b) and the MH loses all subsequent traffic of the session.

In most cases, data encryption and use of IP Security (IPsec) protocol cannot prevent such attack on data integrity and confidentiality, since route optimization signaling are transparent to IPsec, thereby redirecting the traffic even though the attacker cannot read the encrypted data. To launch traffic redirection attacks, the attacker needs the knowledge of the IP addresses of the communicating nodes.



(a) (b)

Figure 9.4: Man-in-the-middle attack (a) The attacker sends fabricated BU to the CN to modify the binding cache of the MH to its own (Attacker) IP address and CN accepts the BU (b) Traffic is redirected to the Attacker who learns the confidential information of the packet and may modify the packet before forwarding to the MH without the knowledge of the involved parties.

Therefore, nodes with well-known IP addresses, such as public servers, DNS servers or file servers are more vulnerable to such attacks.

Remedy: Nodes with frequently changing addresses or nodes that use randomly generated addresses may mitigate such attacks. However, this addition of security mechanisms to the BU process makes the mobility protocol slower and more complex.

9.3.2 Man-in-the-middle attack

The attacker might send binding update message to the CN telling it to modify MH's IP with its own (attacker's) IP address. If the CN accepts such binding update, CN will start sending the packets to the attacker instead of the MH. The attacker will then be able to learn the confidential contents of the message, may modify the packet before forwarding it to the MH. Thus, the attacker might act as a MITM getting all-important private data destined to the victim (MH) without the knowledge of the CN and the MH.

Figure 9.4 shows the MITM attack launched on the MH to steal information of the ongoing communication between the CN and MH. First, the attacker sends an malicious BU to the CN saying that the MH's CoA has changed and it is now the attacker's IP address. If CN accepts such fabricated BU from the attacker, it will confirm with a binding acknowledgement (see Fig. 9.4(a)). Since the CN has updated its binding cache in response to the malicious BU, it will start sending traffic towards the attacker rather than the MH as shown in Fig. 9.4(b). The attacker thus intercepts the packet to learn confidential contents of the message, may modify the packet content before forwarding it to the MH.

Another way of MITM attack is to send spoofed binding update long before CN and MH are in contact, with attacker's address as the CoA and MH's HoA in the home address option. If such binding update remains in CN's binding cache, the CN will send traffic through the attacker (assuming it to be the HA), thereby allowing the attacker to act as a MITM. Such attack can be mitigated by setting shorter lifetime of binding update.

9.3.3 Replay attack

This kind of attack takes advantage of previously sent (authenticated or unauthenticated) binding updates by recording it and later on, replaying it when the victim (MH) moves to some new location, thereby interrupting the CN-MH communication. The attacker may get the opportunity to record BUs while residing in the same radio access network where MH is located. Replay attack can also work on authenticated binding updates since the malicious agent only replays it to deceive the CN. Therefore, it is difficult to protect against such attacks.

A reply attack is shown in Fig. 9.5. The MH was first in subnet A in Fig. 9.5(a) and it sends a binding update to CN to update MH's CoA in CN's binding cache. Any attacker listening to such BU can record the BU and use that for replay attack in future. In Fig. 9.5(b), the MH has moved to some new subnet (B). Now the

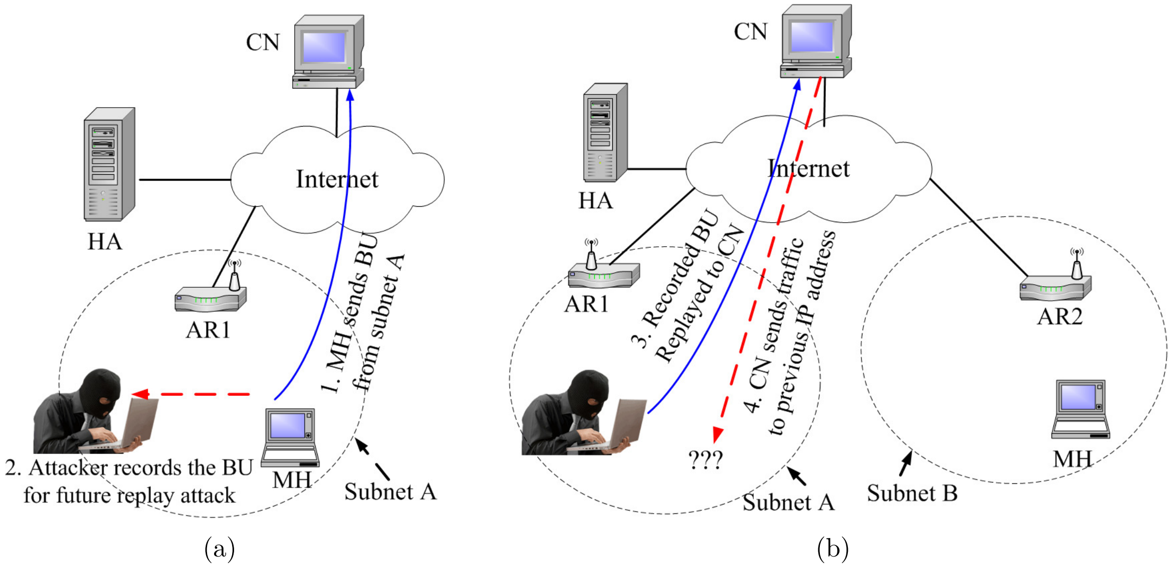

CN
Internet
HA
AR1
1. MH sends BU from subnet A
2. Attacker records the BU for future replay attack
MH
Subnet A
(a)
CN
Internet
HA
AR1
3. Recorded BU Replayed to CN
4. CN sends traffic to previous IP address
???
Subnet A
Subnet B
AR2
MH
(b)

Figure 9.5: Replay attack (a) When the MH is in subnet A, MH sends the BU to the CN informing its newly acquired care-of address and the attacker records the BU message for prospective replay attack in the future (b) When MH moves to other subnet B, the attacker sends the recorded BU of the MH claiming that MH is in subnet A, thus disrupting traffic away from the MH to some non-existing host.

attacker may use the previously recorded BU and send it to CN to deceive it. If the CN accepts such replayed message, CN would then start sending packets to the old (subnet A) address thinking that MH has moved to subnet A again which is not true. Thus, traffic from CN are redirected to a non-existing IP-address, thereby disrupting the communication.

9.3.4 Bombing attack

In this type of attack, huge amount of unsolicited data traffic is redirected to the victim node (or a network) to degrade its performance as well as bandwidth wastage. The attacker may exploit real-time streaming servers for this kind of attack. First, the attacker establishes a connection with streaming server, and starts to download a stream of data. After getting the sequence number, the attacker might claim that it has moved to a new location. The attacker might use the IP address of the victim

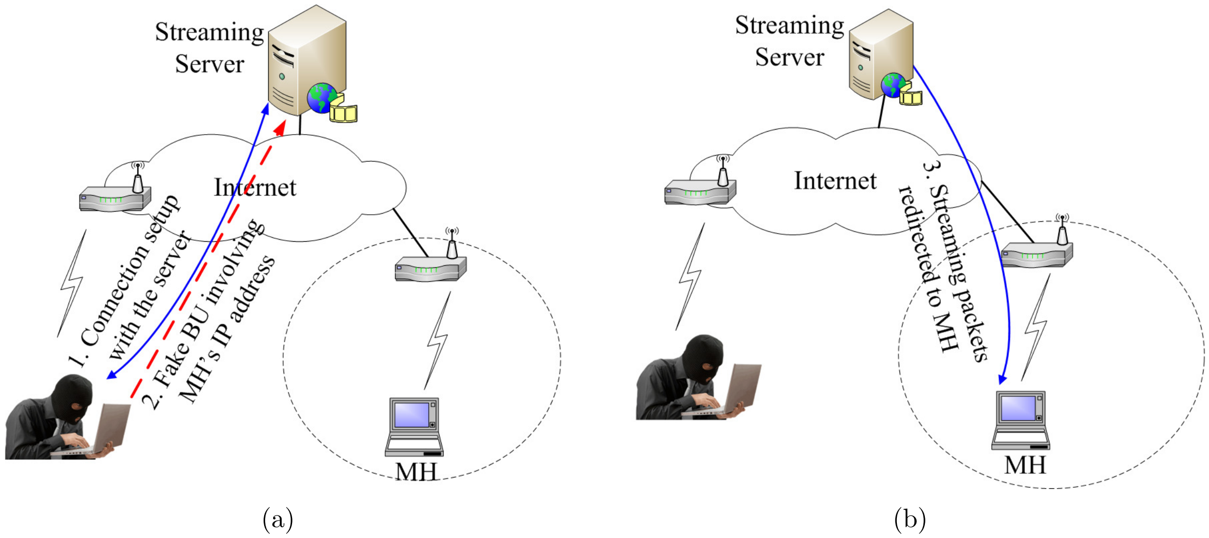

Streaming Server
Internet
1. Connection setup with the server
2. Fake BU involving MH's IP address
MH
(a)
Streaming Server
Internet
3. Streaming packets redirected to MH
MH
(b)

Figure 9.6: Bombing attack (a) The attacker establishes a connection with a streaming server, later on the attacker sends a fake BU involving the IP address of the MH, (b) The streaming data packets are redirected to the MH that the MH has not requested for.

node in the binding update. As a result, subsequent packets from the server will be directed to the victim node.

Figure 9.6 shows the bombing attack on a MH which overwhelms MH with unsolicited data packets and degrade its performance. In Fig. 9.6(a), the attacker establishes a connection with a streaming server and after some time, it sends a false BU to the server claiming that its CoA has changed. In the BU message, the attacker uses the IP address of the victim MH. As a result, the traffic from the streaming server has been redirected to the MH causing its performance degradation and bandwidth wastage.

In such attacks, the victim node will not accept those unsolicited (streaming data) packets and therefore, will not send the acknowledgement, thereby stopping the communication. However, the attacker can ensure a a continuous flow of data streams sent to the victim by sending spoofed acknowledgement (towards the server) as the attacker knows the initial sequence number. One possible remedy of this attack could be to use the TCP RESET signal by the victim node to immediately stop such large flow of data stream. This may not be possible since the victim will always drop the packets immediately without even processing the appropriate header to know the actual source address.

The bombing attack can be very serious since it can target any Internet node or a whole subnet with enormous amount of unwanted data and the target node cannot do anything to stop such data flow, thereby losing its bandwidth without any clue to such attacks. This attack may become severer and harmful to the Internet if it is used in combination with distributed denial-of-service (DDoS) attacks.

9.3.5 Reflection attack

In some earlier design of Mobile IP, the CN could initiate route optimization signaling whenever the CN receives packet through HA. This can lead to reflection attack. Route optimization was initiated to the address that was included in the Home Address option. An attacker can take advantage of this and can send traffic with a care-of-address of the victim and the victim's address in the Home Address option, thereby redirecting RO signaling to the victim. Figure 9.7 shows the reflection attack where the attacker sends a false initial message to the CN, thereby inducing CN to send two messages to the MH. As a result, the MH receives every packet sent by the attacker twice due to the reflection. Thus the attacker is able to amplify a packet flooding attack against a target MH by a factor of two. Moreover, the identity of the attacker of such reflection attacks remains undetected as both the messages arriving at the target have the CN's address as the source address.

9.3.6 Home Agent poisoning

The HA keeps the mapping of HoA to CoA of the MH. Therefore, in every subnet crossing location updates are sent to the HA to update the database entry accordingly. The entry can be corrupted if spoofed BU is accepted by the HA. This will affect all subsequent communication with that host whose entry has been corrupted

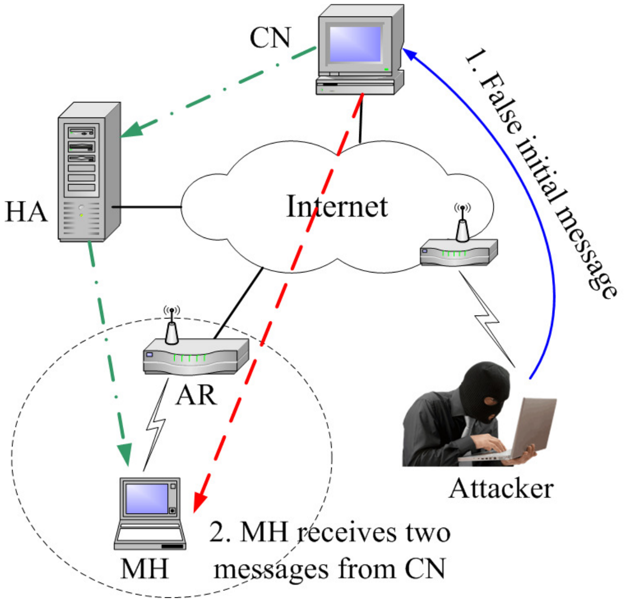


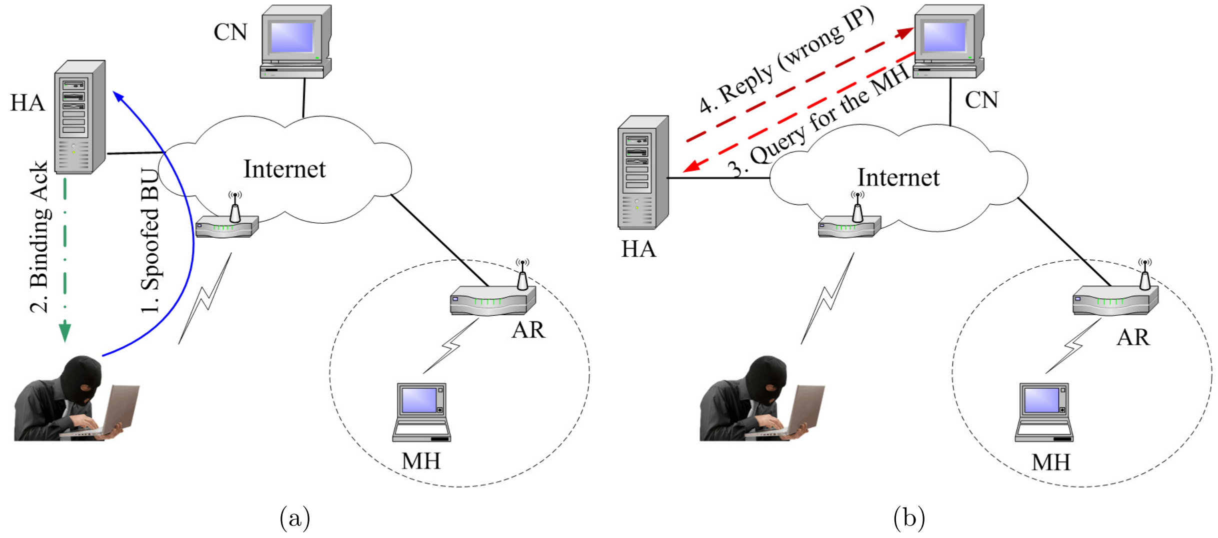
Figure 9.7: Reflection attack.

and no Internet node will be able to reach the victim node. Figure 9.8 shows the HA poisoning. The attacker sends spoofed BU to the HA (Fig 9.8(a)) and the HA accepts the BU. Therefore, the subsequent query to the HA by any CN (for the MH) will produce wrong reply as shown in Fig. 9.8(b).

To present this kind of attack, the HA must verify two things. First, the node sending BU is a legitimate client of the HA and secondly, this node sending BU has the right to change its CoA. Otherwise, any legitimate client of the HA may launch attack and modify the binding cache of some victim node under the HA.

9.3.7 Blocking MH to send legitimate BU

When the MH enters a new radio access network, it obtains new IP address and sends BUs to the HA and the CNs. If the attacker is on the same radio link (as the victim MH), the attacker can block MH from sending legitimate BU by launching brute-force attack on the radio link or by a flooding attack. Thus when the MH



(a)

(b)

Figure 9.8: HA poisoning (a) Spoofed BU send to the HA and the HA updates the entry in the location database for the MH (b) When the CN queries the HA for the IP address of the MH, it receives the wrong IP.

gives up sending BUs to the CN, the attacker can send fabricated BUs to the CNs and the HA, thereby redirecting MH's traffic towards the attacker. This could lead to MITM attack as well.

9.3.8 Resource exhaustion attack

Attacker establishes connections with the MH with thousands of fake IP addresses. Thus whenever, the MH moves to some new location, the MH has to send to send BU to these imaginary hosts, thus huge processing power of the MH is wasted while dealing with these unnecessary BUs. This attack cannot be prevented with authenticated BUs. These fake connection will require the victim to keep states for each one of them, wasting its memory as well, resulting in denial of service attacks.

Moreover, the attacker may trick MH to participate in unnecessary complex cryptographic operations, using up the resources of the MH. These attacks become severer for strong and expensive defensive mechanism.

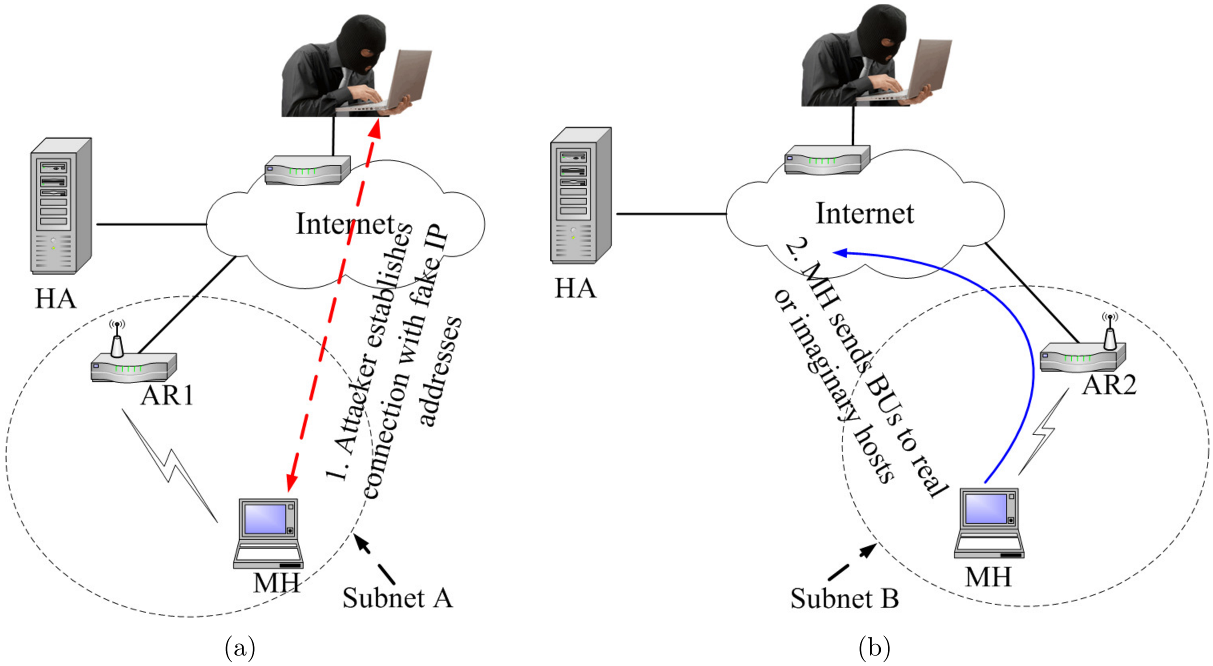

HA
Internet
AR1
MH
1. Attacker establishes connection with fake IP addresses
Subnet A
(a)
HA
Internet
2. MH sends BUs to real or imaginary hosts
AR2
MH
Subnet B
(b)

Figure 9.9: Resource exhaustion attack (a) The attacker establishes unnecessary connections with the MH using fake IP addresses (b) MH sends BUs to all the fake IP addresses thus wasting its processing power as well as memory.

9.3.9 Forcing sub-optimal routing

To prevent itself from attacks that aims at exhausting MH's resources (resource exhaustion attack), the MH can turn off route optimization to the CN. Thus, the traffic from the CN will follow sub-optimal route through the HA, causing more delay for packet delivery. This kind of attack is most effective for delay sensitive traffic, such as real-time audio/video traffic.

9.3.10 Exploitation of routing header

Routing headers pose general security threats that might be exploited by the malicious agents. When a packet having one or more routing headers reaches the IPv6 node specified in the destination address (IP header) field, the top routing header is popped out and the IP header of the packet is modified with the address from the routing header as the new destination address. The packet is then routed to this

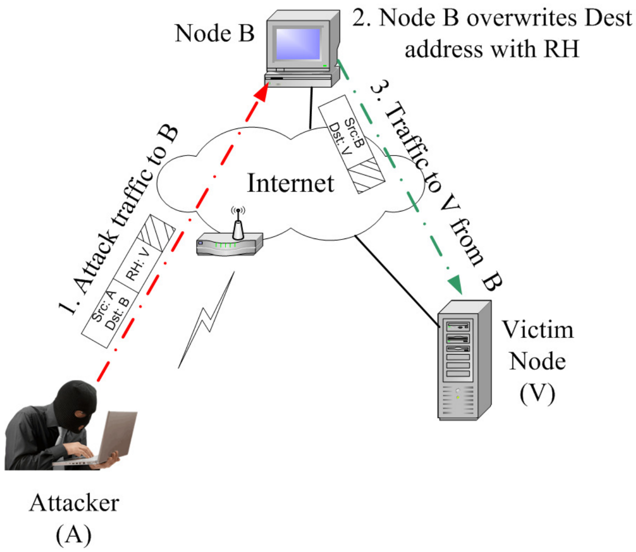


Figure 9.10: Exploitation of routing header to hide the attacker's source IP.

new destination. Thus, an attacker can misuse this routing header, and launch attack traffics indirectly going through multiple intermediate destinations. Therefore, finding the original source of the attack traffic might be very difficult for the victim node. In addition, general protection mechanism (such as ingress filtering) may not work to counter such attack since this attack allows skipping between the inside and outside of an administrative domain.

Figure 9.10 shows the exploitation of Routing Header (RH) by the attacker (A) that sends attack traffic to a victim node V, routing through an intermediate node B. The attacker sends attack traffic (with its own source IP) destined to the node B. However, the attacker includes a RH which has the IP address of the victim node V. When the packet arrives at the node B, since it contains a RH, node B overwrites the destination address field in the IP header with the content of RH (which is V) and the source address field is now B. Then the packet is sent to the victim (V). When it reaches V, the victim node cannot find out the actual source of this attack packet. Thus, the attacker can successfully hide his identity while sending malicious traffic through one or more intermediate nodes.

9.3.11 Exploitation of home address destination option

In an IPv6 packet, destination option field contains information that are only processed/examined by the destination node. Home address of any MIPv6-enabled node does not change and this HoA is used by the transport layer to identify a TCP connection between the MH and the CN. When a mobile node visits any foreign network, it obtains its care-of-address which is a topologically correct address. This address cannot be used in transport layer since it will require breaking of the TCP connection when the MN gets a new CoA. For RO traffic from the MN, the MN uses the CoA as the source address field (in route optimized packet towards the CN). However, the MN must also identify itself and does so by including its home address in the home address destination option field. After receiving such packet from the MH, CN overwrites the source address field with the MH's HoA and forwards it to the transport layer. Thus, the transport layer never have to deal with the care-of-address.

An attacker can exploit this home address destination option to disguise the source address of the attack. The attacker can send malicious traffic (from outside) with a source address (in HoA option) within an administrative domain. The egress filter would allow such traffic as it is coming from outside. However, when the victim node receives the packet, it overwrites the source address with the address contained in the HoA destination option. Therefore, it appears that the attack is coming from the inside of the administration domain, which is misleading. Figure 9.11 shows such an illustration. The attacker sends malicious traffic to the victim node V. However, the attacker uses the HoA address of the node B in the HoA option field. When V receives the packet, it overwrites the source IP with the HoA option field (which contains B's HoA) as if node B has launched the attack. Thus, the attacker can

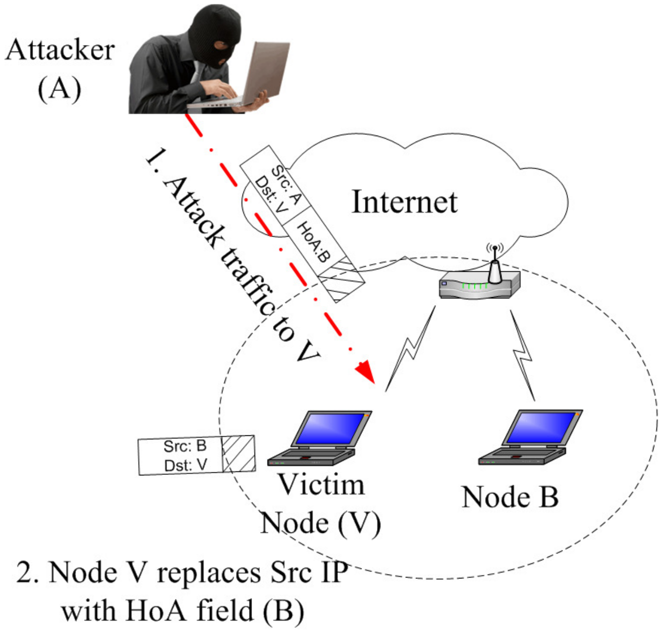


Figure 9.11: Exploitation of HoA option to hide the attacker's source IP.

easily disguise its identity bypassing the firewall and it seems that somebody within the V's administrative domain has launched the attack.

9.3.12 Acting as the HA and learning about home network

Each MH learns about its HA through Home Agent discovery message. However, MIPv6 does not protect this message with IPsec since this message is sent to anycast address in the home network. Therefore, any attacker in the home network can take this opportunity to claim itself as the HA and can initiate security associations with the MH.

In the event of network renumbering in home network, the HA can send ICMPv6 unsolicited mobile prefix reply message to the MH to update it. If such message is not protected properly, any malicious agent may send such message, thereby updating the MH's home prefix to the attacker's network. Moreover, the attacker may learn about the topology of the MH's home network by soliciting ICMPv6 mobile prefix message to the HA. This knowledge may be exploited in future attack (such as bombing attack) towards the home network

9.4 Defense Mechanisms

In this section, some of the defense mechanisms that can be used to prevent the prospective attacks have been discussed. The main considerations while designing security solutions are summarized as follows:

- Focus on the attacks that are introduced due to IP-mobility protocol.

- Low processing requirement: The processing overhead required for cryptographic operations and/or authentication protocols are relatively high for low-power mobile devices. Therefore, defense mechanisms that are simple and computationally less expensive are suitable to be implemented in mobile nodes with low processing power.

- Infrastructure less approach: To protect against malicious BU leading to traffic redirection and MITM attacks, authentication of BU is essential. However, use of strong cryptographic (authentication) protocols require the existence of certification infrastructure. As there is no distinction between a fixed IPv6 node and a mobile node, this certification infrastructure is required to authenticate all IPv6 nodes across the public network. However, at present there is no such existing infrastructure that can be used to authenticate all IPv6 nodes. The deployment of such global infrastructure is neither realistic nor feasible in the current Internet. Therefore, infrastructure-less solutions are more realistic solution.

- Low latency efficient solution: The main focus of the mobility protocol is to facilitate uninterrupted ongoing communications between the MH and CN. If the security protocols requires significant amount of time for computation, the connection between the parties may be broken. Therefore, it is desirable that the security protocols are fast enough to meet this goal.

9.4.1 IP security protocol

IP security (IPsec) [61] is a suite of protocols designed to provide inter-operable, high quality, cryptographically-based security for IPv4 and IPv6. IPsec services [61] are provided through the use of two traffic security protocols, namely the Authentication Header (AH) and the Encapsulating Security Payload (ESP), and through the use of cryptographic key management procedures and protocols.

In any mobility protocol, it is assumed that MH has a prior trust relationship with the HA and IPsec (AH/ESP) protocol is suitable to be used to authenticate binding updates between MH and the HA. However, it might not be so for the BUs between the MH and the CN due to the absence of such trust relationship since the CN can be any IPv6 node in the Internet. Moreover, these exists no such global infrastructure that can be used to authenticate all IPv6 nodes. Therefore, use of AH protocol to authenticate the BUs between the MH and CN is not feasible. Alternative solutions for securing MH-CN BUs is the use of return routability protocol (discussed later in this section) or any other infrastructureless authentication.

In IPsec protocol, a preconfigured Security Association (SA) is established between the MH and the HA to authenticate the binding update and the following binding acknowledgement. SAs can be established through Internet Key Exchange (IKE) [62] with certificate authentication discussed in the next subsection. Each SA records the algorithm and parameters controlling security operations. An index parameter, called the Security Parameters Index (SPI) is used in security associations. They are referenced by the sending host and established by the receiving host. SAs are unidirectional and two SAs must be established between the MH and the HA for the bi-directional tunnel required for mobility signaling. The SAs are based on HoA instead of CoA. Therefore, SAs are not required to be changed when the MH moves and attaches to a new access network. Once the security association has been performed, the MH and HA are ready to use IPsec protocol. Therefore, when MH moves to a new subnet, it sends BU message.

9.4.1.1 Authentication Header protocol

IPsec Authentication Header (AH) protocol [63] is one of IP security protocols that can ensure that the binding update is originated from the MH, not from malicious agent or attacker. AH protocol guarantees connectionless integrity and data origin authentication of IP packets. However, AH protocol cannot provide confidentiality or privacy of the contents. Though confidentiality is not strictly required to protect MH-HA signaling, other mobility signaling, such as in return routability protocol require confidentiality to protect keys exchanged between MH, HA and CN. Hence, AH protocol is not widely used. Instead, IPsec ESP protocol is used since it supports encryption of messages in addition to ensure authenticity.

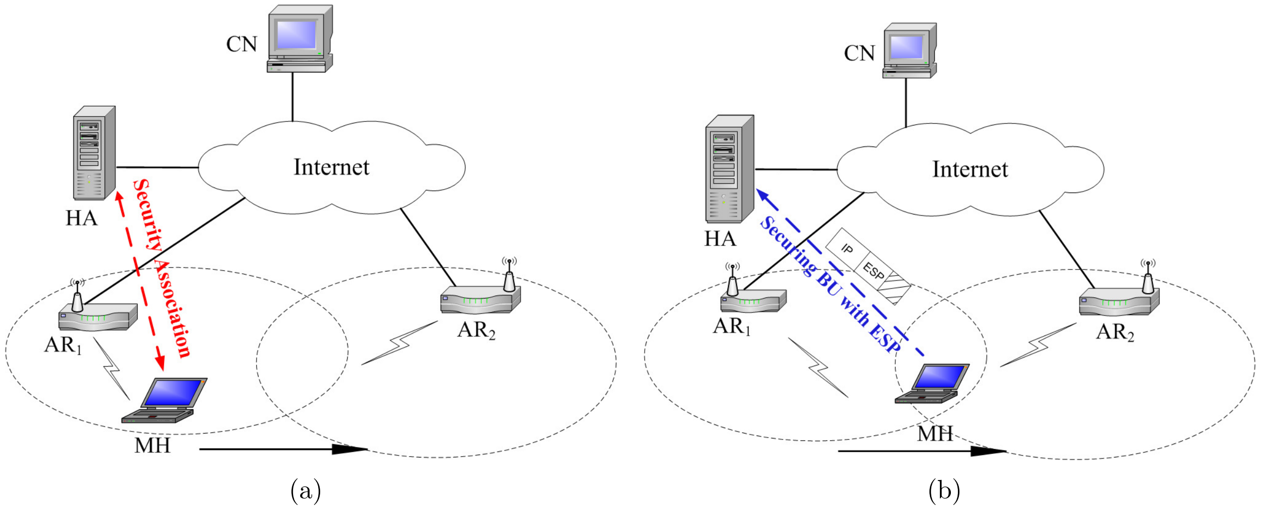


Figure 9.12: Protecting security and integrity using ESP (a) Security association performed between MH and the HA (b) The binding update sent by MH is protected by ESP header.

9.4.1.2 Encapsulating Security Payload protocol

IPsec ESP protocol [64] can be used since ESP can provide confidentiality, data origin authentication, connectionless integrity, anti-replay service and traffic flow confidentiality. At the time of security association, the set of services can be chosen.

ESP protocol ensures confidentiality of data by encrypting the datagram. An encryption algorithm combines the data in the datagram with a key to transform it into an encrypted form. This is then repackaged using a special format (with ESP header, trailer and authentication data) and transmitted to the destination. After receiving the encrypted packet, the destination node decrypts it using the same algorithm. ESP supports its own authentication scheme or can be used in conjunction with AH. The ESP header is inserted after the IP header and before the next layer protocol header.

In IP-mobility protocols, the binding update, binding ack, ICMPv6 prefix discovery solicitation and corresponding reply message can be protected by IPsec ESP protocol. MH may use ESP for both authentication and encryption to protect MH-HA communication against traffic analysis and to provide privacy for traffic routed over public Internet.

Figure 9.12 shows the use of ESP header for security data packets between the MH and HA. A security association is performed between the MH and HA to choose security algorithm and the related parameters in Fig. 9.12(a). After the security association, the MH sends BU packet to the HA with proper encryption along with the ESP header as shown in Fig. 9.12(a), thereby ensuring data integrity and confidentiality.

9.4.2 IKE-based schemes

IKE or IKEv2 [62], a key distribution mechanism for Internet community, is commonly used to performing mutual authentication and establishing and maintaining security associations for IPsec protocol suite. To ensure confidentiality, data integrity, access control, and data source authentication to IP datagrams, IPsec maintains state information at the two ends of the data communication. IKE helps to dynamically exchange the secret key that is used as the input to the cryptographic algorithms. Security associations are established using the Internet Security Association and Key Management Protocol (ISAKMP). Thus, IKE establishes a secured framework to distribute public keys and defines ways to generate those keys.

IKE works in two phases. In phase 1, two communicating peers establish a secure authenticated communication channel, namely ISAKMP security associations using the DiffieHellman key exchange [65] algorithm to generate a shared secret key to encrypt further IKE communications. In phase 2, SAs are negotiated on behalf of services, e.g., IPsec ESP that needs key or parameter negotiation.

9.4.2.1 Limitations of IKE

Although IKE provides very strong security, it has the following limitations:

- IKE-based scheme require the existence of certification infrastructure for its operation. This requirement cannot be met to protect against spoofed binding updates between the MH and the CN (fixed or mobile) which can be any Internet node. This is because currently there exists no infrastructure-based solution that can authenticate all IPv6 nodes in the world. It is not also realistic to propose and build such global infrastructure.

- IKE requires very complex and power-consuming operations. This may be a major issue for low-end mobile devices when each of the packets (both outgoing and incoming) are required to be processed by the cryptographic engine in the MH.

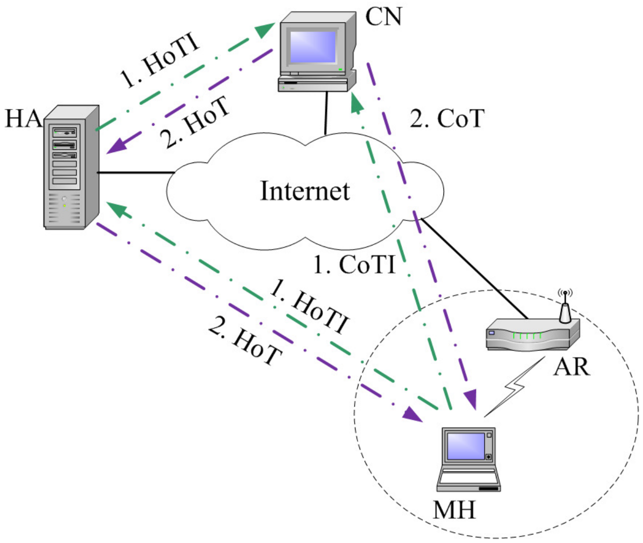


Figure 9.13: Return routability test in Mobile IPv6.

9.4.3 Return Routability protocol

One of the major security concerns for mobility protocols is the use of unauthenticated and forged binding updates. To prevent such attacks, a node sending a binding update must prove its right to redirect the traffic. The solutions proposed in MIPv6 [17] for this kind attack is Return Routability (RR) test. This approach of RR is used before each binding update message is sent to the CN, and they are exchanged among the MH, HA and CN. Figure 9.13 shows the message exchange in RR test. The MH initiates RR by sending Home Test Init (HoTI) and Care-of Test Init (CoTI) message to the CN; among these two messages, the HoTI goes to CN via the HA whereas the CoTI goes directly to the CN. The CN then send corresponding challenge packets Home Test (HoT) and Care-of Test (CoT) destined to the MH. After such message exchange, the CN accepts BU from the MH that is able to receive those challenge packet (HoT and CoT)

The first two messages of the test include two 64-bit cookies, the HoTI cookie and CoTI cookie (see Fig. 9.14). These cookies are randomly generated 64-bit numbers

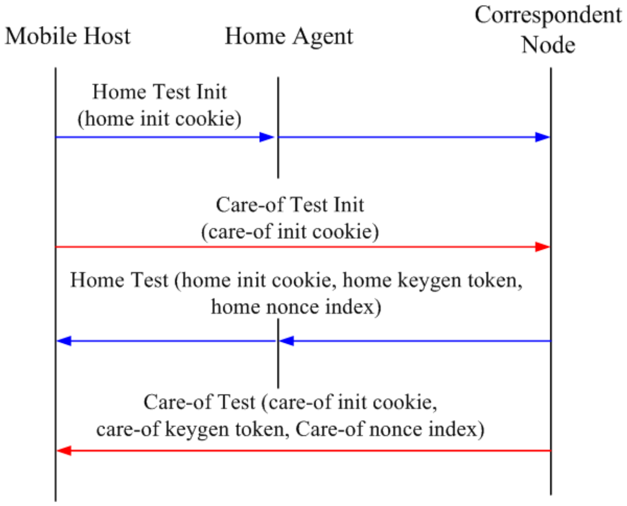

Mobile Host
Home Agent
Correspondent Node
Home Test Init
(home init cookie)
Care-of Test Init
(care-of init cookie)
Home Test (home init cookie, home keygen token, home nonce index)
Care-of Test (care-of init cookie, care-of keygen token, Care-of nonce index)

Figure 9.14: Message exchange in return routability test.

and they must be returned by the CN in the reply messages, that is, Home Test (HoT) and Care-of Test (CoT) messages. Each CN is assumed to maintain a 20-byte secret key, K_{cn} which is not shared with anyone and this value of K_{cn} is used as a parameter for the key generating function HMAC_SHA1() which is a specific construction for calculating a message authentication code (MAC) involving a secure cryptographic hash function SHA-1. K_{cn} is the first parameter of this function and the second parameter is composed of the concatenation of the Home (or Care-of) address, nonce index and a byte x. This byte is 0 for home address and 1 for care-of address. The first 64-bit of the output of the function is used as the key generation (keygen) token for the HoT and CoT message.

After receiving both the HoT and CoT messages, the MH first matches the cookies to make sure that they are same as those sent in the HoTI and CoTI messages. The mobile host then hashes both the (home and care-of) keygen tokens together and forms a 20-byte K_{bm} using the SHA1 function. The mobile host records the value of K_{bm} and the nonce indices included in the HoT and CoT messages associated with the correspondent host, for use in the binding update.

The RR protocol is considered to a relatively weak authentication protocol (compared to IKE-based scheme) to be used between the MH and the CN and it requires no certification infrastructure. The pros and cons of the protocol are explained in the following subsections.

9.4.3.1 Advantages

The use of the RR protocol has the following benefits:

- The RR protocol limits the number of potential attackers that can hijack an ongoing session. If RR is not used, any IPv6 node can spoof BUs to redirect traffic as shown in Figs. 9.3 and 9.4. The use of RR protocol can significantly scale down such damages though some attack is possible (explained in the limitation section)
- The RR protocol requires less CPU processing power and it only uses relatively inexpensive encryption and one-way hash functions unlike other complex cryptographic operations.
- The RR protocol is also stateless as the CN does not store a separate state for each mobile host. Instead, it stores a single periodically-changing randomly-generated secret key K_{cn} for this purpose and remains stateless until CN has authenticated the MH.

9.4.3.2 Limitations

Following are the limitations of the RR protocol:

- The vulnerabilities of the RR method exists on the path between the HA and the CN. As CN can be any node in the Internet, no prior relationship or security association exists between these nodes. Attackers who are on this path or have access to the packets sent on this path can learn the secret that is necessary for spoofing the BU. Such attacks include various DoS attacks, impersonation and eavesdropping, etc.

- Another vulnerability is possible when the CN is another mobile node at an unsecured access network. In that case, an attacker in such network may learn the keygen tokens and can send spoofed binding updates.
- The return routability are subject to race condition though the chance is very low. Return routability process starts after the MH has sent the binding update to the HA. The race condition is possible if this binding update is delayed to reach the HA whereas the HoT message is returned by the CN to the HA. This results in tunneling the HoT message to the wrong care-of address by the HA.

Thus, RR protocol is a relatively weak routing-based authentication method and it does not protect against all possible attacks, rather aims at limiting the number of potential attackers for a particular target, and number of targets a potential attacker can threaten.

9.4.4 Protection for routing headers related issues

Several security vulnerabilities are related to routing header and home address destination option that are used to optimize route between MH and CN. The protection mechanisms to prevent their malicious use are summarized in the following subsections.

9.4.4.1 Routing header

To protect misuse of routing headers to deceive IPv6 nodes and sending packets hiding the source's identity, MIPv6 defines a new routing header, called type 2 routing header. This new routing header enables packets to be routed directly from a CN to the MH's CoA. The MH's CoA is inserted into the IPv6 destination address field. Once the packet arrives at the CoA, the MH retrieves its home address from the routing header, and this is used as the final destination address for the packet.

Mobile IPv6 follows several rules to process this routing header and prevent misuse of the routing header. Firewalls and/or other access control devices applies these rules allowing type 2 routing headers while blocking other routing headers. Following are the list of rules/restrictions for type 2 routing headers:

- Only one routing header per packet is allowed.
- All IPv6 nodes that process this routing header must verify that the address contained within is the node's own HoA.
- The IP address contained in the routing header must be a unicast routable address since it is the MH's HoA.
- The scope of the HoA (within the routing header) must not be smaller than the scope of the CoA.

A node must drop the packet if any of these conditions are not met while processing the type 2 routing header.

9.4.4.2 HoA option

To protect against attacks regarding HoA destination option, similar restrictions are used:

- The IP address contained in HoA option must be a unicast routable address.
- If it is non-BU packet, the CN must not accept any packet with a HoA that dose not have any current binding in its cache.
- For BU packet, it must be protected by ESP (in case for the HA) or by some security parameters established by RR protocol (in case for the CN).

- The receipt of a HoA option must not trigger any change in the routing table or binding cache in a node.

9.4.5 Additional measures to mitigate attacks

There are a few measures that can be taken by the entities to protect against possible attacks. These are described as follows.

9.4.5.1 Keeping nodes stateless

An IPv6 node may not save any state information for receiving and replying to BU messages. This stateless approach can prevent the CN from DoS attacks by malicious agents causing resource (CPU and memory) exhaustion. To make CN stateless, the BU will have to contain enough information so that accounting can be done for legitimate BUs.

9.4.5.2 Keeping shorter lifetime for binding entry

To mitigate the attack based on the spoofed BU, one possible approach is to limit the binding entry lifetime. As a result, binding entry is removed from the cache of the CN if no further BU is received. Therefore, the attacker cannot take advantage of the old binding entry when the MH is inactive for some time.

The problem of such quick expiration (of binding entry) is the wastage of bandwidth and transmission power of the MH and the CN (or HA) in legitimate situations. These messages are absolutely unnecessary resulting in overhead on the HA (or CN), sometimes leading to resource exhaustion.

9.4.5.3 Use of Cryptographically Generated Address

The use of Cryptographically Generated Address (CGA) [66] can reduce the chance of attack on a victim node. This idea was first introduced in a BU authentication protocol known as CAM [67]. In this approach, the least significant 64-bits of the IP address (the interface identifier) is selected by computing a 64-bit one-way hash of the node's public signature key. In the CGA approach, the MH signs the binding update with its private key and sends the public key along with the signed data. The recipient of the binding update hashes the public key and compares the hash to the address before verifying the signature on the location data. This prevents anyone other than the node itself from sending location updates for its address. The main advantage of this approach is that it provides public-key authentication of the IP address without any trusted third parties.

9.5 Comparison among the Defense Mechanisms

In this section, we analyze the defense mechanisms along with their capabilities and limitations. Table 9.1 lists the defense mechanisms and their capabilities to prevent various attacks along with their merits and demerits.

9.5.1 MH-HA security

Mobility protocols assume that there exists a trust relationship between the MH and the HA. Based on this assumption, IPsec ESP protocol can be used to protect against the security threats between the MH and HA. The required security associations in IPsec protocol can be established through IKE-based scheme which helps in sharing secret keys required for cryptographic algorithms. Thus, use of IPsec and IKE can secure the control traffic as well as the data traffic between the MH and HA. Thus it protects against certain types of traffic analysis and provides data confidentiality.

However, use of ESP does not protect against misbehaving MH that may use spoofed CoA in BU to launch DoS attacks to the HA. Moreover, the authentication based on IPsec and IKE can be applied for securing control and data traffic between the MH and CN (which might be any IPv6 nodes in the public Internet). This

Table 9.1: Security threats and defense mechanisms for IP-mobility protocols.

Defensive mechanism	**Protection from**	**Benefits**	**Limitations**
IPsec and IKE	Attack on traffic between MH and HA	Strong security measures, Ensures authentication and data confidentiality	High CPU overhead, does not protect against misbehaving MH, assumes a trust relationship between MH and HA
Return routability	Attack on control traffic between MH and CN	Infrastructure-less scheme, less CPU processing	Weak authentication, Does not protect from attackers on the path between HA and CN, can malfunction due to race condition
Keeping nodes stateless	Resource exhaustion attacks	Helps in avoiding DoS attacks	May introduce additional delay for legitimate traffic
Use of Sequence number	Replay attack	It is enough if sequence number can be stored in stable storage	Replay attack is possible after reboot or turnover at 16 bit boundary if stable storage is not available
Use of access control rules for Routing headers	IP spoofing, MITM, traffic redirection	Attacker cannot hide identity deceiving firewalls	additional processing for gateway router
Use of shorter binding lifetime	Replay attack, HA poisoning	Up-to-date entry in binding cache	Frequent refreshing updates wastes bandwidth
Use of CGA	Bombing attack, MITM, traffic redirection	Prevents IP-spoofing and hard for attackers to target	Higher complexity and higher CPU processing
Blocking legitimate BU	MH may keep on trying to send BU	Can avoid HA poisoning or session hijacking	Too much overhead on the MH

is because currently there exists no global infrastructure-based solution that can authenticate any IPv6 node. It is not also realistic to propose and build such a global infrastructure. In addition, processing overhead of the IPsec/IKE-based schemes are quite high and can be a big issue for mobile devices with limited processing capability.

9.5.2 MH-CN security

Since no global authentication infrastructure exist to authenticate all IPv6 nodes, mobility protocols encourage use of return routability protocol, an end-to-end authentication technique to mitigate attacks on control packets (BU) between the MH and CN. This makes sure that the MH sending the BU has the right to use the CoA. However, this authentication scheme is not as strong as IPsec/IKE-based scheme. Vulnerabilities are possible if the attacker is on the path between HA and CN. The RR protocol actually aims at limiting the number of potential attackers for a target node.

9.5.3 Protection against routing header related issues

To protect against manipulation using routing headers, several rules are applied while processing such headers. This ensures sure that attackers cannot deceive IPv6 nodes hiding their own identity. However, additional processing is required by the gateway routers which may slow down their operation.

9.5.4 Other protection mechanisms

To protect against resource exhaustion attacks, the MH or CN may abstain from keeping states information. Therefore, the MH or CN will not have to keep track of the current states of the half-open requests, thereby saving its resources and preventing possible DoS attack on the victim node. However, legitimate control traffic will require more time for processing.

Use of sequence number in the control traffic (BU) may foil the replay attack if the sequence number can be stored between reboots. In addition, the binding entry lifetime should be kept small to prevent HA poisoning by overwriting the entry with wrong CoA, thereby leading to MITM or traffic redirection attack.

Nodes with fixed IP addresses are more vulnerable to attacks, such as, bombing attack, traffic redirection attack, MITM attack. To ensure that a IPv6 node does not become a victim such attacks, it may choose to change its IP address periodically or use CGA.

To mitigate the attack on the MH's radio access network, the MH may keep trying to send BU message in spite of failures in several attempts until an acknowledgement is received. This will ensure the binding entry in the HA or the CN is not corrupted by the attackers. However, this will impose additional overhead on the low-end mobile devices.

9.6 Security Solutions for SIGMA

SIGMA was proposed as a transport layer mobility solution that ensures seamless connectivity of mobile nodes with the Internet. It exploits the multi-homing feature of STCP to achieves seamless handover and uses a make-before-break strategy so that the ongoing communication is not interrupted during handover. The MH acquires a second IP-address into the association and continue communication without much loss of data after the handover. Thus, the main advantages of SIGMA are the least handoff latency and packet loss compared to Mobile IPv6.

In SIGMA, traffic between MH and the CN follow a direct route avoiding the triangular routing of Mobile IP. However, this performance improvement of SIGMA introduces several security vulnerabilities. SIGMA requires binding updates to be exchanged between MH and CN to maintain direct routing between them. Moreover, MH must update the LM after moving to a new access network. These binding updates along with other mobility signaling may result in several security issues, including session hijacking, bombing attack, various denial of Service attacks, etc. To secure the MH, CN or the LM from these attacks, security measures must be enforced.

9.6.1 Securing MH-LM signaling

To secure the signaling between the MH and the LM, the binding update must be protected. In this section, we explain the mechanism that can be used to protect MH-LM communication against possible attacks. The MH-LM communication can be protected by adopting defense mechanism based on the ESP protocol.

In SIGMA, the MH takes advantage of multiple IP addresses into the SCTP association to ensure seamless handover. The MH has an ID for identifying its ongoing session to the upper layer protocols and this ID is unique and does not change with the change of the CoA. When the MH reaches the overlapping area between two cells, the MH acquires the second IP address from the new access network and informs the LM to "add" the second IP address into the association though the SCTP dynamic address reconfiguration option [6] with a value of ADD_IP. When the MH moves further inside into the new access network, the added (second) IP address is set as the primary address and the MH informs it through another BU with SET_PRIMARY option. At this point, if the old IP address is no longer valid, it is deleted from the association by using DELETE_IP option.

To secure the above mentioned MH-LM signaling from security threats, defense mechanism based on the ESP protocol can be used. Figure 9.15 shows the exchange of secured BU between the MH and the LM while the MH reaches the overlapping region and acquires a new CoA. It is assumed that the MH has a prior relationship with the LM and has a pre-established security associations with the LM.

Figure 9.16 shows the format of the secure BU packet (exchanged between MH and LM) for SIGMA protocol where the source address is the MH's CoA address and the destination address is the LM's address. The MH identity option field contains

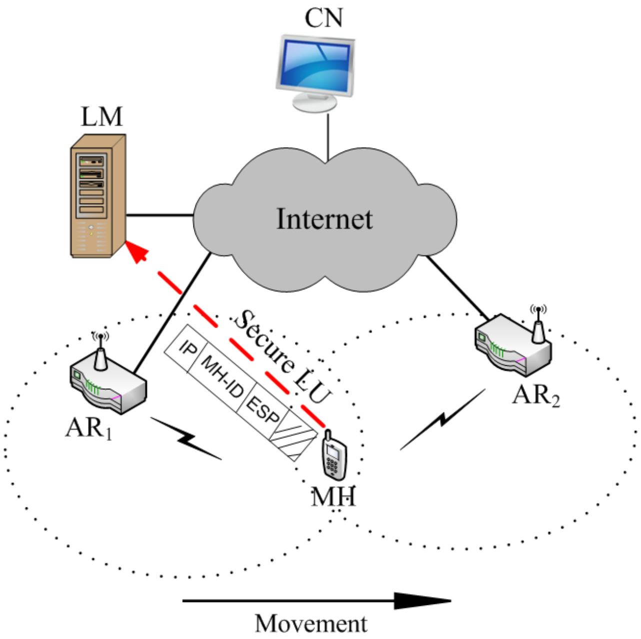


Figure 9.15: Securing the MH-LM signaling in SIGMA protocol.

the MH's ID (that might be a fixed domain name or some other identification [3]) which replaces the source address field (of IPv6 header) when it reaches the LM and thus LM identifies that the binding update came from the MH. The LM then checks the type field in BU options to see whether this BU is intended for add, delete or set primary options. The BU type field is subdivided into three bits: A, S, and D bits, meaning add this address into the association, set this address as the primary address, and delete this address from the association, respectively. The exact meanings of the all the valid bit patterns of the type field are listed in Table 9.2. For example, if the value of the type field is "100", the LM adds the CoA into the SCTP association.

As we can see in Fig. 9.16, the CoA appears in the IPv6 header and is not covered by ESP header. However, the LM must authenticate that the BU came from an authorized MH that can change its CoA. Therefore, the Alternate CoA

Table 9.2: Secure SIGMA binding update type bits and their meanings.

Type bits **A**	 **S**	 **D**	**Meaning**
1	0	0	Add this CoA into the association and do not set it primary
0	1	0	Set this CoA as the primary address
1	1	0	Add this CoA into the association and set it primary
0	0	1	Delete this CoA from the association

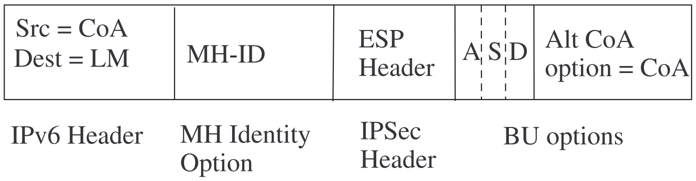
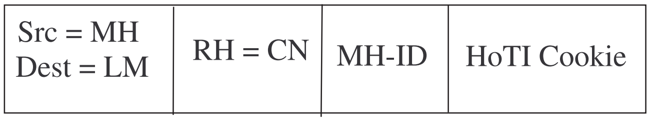


Figure 9.16: Binding update (MH-HA) packet format with IPsec Header.

mobility option are used in SIGMA that contains the MH's CoA in the BU as shown in Fig. 9.16.

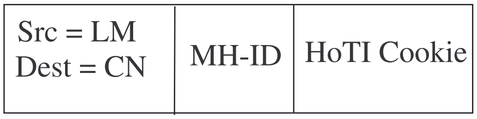
Our proposed mechanism to secure MH-LM signaling is based on the ESP protocol, and is therefore very difficult to break by the intruders. It protects certain types of traffic analysis and also provides confidentiality.

The session between the MH and CN may still be stolen if the wireless access network is jammed by the malicious agent, thereby blocking the MH from sending legitimate secure BU packets to the LM to update the CoA field into the association. The refreshing BU timer (or lifetime), therefore, may come into the rescue of such attacks. The LM may drop/invalidate a binding entry from its binding cache if there is no BU or refreshing BU received within certain time period. This can protect against future attacks targeting the location database in the LM. Alternatively, if MHs are challenged and blocked to send binding updates to the LM in a hostile network (e.g., in battlefield), it is better to switch to a more reliable access link (such as satellite link) to prevent possible attacks.



Src = MH
Dest = LM
	RH = CN	MH-ID	HoTI Cookie

Figure 9.17: HoTI message format sent from the MH to the LM.



Src = LM
Dest = CN
	MH-ID	HoTI Cookie

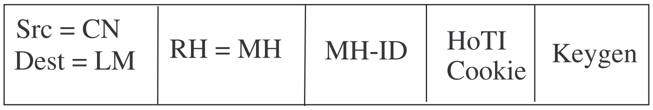
Figure 9.18: HoTI message format sent from the LM to the CN.

9.6.2 Securing MH-CN signaling

In SIGMA, the MH must notify the CN about the change of IP address into the SCTP association through sending BU to the CN. This BU must also be protected. Compared to MIPv6, the defense mechanism in SIGMA can have more time in setting up the new CoA into the association as the communication can continue with the old address.

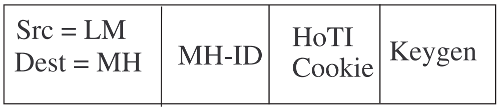
As mentioned earlier, it is not possible to build a global infrastructure to authenticate every IPv6 node of the public Internet, IPsec protocol cannot be used for securing MH-CN signaling. Therefore, we propose the involvement of the LM in securing MH-CN signaling. The proposed solution is based on the idea of return routability protocol (see Fig. 9.13).

Before sending the BU to the CN, the MH should initiate the process by simultaneously sending two messages (HoTI and CoTI) to the CN; one of them through the LM and the other directly. Each of the messages has a 64-bit cookie (randomly generated) and the MH ID. The HoTI message are sent with a routing header that contains CN's IP address as shown in Fig. 9.17. After it reaches the LM and then



Src = CN
Dest = LM
	RH = MH	MH-ID	HoTI
			Cookie
				Keygen

Figure 9.19: HoT message format sent from the CN to the LM.



Src = LM
Dest = MH
	MH-ID	HoTI
		Cookie
			Keygen

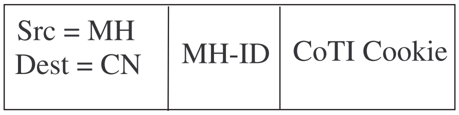
Figure 9.20: HoT message format sent from the LM to the MH.

the LM forwards it to the CN by removing the routing header field and modifying the source and destination IP field (src: LM, dest: CN) as shown in Fig. 9.18.

The CN receives the HoTI messages and uses a hash function to generate the keygen for the HoT message as in RR test (see Section 9.4.3). The CN replies with HoT message to the LM (see Fig. 9.19) and the LM forwards to the MH (Fig. 9.20).

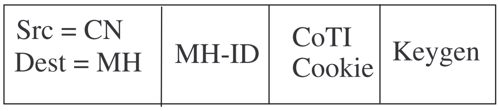
Regarding the Care-of Test Init message, the MH sends this message along with the MH's ID and 64-bit CoTI cookie as shown in Fig.9.21. The CN follows a procedure similar to HoT to compute the CoT keygen value and sends the reply along with the CoTI cookie to the MH as shown in Fig. 9.22.

The approach used to secure the MH-CN signaling in SIGMA does not require any complex cryptographic operation, thereby saving CPU power for low-end mobile



Src = MH
Dest = CN
	MH-ID	CoTI Cookie

Figure 9.21: CoTI message format sent from the MH to the CN.



Src = CN
Dest = MH
	MH-ID	CoTI
		Cookie
			Keygen

Figure 9.22: CoT message format sent from the CN to the MH.

devices. It also does not depend on any certification infrastructure. However, if the attacker is on the path between the CN and the LM, it can learn the secret that is necessary to send spoofed BU. Thus, it is a relatively weak defense mechanism with less CPU and memory requirement.

9.7 Summary

In this chapter, we have identified major security threats for the IPv6 network introduced due to IP-mobility protocol. We have explained in details possible security vulnerabilities on various components of the network, and their possible impacts on the Internet. We have also analyzed the existing defense mechanisms along with their capabilities and limitations to prevent or mitigate these security threats. Finally, we have proposed effective defense mechanism for SIGMA to protect it entities against threats from malicious agents. Our comprehensive analysis on the security threats for mobility protocols and corresponding defense mechanisms can help mobile users identify possible vulnerabilities in their networks and choose a suitable solution to protect against threats.

In the next chapter, we develop analytical model to derive mobility patterns of nodes or networks onboard vehicles roaming in city streets.

Chapter 10

Vehicular Mobility Model in City Streets

Performance evaluation of mobility protocols require knowledge of different stochastic properties of the node mobility. Mobile networks can be formed in vehicles (with onboard IP-enabled devices) that roam around the city with certain mobility pattern. However, there exists no previous work that analyzed the stochastic properties of vehicular mobility models in city streets. In this chapter, we have used a realistic mobility model for vehicular movement in the city environment and have analytically derived its fundamental stochastic properties to analyze the model thoroughly. The analytically derived properties have been validated by ns-2 simulations whose parameters have been taken from real street map data. Our developed model can be used as a generic framework for comprehensive analysis of other mobility models.

10.1 Introduction

Mobility models that can mimic the movement pattern of vehicles, are important building blocks for simulation-based studies for mobile networks (onboard a vehicle). Mobility models can help in testing and evaluating protocols related to mobile networks. Mobility should be modeled in a realistic manner to ensure match between simulation results and data from real-world deployment of vehicular mobile networks.

Vehicular mobility models [68] can be grouped into four categories: the first is synthetic model that is based on mathematical models while others are survey-based models, trace-based models and simulator-based models. Among these four variants, we are interested in synthetic models since mathematical models can help in obtaining the general trend and behavior of the mobility model. Moreover, sensitivity of certain system parameters can be easily analyzed in mathematical models which might not be possible for other types of mobility models.

In real life scenarios, while moving along city streets, vehicles do not have the freedom to travel freely along any arbitrary direction with a random speed. There are buildings, rivers, trees, etc, and the vehicular mobility environment is constrained by city streets. Moreover, vehicles must follow traffic regulations, such as stop lights, speed limits, etc. Therefore, many widely used random (synthetic) mobility models [69–72], for example, random waypoint and random direction models are not suitable at all to represent vehicular movement. In this chapter, we have developed a mathematical model to mimic the movement pattern of vehicles roaming in a city, and termed it as *Vehicular Mobility model in City Streets (VMoCS).* This model takes into account the city street map and street constraints (such as, stop lights, speed limits, acceleration, deceleration, etc.) as expected in city streets. In this model, a vehicle travels from a random initial point (at crossroad) to another destination point, and such a movement is termed as an *epoch.* We have derived several stochastic properties of the VMoCS model. To be specific, we have derived closed form expressions for the distance traveled by a vehicle during each cycle, variance of epoch length, mean epoch time, number of subnet crossing per epoch and the duration a vehicle resides under an access network (subnet). These stochastic properties are very crucial for mobility protocols since they give an estimate of the amount of signaling required for mobile networks moving around the city. Without the proper knowledge of these properties, such mobility models cannot be accurately utilized by mobility simulation tools required for the performance evaluation of vehicular mobile networks.

Several works on vehicular mobility models, both analytical [40, 73–78] and simulation-based [79–84] have been reported in the literature. However, the analytical models in [40, 73–78] lack expressions for various stochastic properties (such as mean epoch length, epoch time, and subnet residence time) that are very crucial for the utilization and applicability of any mobility model. On the other hand, as this work deals with synthetic models, simulation-based models [79–84] are not the focus of this work. Moreover, simulation studies cannot always be tested with a large range of parameter values due to resource limitation and may sometime fail to model the general trend of the problem. In contrast, analytical models represent general scenarios which provide better insights into the behavior of the system being analyzed.

To the best of our knowledge, there exists no previous work that quantitatively analyzed the stochastic properties of vehicular mobility models in city streets. Our earlier work on city mobility model [85] lacks real driving strategies, such as variation of speed in different street segments, and chance of delays in crossroads. In addition, there was no validation presented earlier. This work aims at capturing these real driving strategies into the analytical model and presents a complete and detailed mathematical analysis for vehicular mobility in city streets. We have also verified the correctness of our analytical model by comparing numerical results with simulations. Simulation parameters have been derived from real street map data (in contrast to randomly picked values) so that simulation generates results in accordance to real-world street scenarios rather than non-existing scenarios.

Our objective in this chapter is to develop an analytical model for vehicular mobility pattern that considers city street constraints and real driving strategies and to obtain its key stochastic properties for deeper understanding of its behavior. Our contributions of the work presented in this chapter are (i) developing an analytical model to formulate various stochastic properties, such as expected epoch length, variance of epoch length, expected epoch time, expected number of subnet crossing, and subnet residence time of VMoCS mobility model, and (ii) validating the analytical results through ns-2 simulations that uses real street map data.

The mathematical analysis presented in this chapter explains analytically how certain parameters of VMoCS model can influence its epoch length, epoch time, number of subnet crossings, and subnet residence time, thus giving a deeper understanding of the behavior of the model. The results obtained using VMoCS model can be used for estimating various performance metrics of mobility enabled devices in vehicles while roaming in city streets.

The rest of the chapter is organized as follows. In Section 10.2, the VMoCS mobility model is explained and its various stochastic properties are derived and analyzed. The simulation results are presented in Section 10.3. Section 10.4 concludes the chapter.

10.2 VMoCS Mobility Model

The street map considered in VMoCS model is represented by a grid shown in Fig. 10.1. Each vehicle starts movement from a randomly picked crossroad. It then chooses a destination point (another crossroad in the street map). Movement to the destination involves (at most) one horizontal and one vertical movement. Upon reaching the destination crossroad, the vehicle again chooses another destination point and the process is repeated. Each such cycle is termed as an epoch.

Following are the assumptions of the VMoCS model:

- Starting and destination points are assumed to be crossroads.

- Each street and avenue have separate speed limits.

- Streets and avenues are parallel to axes.

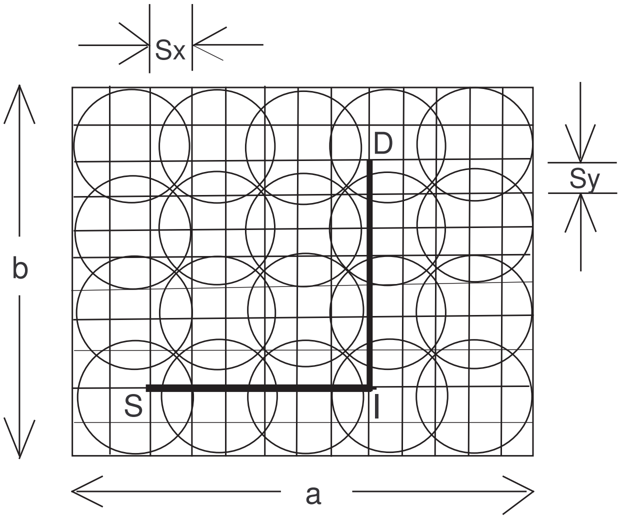


Figure 10.1: Road network in VMoCS mobility model.

- Stop light is present at every crossroad. A vehicle encounters the stop light at a crossroad with a probability.
- After each stop signal, each vehicle accelerates up to the speed limit of the street/avenue (in a fraction of the street segment) and moves with the speed unless a stop light is encountered.
- In case of encountering a stop light, the vehicle stops by decelerating in a fraction of street segment.

The street map shown in Fig. 10.1 is a rectangular shaped area of dimension $a \times b$. Let there be N_s horizontal roads (streets) and N_a vertical roads (avenues) and streets be S_y distance apart and avenues be S_x distance apart. So number of streets and number of avenues are

$$N_a = \frac{a}{S_x} + 1 \tag{10.1}$$

$$N_s = \frac{b}{S_y} + 1. \tag{10.2}$$

In the i-th epoch, the vehicle moves from point S^i to point D^i via intermediate point I^i, involving (at most) one horizontal and one vertical movement. So the movement pattern in each epoch of VMoCS model can be represented by (S^i, V_x^i,I^i), (I^i, V_y^i, D^i), where V_x^i and V_y^i are speed limits of the two streets. In an epoch where the starting point and destination point are on the same (horizontal or vertical) road segment, the intermediate point I^i coincides with D^i.

In the following subsections, we derive five stochastic properties of the VMoCS mobility model. First, we derive the expression for expected epoch length that gives an estimate of the distance traversed by the vehicle in an epoch. Then the expression for variance of epoch length is derived and it gives us a measure of how the expected epoch length varies, which in turn gives some indication of the randomness of the VMoCS model. Next, the mean epoch time is derived considering the restrictions in city streets (such as speed limit, stop light). Next, we derive the number of subnet crossing in an epoch considering circular shaped cells. Finally, the mean residence time in a cell (or subnet) is computed.

10.2.1 Epoch length

In each epoch, the vehicle moves along (at most) two road segments. Let L_x^i and L_y^i be the lengths of those two road segments during i-th epoch. So $L_x^i = |I_x^i - S_x^i|$ and $L_y^i = \left|D_y^i - I_y^i\right|$, where P_x and P_y are the x and y-coordinates of the point P. So the total distance covered by the vehicle during i-th epoch is as follows:

$$L^i = L_x^i + L_y^i. \tag{10.3}$$

It should be noted that the destination point (D^i) of one epoch is the starting point (S^{i+1}) of the next epoch. Hence, successive epochs are not totally independent of one another. But distances covered by odd epochs (i.e., first, third, fifth and so on) are independent of one another. Therefore, L^1, L^3, L^5, $\ldots$, L^{2n+1} are independent of one another. A similar situation holds for L^2, L^4, L^6, $\ldots$, L^{2n}. We can combine these two series and it does not change the asymptotic nature of the total distance covered by the vehicle [69].

Let us number each of the avenues as *1, 2, ...,* N_ath avenue from left to right and each of the streets as *1, 2, ...,* N_sth street from top to bottom. So while considering the horizontal movement path along *SI* street segment (Fig. 10.1), both the points S and I can be selected from a street having a maximum of N_a discrete points. The probability mass function (pmf) of a randomly selected point's location on a horizontal line of N_a discrete points is given by,

$$f_{P_x}(x) = \begin{cases} \frac{1}{N_a}, & \text{when } x \ \in \ \{1, 2, ...N_a\} \\ 0, & \text{otherwise.} \end{cases} \tag{10.4}$$

Since the random selection of both the points is independent of each other, their joint pmf is as follows:

$$\begin{aligned} f_{P_{x_1},P_{x_2}}(x_1, x_2) &= f_{P_{x_1}}(x_1).f_{P_{x_2}}(x_2) \\ &= \begin{cases} \frac{1}{{N_a}^2}, & \text{when } x_1, x_2 \ \in \ \{1, 2, ...N_a\} \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

10.2.1.1 Expected epoch length

The expected epoch length of VMoCS mobility model gives a measure of the distance covered by a vehicle in an epoch on the average which is required while analyzing vehicular mobility and handover protocols.

Let us first find out the expected length along the street (i.e., along horizontal direction). Let it be L_x. The values of L_x can be found from the following $N_a \times N_a$ matrix whose entries are given by,

$$M(i, j) = \ S_x \mid i - j \mid, \text{where } 1 \leq i, j \leq N_a. \tag{10.5}$$

$$\begin{bmatrix} 0 & S_x & 2S_x & .. & (N_a-1)S_x \\ S_x & 0 & S_x & .. & (N_a-2)S_x \\ 2S_x & S_x & 0 & .. & (N_a-3)S_x \\ 3S_x & 2S_x & S_x & .. & (N_a-4)S_x \\ . & . & . & .. & . \\ . & . & . & .. & . \\ (N_a-2)S_x & (N_a-3)S_x & (N_a-4)S_x & .. & S_x \\ (N_a-1)S_x & (N_a-2)S_x & (N_a-3)S_x & .. & 0 \end{bmatrix}$$

The possible values of L_x are 0, S_x, $2S_x$, $3S_x$, $\ldots$, $(N_a-1)S_x$ and it depends on the location of the starting and destination points of an epoch. Therefore, the expected value of L_x can be obtained as follows:

$$\begin{aligned} E(L_x) &= \frac{1}{{N_a}^2}\Bigg(N_a \times 0 + 2\Big((N_a-1)S_x + (N_a-2)2S_x + .. + 2(N_a-2)S_x + 1(N_a-1)S_x\Big)\Bigg) \\ &= \frac{2S_x}{{N_a}^2}\sum_{i=1}^{N_a-1}(N_a-i)i \qquad (10.6) \\ &= \frac{S_x({N_a}^2-1)}{3N_a}. \end{aligned}$$

Substituting $a = S_x(N_a-1)$ in Eqn. (10.6), we get

$$E(L_x) = \frac{a(N_a+1)}{3N_a}. \qquad (10.7)$$

Similarly, for movement along avenues, the expected value of L_y can be obtained as follows:

$$E(L_y) = \frac{b(N_s+1)}{3N_s}. \qquad (10.8)$$

Therefore, the expected epoch length can be obtained by adding Eqns. (10.7) and (10.8) as follows:

$$E(L) = \frac{a(N_a+1)}{3N_a} + \frac{b(N_s+1)}{3N_s}. \qquad (10.9)$$

In Fig. 10.2, expected epoch length (per unit length of road network), i.e., $E(L)/a$ of VMoCS model and classical Random Waypoint (RWP) model [69] is plotted against b/a. For VMoCS model, number of streets (N_s) is obtained by $N_s \approx (b/a)N_a$ (see Eqns. (10.1) and (10.2)). For the RWP model, the expression for the expected

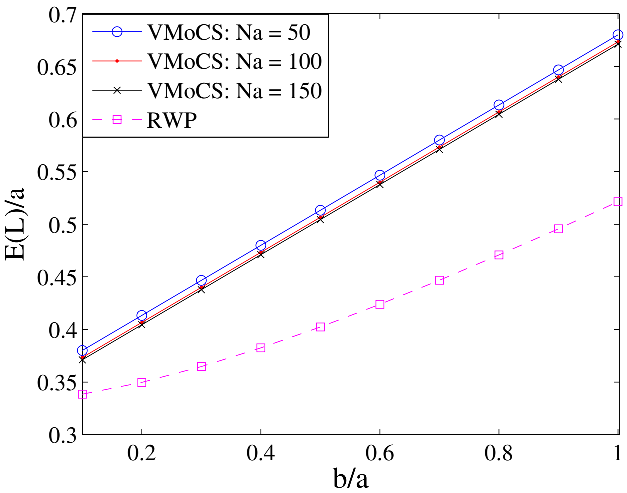


Figure 10.2: Expected epoch length of VMoCS model within an $a \times b$ rectangular road network.

epoch length has been taken from [69]. Here, the general trend of all these graphs is that the value of $E(L)/a$ increases with the increase of b/a ratio. That means expected epoch length (per unit length of road network) increases as the grid (for road network) becomes more homogeneous (towards a square shape). This is because for a fixed value of a, increased value of b/a implies larger grid dimension, resulting in higher mean epoch length. For example, the expected epoch lengths of VMoCS model for an $a \times a$ square shaped grid are $E(L) = 0.68a$ and $0.6711a$ for $N_a = 50$ and 150, respectively whereas that of RWP model is $E(L) = 0.5214a$. For an $a \times (a/2)$ rectangular road network, $E(L)$ is $0.5133a$ and $0.5044a$ for $N_a = 50$ and 150, respectively for VMoCS model whereas that of RWP model is $E(L) = 0.4024a$. Thus, it is found that the expected epoch length of RWP model is much less than that of VMoCS model (irrespective of the value of N_a). This is because in RWP model mobile node uses straight (unrestricted) route between source and destination point (in an epoch) which is not the case for vehicle movement in VMoCS model.

For large values of N_a and N_s, Eqn. (10.9) reduces to

$$E(L) = \frac{a}{3} + \frac{b}{3}. \tag{10.10}$$

For a square shaped grid of dimension $a \times a$, $E(L) = 2a/3$.

10.2.1.2 Variance of epoch length

The variance of epoch length implies how the epoch lengths vary about the mean epoch length. To compute the variance, let us first compute the second moment of L_x as follows:

$$\begin{aligned} E(L_x^2) &= \frac{1}{{N_a}^2}\Big[N_a \times 0^2 + 2\Big((N_a-1)S_x^2 + (N_a-2)(2S_x)^2 + \ldots + 1\{(N_a-1)S_x\}^2\Big)\Big] \\ &= \frac{2S_x^2}{{N_a}^2}\sum_{i=1}^{N_a-1}(N_a-i)i^2 \\ &= \frac{S_x^2(N_a^2-1)}{6}. \end{aligned} \tag{10.11}$$

Again, substituting $a = S_x(N_a-1)$ in Eqn. (10.11), we get

$$E(L_x^2) = \frac{S_x(N_a+1)a}{6}. \tag{10.12}$$

Similarly,

$$E(L_y^2) = \frac{S_y(N_s+1)b}{6}. \tag{10.13}$$

Therefore, the second moment of L can be obtained by adding Eqns. (10.12) and (10.13) as follows:

$$E(L^2) = \frac{S_x(N_a+1)a}{6} + \frac{S_y(N_s+1)b}{6}. \tag{10.14}$$

In Fig. 10.3, $E(L^2)/a$ values of VMoCS model are shown for various number of avenues and for $S_x = S_y = 50$ to 200 meters. For an $a \times a$ square shaped road network, $E(L^2) = 3.4a$ km and $13.4a$ km for $\{N_a, S_x\} = \{50, 200 \text{ m}\}$ and $\{200, 200 \text{ m}\}$, respectively. For an $a \times (a/2)$ rectangular shaped network, $E(L^2) = 2.1333a$ km and $8.3833a$ for $\{N_a, S_x\} = \{50, 200 \text{ m}\}$ and $\{200, 200 \text{ m}\}$, respectively. Thus, the second moment of L increases with the increase of S_x and N_a which follows Eqn. (10.14).

Now the variance of L_x can be obtained as,

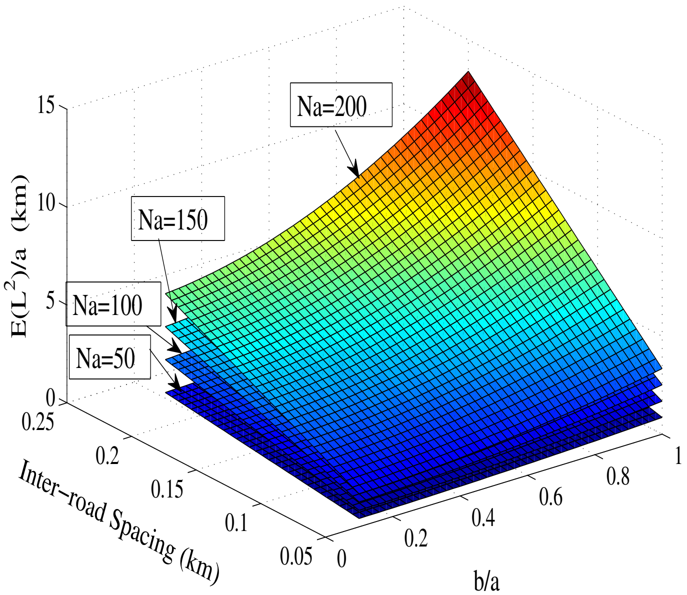


Figure 10.3: $E(L^2)/a$ of VMoCS model within an $a \times b$ rectangular road network for various numbers of avenues.

$$\begin{aligned} V(L_x) &= E(L_x^2) - (E(L_x))^2 \\ &= \frac{S_x^2(N_a^2-1)}{3}\left(\frac{1}{2} - \frac{N_a^2-1}{3N_a^2}\right). \end{aligned} \tag{10.15}$$

For large value of N_a, $N_a^2 - 1 \approx N_a^2$. Hence, Eqn. (10.15) reduces to

$$V(L_x) = \frac{S_x^2(N_a^2-1)}{18} = \frac{E(L_x^2)}{3}. \tag{10.16}$$

Similarly,

$$V(L_y) = \frac{S_y^2(N_s^2-1)}{18} = \frac{E(L_y^2)}{3}. \tag{10.17}$$

Therefore, the variance of epoch length can be obtained by adding Eqns. (10.16) and (10.17),

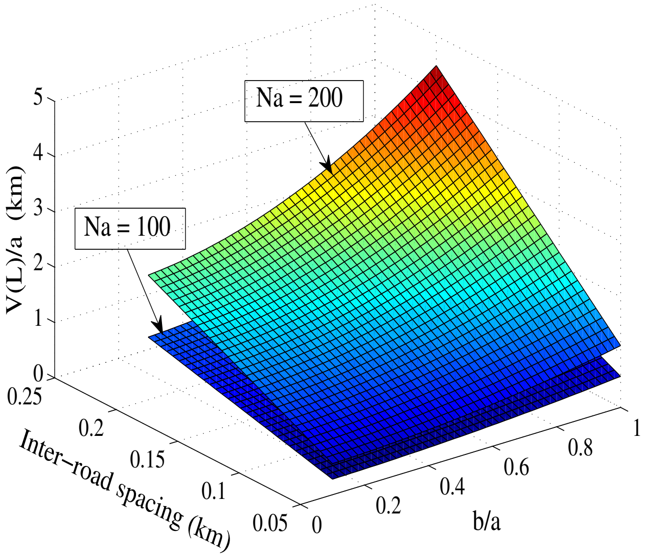


Figure 10.4: $V(L)/a$ of VMoCS model within an $a \times b$ rectangular road network for various numbers of avenues.

$$V(L) = \frac{S_x^2(N_a^2 - 1)}{18} + \frac{S_y^2(N_s^2 - 1)}{18}. \tag{10.18}$$

In Fig. 10.4, $V(L)/a$ values of VMoCS model are shown for various number of avenues and for $S_x = S_y = 50$ to 200 meters. For an $a \times a$ square shaped road network, $V(L) = 2.2444a$ km and $4.4667a$ km for $\{N_a,\ S_x\} = \{100,\ 200\ \text{m}\}$ and $\{200,\ 200\ \text{m}\}$, respectively. For an $a \times (a/2)$ rectangular road network, $V(L) = 1.4056a$ km and $2.7944a$ km for $\{N_a,\ S_x\} = \{100,\ 200\ \text{m}\}$ and $\{200,\ 200\ \text{m}\}$, respectively. Thus, it is found that the variance of VMoCS model increases with the increase of S_x (inter-road spacing) and N_a (number of avenues), since $V(L)$ is proportional to the square of S_x and N_a.

10.2.2 Epoch time

The next step is to find out the total time required for a vehicle (mobile node) to complete an epoch. We have considered possible restrictions in the city streets, such as stop lights, speed limits. We have assumed an average delay of τ sec at each crossroad with a probability of ϕ of encountering it, that is, the vehicle will encounter a stop light in a crossroad with a probability ϕ.

There is an average of $E(L_x)/S_x$ crossroads while moving along the streets, and $E(L_y)/S_y$ crossroads while moving along the avenues in an epoch. So the total delay regarding stoppages in crossroads can be estimated as follows:

$$T_{xroad} = \phi\tau\left(\frac{E(L_x)}{S_x} + \frac{E(L_y)}{S_y}\right). \tag{10.19}$$

After encountering the stoplight at a crossroad, the vehicle starts accelerating up to the speed limit. Let us assume that the vehicle reaches the speed limit after traversing the χ fraction of the road segment. Similarly, there are deceleration phase before each stoplight which covers some χ (let) fraction of the street segment. Therefore, $(1-2\chi)$ fraction of the road segment is traversed at the maximum speed which is corresponding street's speed limit. Let V_x^p be the speed limit of the p-th street. For the acceleration phase, the average speed is $(0+V_x^p)/2 = V_x^p/2$. Similar is the case for the deceleration phase.

Therefore, time required to travel along the horizontal road (street) in an epoch (excluding the delay in crossroads) is the sum of the following:

- At the beginning of each epoch, the χ fraction of the first road segment (S_x) is traversed at a speed of $V_x^p/2$. This requires a time of $\chi S_x/(V_x^p/2)$.
- In the last (horizontal) road segment, the vehicle stops to take a (left/right) turn to the avenue. Hence, the end portion (χ fraction) of the last road segment (S_x) is traversed at a speed of $V_x^p/2$. This also requires a time of $\chi S_x/(V_x^p/2)$.

- In an epoch, an average of $E(L_x)/S_x$ street segments is traversed by the vehicle. The length of the middle parts of each street segment is $(1-2\chi)S_x$ which is traversed at the speed limit (V_x^p). So the time required to travel the middle parts of all the horizontal road segments in an epoch is $(1-2\chi)S_x(E(L_x)/S_x)/V_x^p$.

- The vehicle encounters stop signal at each crossroad with a probability ϕ. So for $\phi(E(L_x)/S_x)$ street segments, the end part will have deceleration phase and the next start part will have acceleration phase. These two parts will be traversed at an average speed of $V_x^p/2$. So the time required for acceleration and deceleration phase (near the crossroad with stop light present) is $\phi(\chi+\chi)(E(L_x)/S_x)S_x/(V_x^p/2)$.

- In crossroads where vehicles do not encounter stop lights, the crossroad (along with the end and the next start part) is traversed at the corresponding speed limit. So the time required to travel those crossroads (including the end and start part) is $(1-\phi)(\chi+\chi)(E(L_x)/S_x)S_x/V_x^p$.

Therefore, the time required to travel along the horizontal road (street) in an epoch (excluding the delays in crossroads) can be obtained as follows:

$$\begin{aligned}
E(T_x) &= \frac{\chi S_x}{V_x^p/2} + \frac{\chi S_x}{V_x^p/2} + (1-2\chi)\frac{S_x}{V_x^p} \times \frac{E(L_x)}{S_x} + \phi(\chi+\chi)\frac{E(L_x)}{S_x} \times \frac{S_x}{V_x^p/2} \\
&+ (1-\phi)(\chi+\chi)\frac{E(L_x)}{S_x} \times \frac{S_x}{V_x^p} \\
&= \frac{4\chi S_x}{V_x^p} + \frac{E(L_x)}{V_x^p}((1-2\chi)+4\phi\chi+2\chi(1-\phi)) \\
&= \frac{4\chi S_x}{V_x^p} + \frac{E(L_x)}{V_x^p}(1+2\phi\chi).
\end{aligned} \tag{10.20}$$

Similarly, the time required to travel the vertical road segment (excluding the delays in crossroads) in an epoch is

$$E(T_y) = \frac{4\chi S_y}{V_y^q} + \frac{E(L_y)}{V_y^q}(1+2\phi\chi). \tag{10.21}$$

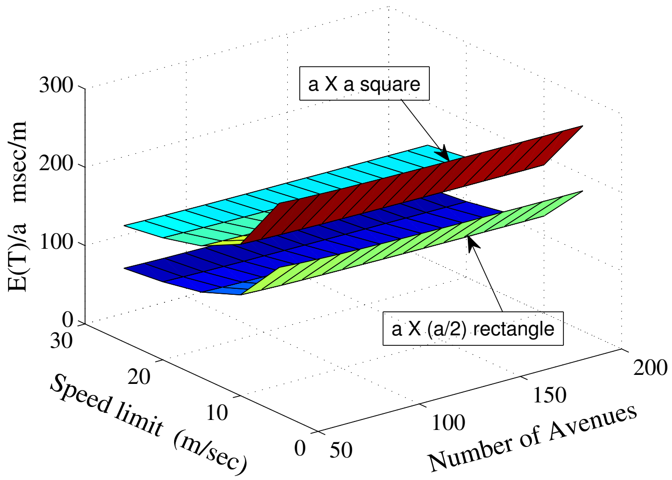


Figure 10.5: Expected epoch time of the VMoCS model as a function of speed limit and number of avenues (streets).

Hence, the total epoch time can be obtained by adding Eqn. (10.19), (10.20) and (10.21):

$$E(T) = T_{xroad} + E(T_x) + E(T_y). \tag{10.22}$$

In Fig. 10.5, the expected epoch time of the VMoCS mobility model is shown for an $a \times a$ square and an $a \times (a/2)$ rectangular road network. Here we have assumed $V_x^p = V_y^q = V^{max}$. We find that the expected epoch time decreases with the increase of speed limit which is quite obvious. On the other hand, expected epoch time increases with the increase in number of avenues (N_a) and streets (N_s). Higher value of N_a means larger grid dimensions for fixed inter-road spacing (see Eqns. (10.1) and (10.2)), resulting in higher epoch time. For an $a \times a$ square road network, $E(T)/a = 262.17$ milisec/meter and 145.75 milisec/meter when $\{V^{max}, N_a\}$ = {5 m/s, 200} and $\{V^{max}, N_a\}$ = { 25 m/s, 200}, respectively. For an $a \times (a/2)$ rectangular road network, $E(T)/a$ = 178.96 milisec/meter and 91.02 milisec/meter when $\{V^{max}, N_a\}$ = {5 m/s, 200} and $\{V^{max}, N_a\}$ = {25 m/s, 200}, respectively.

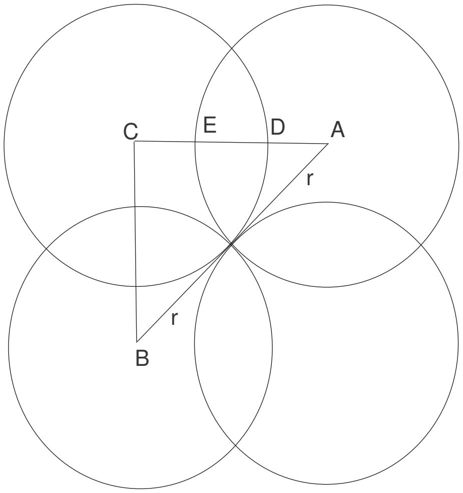


Figure 10.6: Subnet overlapping among the cells.

10.2.3 Number of subnet crossings

Let us consider that the road network of dimensions $a \times b$ (Fig. 10.1) is covered by Access Points (AP); let there be n rows of APs and m APs in each row. In total, there will be mn APs to cover the rectangular area. Let the radio coverage area of each AP be a circular region of radius r and two successive APs overlap at a maximum length of l along its diameter. So we have,

$$a = 2mr - (m-1)l \tag{10.23}$$

$$b = 2nr - (n-1)l. \tag{10.24}$$

Our aim is to figure out the average number of cell boundaries a vehicle crosses during an epoch. Let the radius r of each subnet be greater than the inter-road spacing, i.e., $r > S_x$ and $r > S_y$. In Fig. 10.6, the length of $AB = 2r$, and let $AC = x = BC$, $DE = l$. Since the cells are parallel to axes, we find that $x = \sqrt{2}r = r + r - l$. Hence,

$$l = (2 - \sqrt{2})r. \tag{10.25}$$

Now putting the value of l in Eqn. (10.23), we get,

$$m = \left\lceil \frac{\sqrt{2}a - 2(\sqrt{2}-1)r}{2r} \right\rceil . \tag{10.26}$$

Similarly, we can have

$$n = \left\lceil \frac{\sqrt{2}b - 2(\sqrt{2}-1)r}{2r} \right\rceil . \tag{10.27}$$

For $a = 36$ km, $b = 24$ km, and $r = 0.5$ km, we have, m =51, $n = 34$. Thus, we find that the $a \times b$ rectangular road network is covered by $m \times n$ access points. Therefore, the effective cell length per access point is a/m or b/n.

Let us compute the expected number of subnet crossing during the movement along horizontal direction, i.e., movement from starting point S to the intermediate point I (Fig. 10.1). Let us assume that $a/m = b/n = KS_x = KS_y$. If the distance between these two points is between 0 to $(K-1)S_x$, i.e., less than an AP's effective coverage area, there will be at most one subnet crossing. For any distance between KS_x to $(2K-1)S_x$, there will be at most two subnet crossings, and so on. Thus, if the point S is at first avenue and point I is at N_a-th avenue, then the distance of the road segment will be $(N_a - 1)S_x$ and there will be m subnet crossings. Thus, we can find out the expected number of subnet crossings in an epoch for movement along horizontal direction as,

$$\begin{aligned}
E(C_x) = {} & \frac{2}{{N_a}^2}\Bigg[(N_a-1)+(N_a-2)+...+N_a-(K-1)+2\Big((N_a-K)+(N_a-K-1)+\dots \\
& +(N_a-2K+1)\Big)+...+m\{(N_a-(m-1)K)+..+(N_a-mK+1)\}\Bigg] \\
= {} & \frac{2}{{N_a}^2}\left[KN_a\sum_{i=1}^{m} i - \Big\{\sum_{i=1}^{K-1} i + 2K^2 + 2\sum_{i=1}^{K-1} i + ... + m(m-1)K^2 + m\sum_{i=1}^{K-1} i\Big\}\right] \\
= {} & \frac{2}{{N_a}^2}\left[KN_a\sum_{i=1}^{m} i - K^2\sum_{i=2}^{m} i(i-1) - \sum_{i=1}^{K-1} i\sum_{j=1}^{m} j\right] \\
= {} & \frac{m(m+1)K}{6{N_a}^2}(6N_a-4mK+K+3).
\end{aligned} \tag{10.28}$$

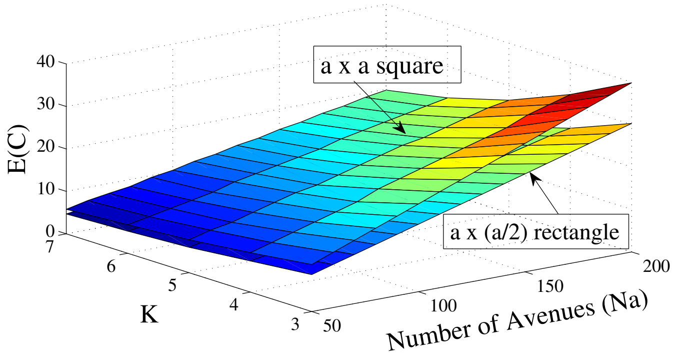


Figure 10.7: Expected number of subnet crossings of VMoCS model as a function of K and number of avenues.

Similarly, expected number of subnet crossings in an epoch for movement along the vertical direction is,

$$E(C_y) = \frac{n(n+1)K}{6{N_s}^2}(6N_s - 4nK + K + 3). \tag{10.29}$$

Therefore, the expected number of subnet crossing in an epoch can be obtained by adding Eqns. (10.28) and (10.29):

$$E(C) = E(C_x) + E(C_y). \tag{10.30}$$

In Fig. 10.7, the expected number of subnet crossings of VMoCS model is shown for an $a \times a$ square road network and an $a \times (a/2)$ rectangular road network. We find that the expected number of subnet crossings decreases with the increase in the value of K which is the ratio of the diameter of an AP to the inter-road spacing, i.e., $2r/S_x$. Higher values of K mean larger coverage area for access points, resulting in fewer number of subnet crossings. On the other hand, expected number of subnet crossings increases with the increase in number of avenues (N_a) and streets (N_s). Higher values of N_a and N_s mean larger grid dimensions for fixed inter-road spacing (see Eqns. (10.1) and (10.2)), resulting in higher number of subnet crossings.

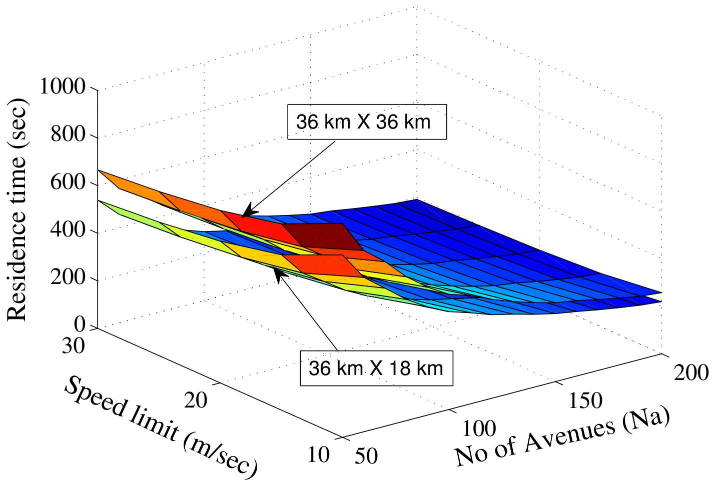


Figure 10.8: Residence time of VMoCS model as a function of speed limit and number of avenues.

For an $a \times a$ square road network, $E(C) = 19.79$ when $K = 7$, $N_a = 200$; and $E(C) = 39.85$ when $K = 3$, $N_a = 200$. For an $a \times (a/2)$ rectangular road network, $E(C) = 15.04$ when $K = 7$, $N_a = 200$; and $E(C) = 30.26$ when $K = 3$, $N_a = 200$.

10.2.4 Subnet residence time

The subnet residence time is the time duration that the mobile node (vehicle) resides under a cell. This can be obtained by dividing the mean epoch time by expected number of subnet crossing as follows:

$$T_r = \frac{E(T)}{E(C)}. \tag{10.31}$$

In Fig. 10.8, the residence time of VMoCS model is shown for 36 km × 36 km square-shaped and 36 km × 18 km rectangular-shaped road network. Here, we used $K = 5$. We find that the residence time decreases with the increase of speed limit as less time is required to cross the coverage area of an AP. Residence time also decreases with the increase in number of avenues (N_a) and streets (N_s). Higher values of N_a (N_s) result in increased $E(C)$ values, but the value of $E(T)$ is not affected much. As a result, the residence time decreases.

For a 36 km × 36 km square road network, $T_r = 899.37$ sec and 190.95 sec when $\{V^{max}, N_a\} = \{10$ m/s, $50\}$ and $\{30$ m/s, $200\}$, respectively. For a 36 km × 18 km rectangular road network, $T_r = 767.18$ sec and 154.28 sec when $\{V^{max}, N_a\} = \{10$ m/s, $50\}$ and $\{30$ m/s, $200\}$, respectively.

10.3 Simulation Study

To validate our mathematical model, we have performed ns-2 simulation [28]. Simulation environment, analysis of the results and a comparative discussion are presented in the following subsections.

10.3.1 Simulation environment

Figure 10.9 shows the topology used for the simulation. The simulation area is a 36 km × 36 km area which is covered by 1764 access routers (arranged in 42 rows and 42 columns as shown in Fig. 10.1). The hierarchical addresses of all the nodes are listed in Table 10.1. CN is the FTP source over TCP whereas the mobile nodes (in vehicles) are the TCP sinks. In each run, each mobile node (vehicle) moves in the environment covered by the 1764 ARs following the constraints of VMoCS mobility model. The HA of each mobile node also acts as an AR, they are assumed to have the address of 1.1.0-1.20.0. In each simulation run, all the 20 mobile nodes moved according to VMoCS model starting from a random location; each mobile node (vehicle) completed 10 epochs in each run, totalling 200 epochs per simulation run. We have taken the average of the results.

To estimate certain model parameters (such as inter-road spacing, speed limit) for VMoCS model, we have considered the manhattan road maps in the City of New York since it matches our assumed road network. It has been found the inter-street

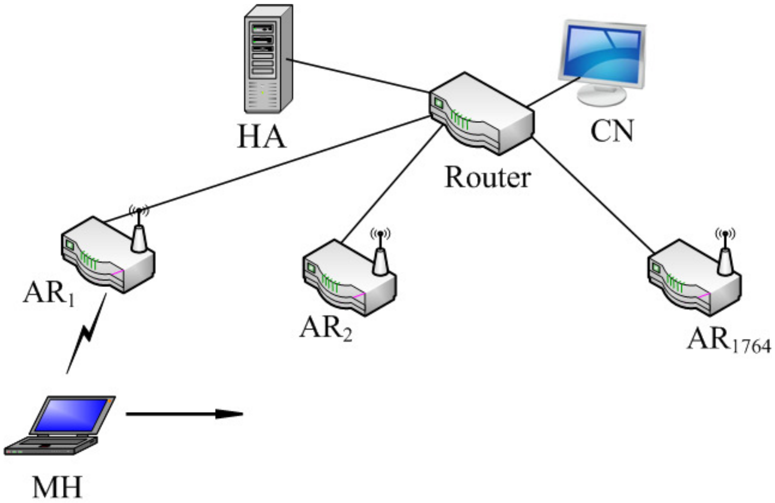


Figure 10.9: Simulation Topology.

spacing in Manhattan is around 80-100 m and inter-avenue spacing is around 320-340 m. The speed limit is roughly 30 miles/hour. We have measured these parameters by analyzing Google Maps [86] data. The signaling time cycle at the crossroads of New York city is between 45 and 120 seconds [87].

Based on the above observations, we have chosen our simulation parameter values. The default values of parameters used for simulations are listed in Table 10.2. We used a square shaped road network with a dimension of 36 km × 36 km. We assumed the inter-street and inter-avenue spacings to be the same (240 m). Therefore, number of avenues in the road network is 151. Similar is the case for number of streets. The speed limit is assumed to be 15 meter/sec (which is around 33.55 miles/hr, reasonable for city streets). Delay in each crossroad is assumed to be 50 sec (maximum value) and there is a 40% chance of encountering stop light in a crossroad. The acceleration portion of the street segment is assumed to be 10% of that segment. Similar is the case for deceleration phase. Each access point's transmission range is 600 m. IEEE 802.11b standard is used for wireless communications. The wireless link bandwidth is 11 Mbps and wired link bandwidth is 10 Mbps.

Table 10.1: Hierarchical addresses used for entities in the simulation.

Node Type	**Hierarchical addresses**
CN	0.0.0
Router	1.0.0
AR_1 - AR_{1764}	1.1.0 - 1.1764.0
MH_1 - MH_{20}	1.1.1 - 1.20.1

Table 10.2: Values of parameters used in the simulation.

Simulation Parameters	**Values**
Simulation area	36 km × 36 km
Inter-road spacings (S_x or S_y)	240 meter
Speed limit of each street/avenue	15 meter/sec
Number of Avenues (Streets)	151
Prob. of encountering stoplight at crossroad (ϕ)	0.4
Average delay in crossroad (τ)	50 sec
Portion of street segment used for acceleration (χ)	0.1
Number of MHs	20
Number of Epochs per MH per simulation	10
Wireless range	600 m
Wired link BW	10 Mbps
Wired link delay	1.8 ms
Wireless (802.11b) link BW	11 Mbps
Number of Access Points	1764

10.3.2 Simulation results

The results obtained from the analysis of the simulation traces are presented in this subsection. We have measured the mean epoch length, epoch time, number of subnet crossings, residence times of each simulation run. In addition, we have measured the average packet drop probability, handoff frequency, average end-to-end delay, number of binding updates.

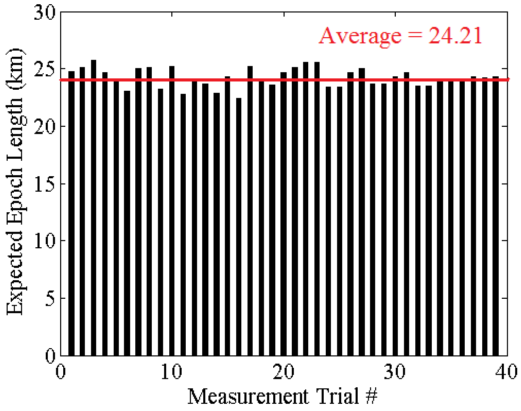


Figure 10.10: Mean epoch lengths for different simulation trials.

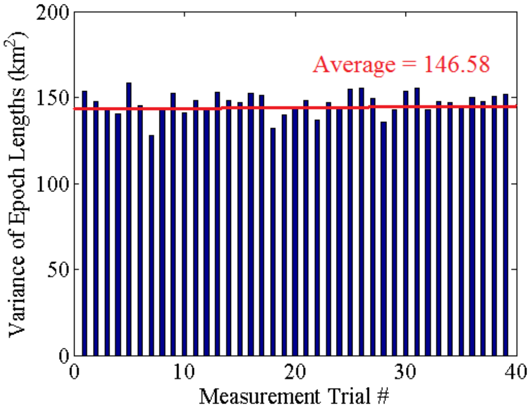


Figure 10.11: Variance of epoch lengths for different simulation trials.

10.3.2.1 Expected epoch length

Figure 10.10 shows the bar chart of epoch lengths for each simulation run (having 200 epochs iterated in each run). There are 40 such simulation results. The average epoch length of all the simulation runs is 24.2113 km which is shown using the red straight line in Fig. 10.10. The variation among the epoch length values reflect the randomness of the mobility model.

According to the theoretical model presented in Section 10.2, the expected epoch length for the road network is given by Eqn. (10.9). Hence, $E(L) = E(L_x) + E(L_y)$ = 2 * 36 * (151 + 1) / (3*151) = 24.1589 km. This value is close to the average epoch length obtained from simulation results with an accuracy of 99.78%.

10.3.2.2 Variance of epoch length

During each simulation run, we had 200 epochs. We recorded the epoch lengths in each trial and computed the mean epoch length of each trial. Using the mean value, we have computed the variance of epoch lengths in each trial as shown in Fig. 10.11. The average variance of all the 40 trials is 146.58 km^2. Hence, the standard deviation for epoch lengths is 12.10 km.

Using the analytical model, the variance of epoch lengths (for $S_x = S_y = 240$ m, and $N_a = N_s = 151$) can be obtained using Eqn. (10.18) and is 145.92 km^2. Therefore, the standard deviation for a 36 km × 36 km square shaped road network is 12.079 km. Thus, the theoretical value matches the simulation one (accuracy of 99.54%).

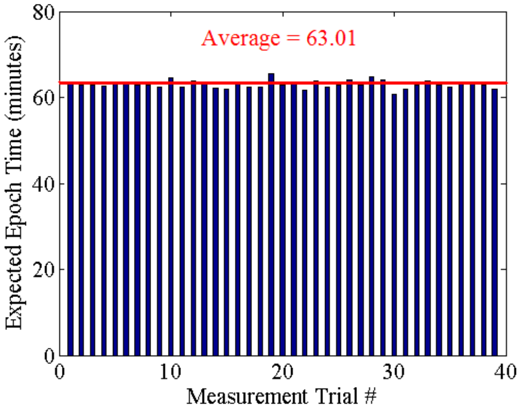


Figure 10.12: Mean epoch times for different simulation trials.

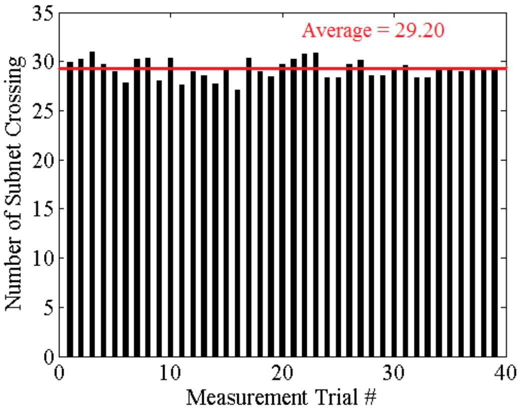


Figure 10.13: Average number of subnet crossing for different simulation trials.

10.3.2.3 Epoch time

Epoch time includes delay in the crossroads and the time to travel horizontal and vertical road segment in an epoch. The theoretical value for delay in crossroads can be obtained using Eqn. (10.19) as $T_{xroad} = 0.4 * 50 * (12079.45 / 240) *2 = 2013.2$ sec. In addition, time required to travel horizontal and vertical street movement can be computed by Eqns. (10.20) and (10.21) which are $T_x = 811.62$ sec and T_y

= 811.62 sec. Therefore, the calculated epoch time (from the analytical model) is $E(T) = 2013.2 + 811.62 + 811.62 = 3636.44$ sec $= 60.61$ min.

On the other hand, the distribution for epoch time obtained from simulations are shown in Fig. 10.12. The average value is 63.01 min which is shown using the red straight line. The simulation value of mean epoch time are close to the analytical one with an accuracy of 96.19%.

10.3.2.4 Expected number of subnet crossings

Using the analytical model, the expected number of subnet crossing can be obtained using Eqns. (10.28), (10.29), and (10.30). For the square shaped road network of 36 km × 36 km, number of streets (avenues) are 151, the value $K = a/(mS_x) = 3.57$. Using the Eqn. (10.28), we find that $E(C_x) = 14.74 = E(C_y)$ due to square road shaped road network. Hence, $E(C) = E(C_x) + E(C_y) = 29.48$.

From the simulation traces, the average number of subnet crossing in each trial is obtained and Fig. 10.13 shows the corresponding bar chart. The average value is 29.208 shown using the red straight line. This average value is close to the analytical value with an accuracy of 99.05%.

Table 10.3: Comparison between analytical and simulation results.

Type of results	E(L)	V(L)	E(T)	E(C)	T_r
Analytical	24.15 km	145.92 km^2	60.61 min	29.48	123.36 sec
Simulation	24.21 km	146.58 km^2	63.01 min	29.20	129.31 sec

10.3.2.5 Subnet residence time

The subnet residence time can be computed using Eqn. (10.31) as $T_r = 60.61$ min / $29.48 = 123.36$ sec. On the other hand, we show the subnet residence times for each simulation trial in Fig. 10.14. The distribution is almost flat and the average value

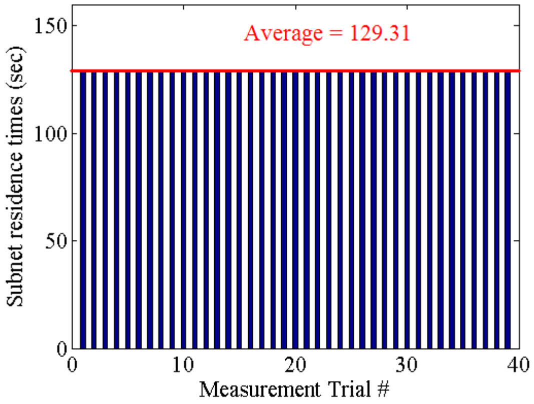


Figure 10.14: Subnet residence times for different simulation trials.

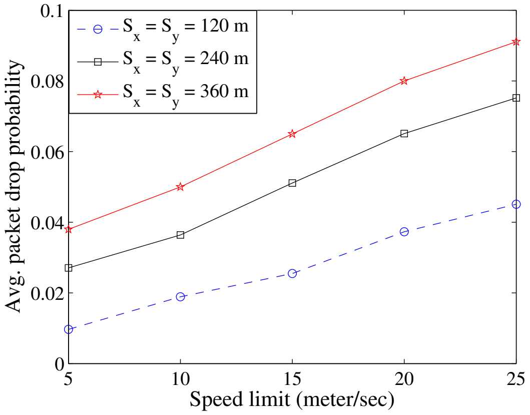


Figure 10.15: Average packet drop probability vs. speed limit for different inter-road spacing.

of all simulation trial is 129.31 sec. Thus, we find that the simulation value matches the analytical one with 95.39% accuracy.

In Table 10.3, the results obtained from the mathematical model (Section 10.2) and the ns-2 simulation are presented for a 36 km × 36 km square-shaped road network with inter-road spacing of 240 m. The theoretical values of mean epoch length, variance of epoch length, epoch time, mean number of subnet crossing and subnet residence time matches the values obtained by simulation, thus validating our analytical model.

The objective of the chapter is to derive several stochastic properties of realistic vehicular mobility model in city streets. These properties have been validated by ns-2 simulation results in Figs. 10.10 - 10.14. However, while doing ns-2 simulations, we have collected additional results related to the packet drop probability, number of handoff, and number of binding updates. These results are part of simulation results that cannot be obtained from our analytical model. These are particularly related to network protocol modeling.

10.3.2.6 Average packet drop probability

To compute the packet drop probability, we have counted the number of packets sent by CN, and number of packets received by the all mobile hosts. We have computed the packet drop probability by taking the ratio of dropped packets to the total packets sent by the CN.

In Fig. 10.15, the average packet drop probability is shown for varying speed limit of the road network. Average drop probability increases for higher speed limit which implies higher mobility rate of mobile nodes (vehicles). Moreover, the drop rate increases for higher inter-road spacings. Higher S_x value implies lower value of N_a for fixed dimension of the road network. This reduces number of possible destination points in an epoch, resulting in higher packet drop rate.

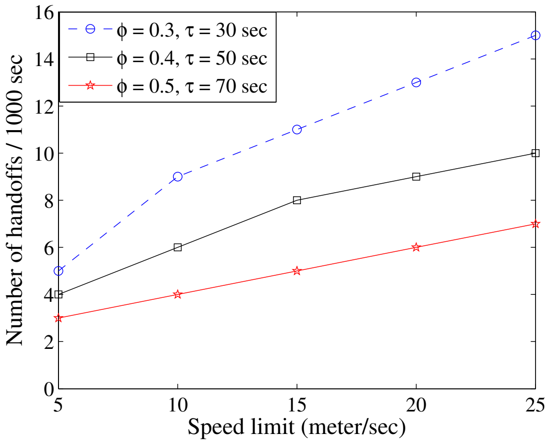


Figure 10.16: Handoff frequency vs. speed limit.

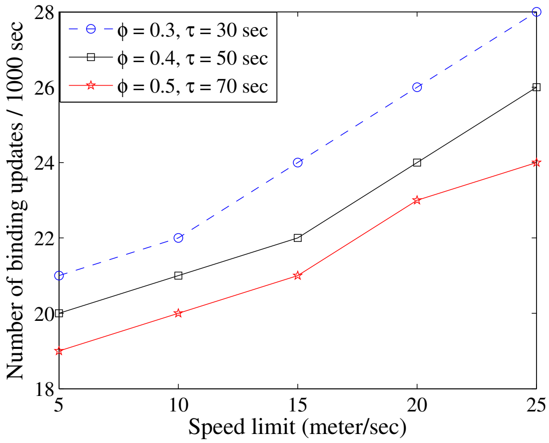


Figure 10.17: Number of binding updates vs. speed limit.

10.3.2.7 Handoff frequency

Figure 10.16 shows the number of handoff per 1000 sec of simulation time for varying speed limit of the road network with different delay time and probability of red light at crossroads. The handoff frequency increases for shorter delays ($\phi = 0.3$, $\tau = 30$ sec) in crossroads (possibly during the off-peak hour) since this allows vehicles to complete the epoch faster. Thus the mean epoch time reduces, resulting in higher handoff frequency. During the peak hours, there are large delays at crossroads with higher probability of encountering stop light ($\phi = 0.5$, $\tau = 70$ sec). This increases the total delay to complete an epoch, causing the handoff frequency to drop.

10.3.2.8 Binding updates

The simulation was run for several speed of mobile network, ranging from 5 to 25 m/s. We have counted the number of binding updates sent by the MH in every minute of simulation time as shown in Fig. 10.17. It may be noted that the binding lifetime was set to 60 sec. Therefore, if there were no regular binding update sent (due to handoff) within a time period of 60 sec, refreshing binding updates were sent by the mobile node to keep the home agent's binding cache valid. It is found that number of binding updates sent by the MH increases with higher mobility rate of the vehicles due to higher handoff rate.

10.4 Summary

In this chapter, we have analyzed a realistic vehicular mobility model that captures the real driving strategies along with possible constraints in city streets, such as stop lights at crossroad, speed limits, etc. We have performed a complete and detailed analysis of the model deriving the expressions for its various stochastic properties: expected epoch length, variance of epoch length, expected epoch time, expected number of subnet crossings and subnet residence time. The analytical model has been validated by ns-2 simulations with parameters chosen from real street map data. Our work can help in estimating various performance metrics of IP-enabled devices or mobile networks on-board vehicles in city streets.

Chapter 11

Conclusion

In this dissertation, a comprehensive evaluation of host and network mobility protocols has been performed. Specifically, we have focused on the cost, scalability, survivability and security analysis of mobility protocols. We have also proposed a dynamic scheduling algorithm to protect the mobility signaling message from getting ignored as a consequence of excessive amount of audio-video streaming data in mobile Internet. Moreover, we have proposed a multi-band mobile router architecture that aims at maximizing bandwidth utilization. Finally, we present a mathematical model to derive various stochastic properties of a realistic mobility model for mobile networks and apply the model in protocol analysis. The analysis presented in this work can help network engineers evaluate different mobility protocols quantitatively, thereby choose one that is more reliable, secure, survivable and scalable.

11.1 Summary

First, we have developed comprehensive cost analysis models to estimate total costs and efficiencies of different host and network mobility protocols and their key entities. We have defined two novel metrics, namely normalized overhead and efficiency, to compare the performance of the mobility protocols and its entities. We have presented numerical results to demonstrate the impact of increased network size, mobility rate, traffic rate and data volume on the total cost, overhead and efficiency of mobility management entities. Our results show that SIGMA (and SINEMO) incurs much lower overhead on its key entities and yields higher efficiency than HMIPv6 (and NEMO) irrespective of session lengths, network size and mobility rate. Based on the cost models, we performed scalability analysis of these mobility protocols. Our results show that the host mobility protocols (HMIPv6 and SIGMA) and network mobility protocols (NEMO and SINEMO) exhibit asymptotically identical scalability feature for complete system though the LM (of SIGMA) is found to be more scalable than the HA (of HMIPv6).

Next, we focus on the multi-class traffic analysis for mobility protocols to protect the all-important signaling traffic from getting lost due to the high volume of real-time and non-real data in the network. We propose a scheduling algorithm that gives highest priority to signaling traffic, thereby ensures its minimum loss. Based on the scheduling algorithm, we have derived closed form expressions for average queuing delay, queue occupancy, and packet drop probability of each class of traffic. Results show the impact of the node density, service rate and traffic distribution on those measures.

We proposed a novel scheduling algorithm for multi-band mobile routers that aims at maximizing utilization of available bandwidth of mobile routers. We have also derived various performance metrics of the multi-band router architecture, that have been validated by extensive simulations. Our results show that the proposed architecture can ensure maximum possible utilization through sharing of capacities among the bands.

We perform quantitative survivability evaluation of NEMO with multiple mobile routers taking into consideration possible natural or man-made failures as well as DDoS attacks. Our results show that increase in the number of mobile routers improves the performance of the mobile network by reducing the mean delay and drop probability while withstanding attack packets.

We proposed a seamless handover scheme for NEMO exploiting the multihoming feature of the mobile router. We have used experimental testbeds to measure its handoff performance (throughput, round trip time, and handoff latency) and compared it with basic NEMO. Our experimental results show that proposed multihomed scheme outperforms basic NEMO in terms of handoff delay, round trip time and throughput.

We identified major security threats for the IPv6 network due to the inclusion of mobility protocols. We have explained in details the security vulnerabilities and their possible impacts on the Internet. We have also analyzed the existing defense mechanisms along with their capabilities and limitations to prevent or mitigate these security threats. Finally, we have proposed effective defense mechanisms for SIGMA to protect against threats from malicious agents.

Finally, we have developed a mathematical model to derive various stochastic properties of a realistic mobility model for mobile networks (onboard a vehicle). We use the knowledge of the node mobility pattern for performance evaluation of host mobility protocol in ns-2 simulations. The analytically derived properties have been validated by simulations whose parameters have been taken from real street map data. Our developed model can be used as a generic framework for the comprehensive analysis of other mobility models.

In summary, we have performed a comprehensive evaluation of host and network mobility protocols. Specifically, we have focused on the cost, scalability, survivability and security analysis of mobility protocols. Our proposed dynamic scheduling algorithm aims at protecting the mobility signaling message and maximizing bandwidth utilization in multi-band system. Moreover, we have presented a mathematical model to derive various stochastic properties of a realistic mobility model for mobile networks. The analysis presented in this work can help network engineers evaluate different mobility protocols quantitatively, thereby choose one that is more reliable, secure, survivable and scalable.

11.2 Future Works

Some future research works are listed as follows:

- All the analysis in this dissertation was performed on terminal-based mobility protocols where the end node is involved in mobility signaling. To relieve low end mobile devices, network-based mobility protocols, such as, Proxy Mobile IPv6 [88] have been proposed where the infrastructure is responsible for mobility management. The analytical models developed in this dissertation can be applied for network-based mobility protocols. Moreover, a seamless network-based mobility protocol can be proposed and analyzed with respect to its total cost, scalability, security and survivability.

- The mobility model proposed in Chapter 10 can be made more realistic by incorporating importance factor for each destination (of a movement). Moreover, the total population of the city can be taken into account and average mobility pattern of the population can be derived. This can also consider peak-hour, off-peak, weekdays and weekends.

- In our analysis, we have assumed the correspondent node to be stationary. However, both the communicating nodes may be mobile, such as, soldiers or units communicating in the battlefield. Analysis of mobility protocols can be performed considering such simultaneous mobility.

- Future research may focus on the issues and challenges when the mobile node hands off between foreign networks operated by multiple service providers.

- Another issue is the survivability of location database. Survivability model can be developed to compute the recovery time of the system and the average number of packets lost due to the failure of mobility database.

Bibliography

[1] C. Perkins, D. Johnson, and J. Arkko, “Mobility support in IPv6,” IETF RFC 6275, Jul 2011.

[2] V. Devarapalli, R. Wakikawa, A. Petrescu, and P. Thubert, “NEtwork MObility (NEMO) basic support protocol,” RFC 3963, Jan 2005.

[3] S. Fu and M. Atiquzzaman, “SIGMA: A Transport Layer Handover Protocol for Mobile Terrestrial and Space Networks,” *e-Business and Telecommunication Networks, Springer*, pp. 41–52, 2006.

[4] P. Chowdhury, M. Atiquzzaman, and W. Ivancic, “SINEMO: An IP-diversity based approach for network mobility in space,” in *Second IEEE International Conference on Space Mission Challenges for Information Technology (SMC-IT)*, Pasadena, CA, Jul 17-21, 2006.

[5] H. Soliman, C. Castelluccia, K. El Malki, and L. Bellier, “Hierarchical Mobile IPv6 mobility management (HMIPv6),” IETF RFC 5380, Oct 2008.

[6] R. Stewart, Q. Xie, and M. Tuexen et al., “Stream Control Transmission Protocol (SCTP) dynamic address reconfiguration,” IETF RFC 5061, Sep 2007.

[7] S. Fu and M. Atiquzzaman, “Handover latency comparison of SIGMA, FMIPv6, HMIPv6, and FHMIPv6,” in *IEEE GLOBECOM*, St. Louis, MO, Nov 28-Dec 02, 2005.

[8] S. Fu and M. Atiquzzaman, “Hierarchical location management for transport layer mobility,” in *IEEE GLOBECOM*, San Francisco, CA, Nov 27-Dec 1, 2006.

[9] A. S. Reaz, M. Atiquzzaman, and S. Fu., “Performance of DNS as location manager for wireless systems in IP networks,” in *IEEE GLOBECOM*, St. Louis, MI, Nov 28 - Dec 2, 2005.

[10] S. Fu and M. Atiquzzaman, “Signaling cost and performance of SIGMA: A seamless handover scheme for data networks,” *Wireless Communication and Mobile Computing*, vol. 5, no. 7, pp. 825–845, Nov 2005.

[11] A. S Reaz, P. K. Chowdhury, and M. Atiquzzaman, "Signaling cost analysis of SINEMO: Seamless End-to-End Network Mobility," in *First ACM/IEEE International Workshop on Mobility in the Evolving Internet Architecture*, San Francisco, CA, Dec 1, 2006.

[12] Christian Makaya and Samuel Pierre, "An analytical framework for performance evaluation of IPv6-based mobility management protocols," *IEEE Transactions on Wireless Communications*, vol. 7, no. 3, pp. 972–983, Mar 2008.

[13] Kumudu S. Munasinghe and Abbas Jamalipour, "Analysis of signaling cost for a roaming user in a heterogeneous mobile data network," in *IEEE Globecom*, New Orleans, LA, Nov 26-30, 2008.

[14] Jong-Hyouk Lee, Sri Gundavelli, and Tai-Myoung Chung, "A performance analysis on route optimization for Proxy Mobile IPv6," in *IEEE International Conference on Communications*, Dresden, Germany, Jun 14-18, 2009.

[15] Jiang Xie and Uday Narayanan, "Performance analysis of mobility support in IPv4/IPv6 mixed wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 2, Feb 2010.

[16] J. Xie and I.F. Akyildiz, "A novel distributed dynamic location management scheme for minimizing signaling costs in Mobile IP," *IEEE Transactions on Mobile Computing*, vol. 1, no. 3, pp. 163–175, Jul 2002.

[17] C. Perkins, D. Johnson, and J. Arkko, "Mobility support in IPv6," IETF RFC 6275, Jul 2011.

[18] R. Stewart, Q. Xie, and M. Tuexen et al., "Stream Control Transmission Protocol (SCTP) dynamic address reconfiguration," IETF RFC 5061, Sep 2007.

[19] C.A. Santivanez, B. McDonald, I. Stavrakakis, and R. Ramanathan, "On the scalability of Ad hoc routing protocols," in *IEEE INFOCOM*, New York, NY, June 23-27, 2002.

[20] Sumesh J. Philip, Joy Ghosh, Swapnil Khedekar, and Chunming Qiao, "Scalability analysis of location management protocols for Mobile Ad hoc Networks," in *IEEE WCNC*, Atlanta, GA, Mar 21-25, 2004.

[21] Lubna K. Alazzawi, Ali M. Elkateeb, Aiyappa Ramesh, and Waleed Aljuhar, "Scalability analysis for wireless sensor networks routing protocols," in *22nd International Conference on Advanced Information Networking and Applications*, Okinawa, Japan, Mar 25-28, 2008.

[22] Youngjune Gwon, James Kempf, and Alper Yegin, "Scalabilty and robustness analysis of Mobile IPv6, Fast Mobile IPv6, Hierarchical Mobile IPv6, and hybrid

IPv6 mobility protocols using a large-scale simulation," in *IEEE ICC*, Paris, France, Jun 20-24, 2004.

[23] Terho Hautala, Timo Braysy, Juha Makela, Janne Lehtomki, and Tommi Saarinen, "Scalability of mobility signaling in IEEE 802.11 WLAN," in *IEEE Vehicular Technology Conference*, Orlando, FL, Oct 6-9, 2003.

[24] Hoang Nam Nguyen and Iwao Sasase, "Downlink queuing model and packet scheduling for providing lossless handoff and QoS in 4G mobile network," *IEEE Transactions on Mobile Computing*, vol. 5, no. 5, pp. 452–462, May 2006.

[25] Mohsin Iftikhar, Tejeshwar Singh, Bjorn Landfeldt, and Mine Caglar, "Multiclass G/M/1 queueing system with self-similar input and non-preemptive priority," *Computer Communications*, vol. 31, pp. 1012–1027, Mar 2008.

[26] Mohsin Iftikhar, Bjorn Landfeldt, and Mine Caglar, "Towards the formation of comprehensive SLAs between heterogeneous wireless DiffServ domains," *Telecommunication Systems*, vol. 42, pp. 179–199, Dec 2009.

[27] Gunter Bolch, Stefan Greiner, Hermann de Meer, and Kishor S. Trivedi, *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*, Wiley-Interscience, New York, NY, 1998.

[28] "The network simulation - ns-2," http://www.isi.edu/nsnam/ns/.

[29] Harkirat Singh, Julan Hsu, Lochan Verma, Scott Seongwook Lee, and Chiu Ngo, "Green operation of multi-band wireless LAN in 60 GHz and 2.4/5 GHz," in *Consumer Communications and Networking Conference (CCNC)*, Las Vegas, NV, Jan 2011.

[30] Eldad Perahia, Carlos Cordeiro, Minyoung Park, and L. Lily Yang, "IEEE 802.11ad: defining the next generation multi-gbps Wi-Fi," in *7th IEEE Consumer Communications and Networking Conference (CCNC)*, Las Vegas, NV, Jan 2010.

[31] Ian F. Akyildiz, David M. Gutierrez-Estevez, and Elias Chavarria Reyes, "The evolution to 4G cellular systems: LTE-advanced," *Physical Communication*, vol. 3, pp. 217–244, Mar 2010.

[32] Sumit Singh, Raghuraman Mudumbai, and Upamanyu Madhow, "Distributed coordination with deaf neighbors: Efficient medium access for 60 GHz mesh networks," in *IEEE INFOCOM*, San Diego, CA, Mar 2010.

[33] Yi bing Lin, Wei ru Lai, and Rong jaye Chen, "Performance analysis for dual band PCS networks," *IEEE Transactions on Computers*, vol. 49, pp. 148–159, Feb 2000.

[34] Klaus Doppler, Carl Wijting, Tero Henttonen, and Kimmo Valkealahti, "Multi-band scheduler for future communication systems," *I. J. Communications, Network and System Sciences*, vol. 1, no. 1, pp. 1–9, Feb 2008.

[35] Lochan Verma and Scott Seongwook Lee, "Multi-band Wi-Fi systems: A new direction in personal and community connectivity," in *IEEE International Conference on Consumer Electronics*, Las Vegas, NV, Jan 2011.

[36] Donald Gross, John Shortle, James Thompson, and Carl M. Harris, *Fundamentals of Queueing Theory*, Wiley-Interscience, Aug 2008.

[37] Konstantin E. Avrachenkov, Nikita O. Vilchevsky, and Georgy L. Shevlyakov, "Priority queueing with finite buffer size and randomized push-out mechanism," *Performance Evaluation*, vol. 61, pp. 1–16, Jun 2005.

[38] Guido Appenzeller, Isaac Keslassy, and Nick McKeown, "Sizing router buffers," *Computer Communication Review*, vol. 34, pp. 281–292, Oct 2004.

[39] Gahng seop Ahn, Andrew T. Campbell, Andras Veres, and Li hsiang Sun, "Supporting service differentiation for real-time and best-effort traffic in Stateless Wireless Ad Hoc Networks (SWAN)," *IEEE Transactions on Mobile Computing*, vol. 1, pp. 192–207, Sep 2002.

[40] Kuo-Tay Chen, Szu-Lin Su, and Rong-Feng Chang, "Design and analysis of dynamic mobility tracking in wireless personal communication networks," *IEEE Transactions on Vehicular Technology*, vol. 51, no. 3, May 2002.

[41] Md. Shohrab Hossain, M. Atiquzzaman, and William Ivancic, "Survivability and scalability of space networks," in *NASA Earth Science Technology Forum*, Arlington, VA, Jun 22-24, 2010.

[42] Yin Fu Huang and Min Hsiu Chuang, "Fault tolerance for home agents in Mobile IP," *Computer Networks*, vol. 50, no. 18, pp. 3686–3700, Dec 2006.

[43] Jenn-Wei Lin and Joseph Arul, "An efficient fault-tolerant approach for Mobile IP in wireless systems," *IEEE Transactions on Mobile Computing*, vol. 2, no. 3, pp. 207–220, Jul-Sep 2003.

[44] J. Jue and D. Ghosal, "Design and analysis of replicated server architecture for supporting IP-host mobility," *ACM Mobile Computing and Communications Review*, vol. 2, no. 3, pp. 16–23, Jul 1998.

[45] J. Faizan Hesham EL-Rewini and M. Khalil, "Introducing reliability and load balancing in Mobile IPv6-based networks," *Wireless Communication and Mobile Computing*, vol. 8, no. 4, pp. 483–500, May 2008.

[46] Hui Deng, X. Huang, Kai Zhang, Zhisheng Niu, and Masahiro Ojima, "A hybrid load balance mechanism for distributed home agents in Mobile IPv6," *Personal, Indoor and Mobile Radio Communications*, pp. 2842–2846, Jan 2003.

[47] Romain Kuntz, Julien Montavont, and Thomas Noel, "Multiple mobile routers in nemo: How neighbor discovery can assist default router selection," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Poznan, Poland, Sep 15-18, 2008.

[48] Poul E. Heegaard and Kishor S. Trivedi, "Network survivability modeling," *Computer Networks*, vol. 53, no. 8, Jun 2009.

[49] S. Fu and M. Atiquzzaman, "Survivability evaluation of SIGMA and Mobile IP," *Wireless Personal Communications*, vol. 43, no. 3, pp. 933–944, Nov 2007.

[50] R. Wakikawa, V. Devarapalli, G. Tsirtsis, T. Ernst, and K. Nagami, "Multiple care-of addresses registration," IETF RFC 5648, Oct 2009.

[51] Jen-Yi Pan, Jing-Luen Lin, and Kai-Fung Pan, "Multiple care-of addresses registration and capacity-aware preference on multi-rate wireless links," in *International Conference on Advanced Information Networking and Applications*, Okinawa, Japan, Mar 25-28, 2008.

[52] Romain Kuntz, "Deploying reliable IPv6 temporary networks thanks to NEMO basic support and multiple care-of addresses registration," in *International Symposium on Applications and the Internet Workshops*, Hiroshima, Japan, Jan 15-19, 2007.

[53] Xiaohua Chen, Hongke Zhang, Yao-Chung Chang, and Han-Chieh Chao, "Experimentation and performance analysis of multi-interfaced mobile router scheme," *Simulation Modelling Practice and Theory*, vol. 18, no. 4, Apr 2010.

[54] Md. Sazzadur Rahman, Outman Bouidel, M. Atiquzzaman, and William Ivancic, "Performance Comparison between NEMO BSP and SINEMO," in *IEEE GLOBECOM*, New Orleans, LA, Nov 30-Dec 4 2008.

[55] Henrik Petander, Eranga Perera, Kun-Chan Lan, and Aruna Seneviratne, "Measuring and improving the performance of network mobility management in IPv6 networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 9, pp. 1671–1681, Sep 2006.

[56] "NEPL (NEMO Platform for Linux) howto," http://www.nautilus6.org/doc/nepl-howto/.

[57] R. Kuntz, "NEMO Basic Support implementation tests at the 6th IPv6 TAHI interoperability test event," http://www.nautilus6.org/doc/tc-nemo-tahi-interop-20050207-KuntzR.txt, Feb 2005.

[58] James Kempf, Jari Arkko, and Pekka Nikander, “Mobile IPv6 security,” *Wireless Personal Communications*, vol. 29, pp. 398–414, Jun 2004.

[59] Dong Hu, Dong Zhou, and Ping Li, “PKI and secret key based mobile IP security,” in *International Conference on Communications, Circuits and Systems*, Guilin, China, June 2006.

[60] Khaled Elgoarany and Mohamed Eltoweissy, “Security in Mobile IPv6: A survey,” *Information Security Technical Report*, vol. 12, no. 1, pp. 32–43, Jun 2007.

[61] Errata Exist, S. Kent, and K. Seo, “Security architecture for the internet protocol,” IETF RFC 4301, Dec 2005.

[62] C. Kaufman, P. Hoffman, Y. Nir, and P. Eronen, “Internet Key Exchange Protocol Version 2 (IKEv2),” IETF RFC 5996, September 2010.

[63] S. Kent, “IP Authentication Header,” IETF RFC 4302, Dec 2005.

[64] S. Kent, “IP Encapsulating Security Payload (ESP),” IETF RFC 4303, Dec 2005.

[65] W. Diffie and M. Hellman, “New directions in cryptography,” *IEEE Transaction on Information Theory*, vol. 22, no. 6, pp. 644–654, Nov 1976.

[66] T. Aura, “Cryptographically Generated Addresses (CGA),” IETF RFC 3972, March 2005.

[67] Greg O’Shea and Michael Roe, “Child-proof authentication for MIPv6 (CAM),” *ACM Computer Communications Review*, vol. 31, no. 2, Apr 2001.

[68] J. Harri, F. Filali, and C. Bonnet, “Mobility models for vehicular ad hoc networks: a survey and taxonomy,” *IEEE Communications Surveys and Tutorials*, vol. 11, no. 4, pp. 19–41, Dec 2009.

[69] Christian Bettstetter, Hannes Hartenstein, and Xavier Prez-Costa, “Stochastic properties of Random Waypoint mobility model,” *Wireless Networks*, vol. 10, no. 5, pp. 555–567, Sep 2004.

[70] Kuo-Hsing Chiang and Nirmala Shenoy, “A 2-d random-walk mobility model for location-management studies in wireless networks,” *IEEE Transactions on Vehicular Technology*, vol. 53, no. 2, Mar 2004.

[71] Tariq Ali and Mohammad Saquib, “Performance evaluation of wlan/cellular media access for mobile voice users under random mobility models,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 10, pp. 3241–3255, Oct 2011.

[72] Enrica Zola and Francisco Barcelo-Arroyo, "Probability of handoff for users moving with the random waypoint mobility model," in *IEEE Conference on Local Computer Networks*, Bonn, Germany, 2011.

[73] Plamen I. Bratanov and Ernst Bonek, "Mobility model of vehicle-borne terminals in urban cellular systems," *IEEE Transactions on Vehicular Technology*, vol. 52, no. 4, pp. 947–952, Jul 2003.

[74] G. Hosein Mohimani Farid Ashtiani Adel Javanmard and Maziyar Hamdi, "Mobility modeling, spatial traffic distribution, and probability of connectivity for sparse and dense vehicular ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1998–2007, May 2009.

[75] Amir Reza Momen and Paeiz Azmi, "A stochastic vehicle mobility model with environmental condition adaptation capability," *Wireless Communications and Mobile Computing*, vol. 9, no. 8, pp. 1070–1080, Aug 2009.

[76] Andrea Clementi, Angelo Monti, and Riccardo Silvestri, "Modelling mobility: A discrete revolution," *Ad Hoc Networks*, vol. 9, no. 6, pp. 998–1014, Aug 2011.

[77] Kahina Ait Ali, Mustapha Lalam, Laurent Moalic, and Oumaya Baala, "V-mbmm: Vehicular mask-based mobility model," in *9th International Conference on Networks*, Menuires, France, Apr 11-16, 2010.

[78] J. Chung and D. Go, "Stochastic vector mobility model for mobile and vehicular ad hoc network simulation," *IEEE Transactions on Mobile Computing (published online)*, Aug 2011.

[79] Francisco J. Martinez, Juan-Carlos Cano, Carlos T. Calafate, and Pietro Manzoni, "CityMob: A mobility model pattern generator for VANETs," in *IEEE ICC Workshops*, Beijing, China, May 2008.

[80] Amit Kumar Saha and David B. Johnson, "Modeling mobility for Vehicular Ad hoc Networks," in *ACM International Workshop on Vehicular Ad Hoc Networks (VANET)*, Philadelphia, PA, Oct 2004.

[81] Amit Jardosh, Elizabeth M. Belding-Royer, Kevin C. Almeroth, and Subhash Suri, "Towards realistic mobility models for mobile ad hoc networks," in *International Conference on Mobile Computing and Networking (MOBICOM)*, San Diego, CA, Sep 14-19, 2003.

[82] J. Harri, F. Filali, C. Bonnet, and Marco Fiore, "VanetMobiSim: Generating realistic mobility patterns for VANETs," in *ACM International Workshop on Vehicular Ad Hoc Networks (VANET)*, Los Angeles, CA, Sep 29, 2006.

[83] Jonghyun Kim, Vinay Sridhara, and S. Bohacek, "Realistic mobility simulation of urban mesh networks," *Ad Hoc Networks*, vol. 7, no. 2, pp. 411–430, Mar 2009.

[84] Hongyu Huang, Yanmin Zhu, Xu Li, Minglu Li, and Min-You Wu, "Meta: A mobility model of metropolitan taxis extracted from gps traces," in *IEEE Wireless Communications and Networking Conference*, Sydney, Australia, Apr 18-21, 2010.

[85] Md. Shohrab Hossain and M. Atiquzzaman, "Stochastic properties and application of city section mobility model," in *IEEE Global Communications Conference (GLOBECOM)*, Honolulu, HI, Nov 30-Dec 4, 2009.

[86] "Google Maps," http://maps.google.com.

[87] "City of New York: Department of Transportation," http://www.nyc.gov/html/dot/html/faqs/faqs_signals.shtml.

[88] S. Gundavelli and K. Leung et al., "Proxy Mobile IPv6," IETF RFC 5213, Aug 2008.

Appendix A

Author's List of Publications

A.1 Peer Reviewed Publications

[1] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic, "Cost and Efficiency Analysis of Hierarchical SIGMA," Accepted for publication in *IEEE GLOBECOM*, Anaheim, CA, Dec 3-7, 2012.

[2] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic, "Performance Comparison between Multihomed Network Mobility Protocols," Accepted for publication in *IEEE GLOBECOM*, Anaheim, CA, Dec 3-7, 2012.

[3] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic, "A Network-based Seamless Handover scheme for Multi-homed devices," Accepted for publication in *IEEE GLOBECOM workshops: MobiWorld*, Anaheim, CA, Dec 3-7, 2012.

[4] Md. Shohrab Hossain and M. Atiquzzaman, "Analysis of Proxy Mobile IPv6: A Network-based Mobility Solution" Accepted for publication in *IEEE International Conference on Computer and Information Technology (ICCIT)*, Chittagong, Bangladesh, Dec 23-25, 2012.

[5] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic, "Performance Evaluation of Multihomed NEMO," *IEEE International Conference on Communications (ICC)*, Ottawa, Canada, Jun 10-15, 2012.

[6] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic, "On the Efficiency of IPv6-based Network Mobility," *IEEE International Conference on Communications (ICC)*, Ottawa, Canada, Jun 10-15, 2012.

[7] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic, "Scalability Analysis of a Multihomed Network Mobility Protocol," *IEEE GLOBECOM workshop on Mobile Computing and Emerging Communication Networks*, Houston, TX, Dec 5-9, 2011.

[8] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic, "Survivability Evaluation of NEMO with Multiple Mobile Routers," *IEEE GLOBECOM workshop on Mobile Computing and Emerging Communication Networks*, Houston, TX, Dec 5-9, 2011.

[9] A. Z. M. Shahriar, Md. Shohrab Hossain and M. Atiquzzaman, "A cost analysis framework for NEMO prefix delegation-based schemes," *IEEE Transactions on Mobile Computing*, Vol. 11, No. 2, pp. 1192-1206, July 2012.

[10] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic, "Cost and Efficiency Analysis of NEMO Protocol Entities," *Journal of Networks*, Vol. 7, No. 3, pp. 427-440, Mar 2012.

[11] M. Atiquzzaman and Md. Shohrab Hossain, "Security Issues in Space Networks," *NASA Earth Science Technology Forum*, Pasadena, CA, June 21-23, 2011.

[12] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic "Cost Analysis of Mobility Management Entities of SINEMO," *IEEE International Conference on Communications (ICC)*, Kyoto, Japan, 5-9 Jun, 2011.

[13] Md. Shohrab Hossain and M. Atiquzzaman, "Asymptotic Scalability Analysis of Mobility Protocols based on Signaling Overhead," *Journal of Telecommunication Systems*, published online: 30 June, 2011.

[14] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic "Security Vulnerabilities and Protection Mechanisms of Mobility Management Protocols," *IEEE Aerospace conference*, Big Sky, Montana, USA, Mar 2011.

[15] Md. Shohrab Hossain and M. Atiquzzaman, "Asymptotic Scalability Analysis of Mobility Protocols based on Signaling Overhead," *International Journal of Communication Networks and Distributed Systems (IJCNDS)*, Vol. 7, Nos. 1-2, pp. 119-134, 2011.

[16] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic "Cost analysis of NEMO Protocol Entities," *International Conference on Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh, Dec 23-25, 2010.

[17] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic "Performance Analysis of NEMO BSP using City Section Mobility Model," *International Conference on Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh, Dec 23-25, 2010.

[18] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic "Cost Analysis of Mobility Entities of Hierarchical Mobile IPv6," *IEEE Military Communications Confence (MILCOM)*, San Jose, CA, Oct 31- Nov 3, 2010.

[19] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic "Scheduling and Queue management for Multi-class Traffic in Access Router of Mobility Protocol,"

12th IEEE International Conference on High Performance Computing and Communications (HPCC), Melbourne, Australia, Sept 1-3, 2010.

[20] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic "Survivability and Scalability of Space Networks," *NASA Earth Science Technology Forum*, Arlington, VA, June 22-24, 2010.

[21] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic "Scalability Analysis of PD-based schemes of NEMO," *IEEE High Performance Switching and Routing (HPSR)*, Richardson, TX, June 13-16, 2009.

[22] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic "Cost Analysis of Mobility Management Entities for SIGMA," *IEEE High Performance Switching and Routing (HPSR)*, Richardson, TX, June 13-16, 2009.

[23] Md. Shohrab Hossain and M. Atiquzzaman, "Stochastic Properties and Application of City Section Mobility Model," *IEEE Global Communications Conference (GLOBECOM)*, Honolulu, HI, Nov 30-Dec 4, 2009.

[24] Md. Shohrab Hossain and M. Atiquzzaman, "Signaling Cost Analysis of Mobility Protocols using City Section Mobility Model," *2nd Internation Conference on Computer Science and Application*, Jeju Island, Korea, Dec 10-12, 2009.

A.2 Papers under Review

[1] Md. Shohrab Hossain, M. Atiquzzaman, and W. Ivancic, "Vehicular Mobility Model in City Streets: Properties and Validation," Submitted to *IEEE Transactions on Vehicular Technology.*

[2] Md. Shohrab Hossain, Husnu Narman and M. Atiquzzaman, "A Novel Scheduling and Queue Management Scheme for Multi-band Mobile Routers" Submitted to *IEEE International Conference on Communications (ICC)*, 2013.

Appendix B

Acronyms

AH Authentication Header

AR Access Router

AZS Anchor Zone Server

BA Binding Acknowledgement

BSP Basic Support Protocol

BU Binding Update

CGA Cryptographically Generated Address

CN Correspondent Node

CoA Care of Address

CoT Care-of Test

CoTI Care-of Test Init

DDoS Distributed Denial of Service

DHCP Dynamic Host Configuration Protocol

DNS Domain Name System

DoS Denial of Service

ESP Encapsulating Security Payload

FTP File Transfer Protocol

HA Home Agent

HiSIGMA Hierarchical SIGMA

HIP Host Identification Protocol

HMIPv6 Hierarchical Mobile IP vesrion 6

HoA Home Address

HoT Home Test

HoTI Home Test Init

HZS Home Zone Server

ICMP Internet Control Message Protocol

IETF Internet Engineering Task Force

IKE Internet Key Exchange

IP Internet Protocol

IPsec IP security

LCoA Local Care of Address

LFN Local Fixed Node

LLM Local Location Manager

LM Location Manager

LMA Local Mobility Anchor

MAP Mobility Anchor Point

MH Mobile Host

MIP Mobile IP

MIPv6 Mobile IP vesrion 6

MITM Man In The Middle

MNN Mobile Network Node

MNP Mobile Network Prefix

MR Mobile Router

MSF Mobility Scalability Factor

NEMO NEtwork MObility

NRT Non-Real Time

PDA Personal Digital Assistant

RA Router Advertisement

RBU Refreshing Binding Update

RCoA Regional Care of Address

RR Return Routability

RT Real Time

RTT Round Trip Time

SA Security Association

SCTP Stream Control Transport Protocol

SIGMA Seamless IP-diversity based Generalized Mobility Architecture

SINEMO Seamless IP-diversity based Network Mobility

SPI Security Parameters Index

TCP Transmission Control Protocol

TNRL Telecommunications and Networks Research Lab

UDP User Datagram Protocol

VMN Visiting Mobile Node