A Two-Layered Mobility Support Architecture:

Fast Mobile IPv6 and Session Initiation Protocol

Prepared by:
Deeya Shakti Nursimloo

Supervised by:
Professor H.A. Chan

This thesis is submitted in partial fulfillment of the academic requirements
for the degree of
Master of Science in Electrical Engineering
in the Faculty of Engineering and The Built Environment
University of Cape Town

30 November 2006

A Two-Layered Mobility Support Architecture:

Fast Mobile IPv6 and Session Initiation Protocol

Prepared by:
Deeya Shakti Nursimloo

Supervised by:
Professor H.A. Chan

This thesis is submitted in partial fulfillment of the academic requirements
for the degree of
Master of Science in Electrical Engineering
in the Faculty of Engineering and The Built Environment
University of Cape Town

30 November 2006

Declaration

I declare that this thesis is my own work. Where collaboration with other people has taken place, or material generated by other researchers is included, the parties and/or materials are indicated in the acknowledgements or are explicitly stated with references as appropriate.

This work is being submitted for the Master of Science in Electrical Engineering at the University of Cape Town. It has not been submitted to any other university for any other degree or examination.

Signed by candidate

Deeya Shakti Nursimloo

Date

Dedication

To my late and beloved father, Appanah Nursimloo

I dedicate this thesis to my late father who had been my inspiration all through.

Acknowledgements

First and foremost, I thank God for constantly supporting me in all circumstances of my life.

I would like to express my sincere gratitude to the following individuals for their assistance during the course of this project:

Professor H.A. Chan, for his unwavering support, guidance and encouragement throughout the project.

My dearest mother, for her love and a great moral support in very aspect of my life.

Rikesh, for being an amazing friend and for always encouraging me in the challenging days.

My friends in the Communications Research Group (CRG), for the useful and interesting discussions on NGN networks, life, and trees. Thank You

Abstract

Real-time communications are likely to play a key role in the convergence of the Internet with various access network technologies. The Next Generation Network aims at providing these real-time services to nomadic users with roaming devices within the Internet infrastructure. Therefore, the objective of the study is to minimize the handover latency and packet loss during any ongoing real-time communication by inter/intra domain mobility.

Mobile IP (MIP) is a well-known network layer protocol that can support transparent macromobility, and variants of the protocol such as Fast Handover for Mobile IP has been proposed to provide an optimized handover scheme for micromobility within IP networks. Fast Mobile IP relies extensively on layer 2 information to anticipate a network handover and to redirect the traffic to the new location the mobile device is about to move to. However, the protocol is burdened with triangular routing that introduces unacceptable delays for real-time communications.

The Session Initiation Protocol (SIP) is an application layer protocol that provides different types of mobility from service to terminal mobility. Terminal mobility in SIP supports real-time communications over User Datagram Protocol (UDP). The main drawbacks in the SIP mobility framework are the call disruption delays incurred when a mobile device is in an overlapped region and the absence of mobility management on the Transport Control Protocol (TCP) connections.

This project proposes an integrated mobility scheme that combines procedures of Fast Handover for Mobile IP and SIP mobility for real-time communications over UDP. An analysis of the existing protocols i.e. network layer Fast Mobile Handover for IP and SIP mobility is presented. The integrated proposed scheme efficiently aims at reducing the handover latency and packet loss for ongoing real-time conversations, i.e. Voice over IP (VoIP) call and the streaming of a video application. Simulation results presented are based on the network simulator ns-2. The simulation results compare and analyze the performance of the proposed integrated scheme to the existing protocols. Thus, the proposed architecture presents a powerful handover mobility support for Next Generation IP-based wireless systems. Recommendations for future work have been presented to further evaluate and optimize the use of the proposed integrated handover scheme.

Synopsis

Mobile communication has grown from a world of cables to wireless communications. The next-generation wireless systems are envisioned to have an IP-based infrastructure platform to support the heterogeneity of the access technologies. Currently, various wireless technologies and networks provide different services to mobile users based on their requirements. The integration of these technologies on a common IP-based platform will empower the mobile users, allowing them to be connected to any system with the best access networks that suit their needs. Mobility management is required to make the roaming seamless among the heterogeneous networks and to minimize service disruptions of real-time applications during the handover. In this study, an integrated Fast Mobile IP and SIP handover management architecture has been presented. The building blocks of the project exploited the complementary capabilities of both protocols to reduce handover latency and packet loss.

In a heterogeneous environment, mobility enabled protocols are considered to achieve global roaming among the various access technologies. Mobility management consists of handover management and location management that will be needed to interwork the mobility protocols to support seamless roaming.

Network-Layer protocol, Mobile IP (MIP) and application-layer Session Initiation Protocol (SIP) each have dominating roles in IP-based mobility management. Fast Handovers for Mobile IPv6 protocol is one of the proposed enhancements of Mobile IP within the IETF mobileip working group. Its performance is based on the capability of supporting two types of handover: reactive and pro-active handover mechanisms. These mechanisms aim to reduce service degradation that a mobile device could suffer due to a change in its point of attachment. SIP is an essential protocol that can allow the provisioning of services in IP-based networks. Therefore, there is a need to seamlessly interwork Fast Mobile IP and SIP for the mobility of real-time services.

This study investigated on an integrated approach of both protocols based on the context of the applications utilized. Furthermore, to reduce system redundancies and signaling loads, several functionalities of Fast Mobile IP and SIP have been integrated to optimize the architecture. The proposed framework would be able to support various types of mobility scenarios by making use of Fast Mobile and SIP in a jointly optimized and effective way.

An evaluation framework was designed in network simulator ns-2 to support the proposed framework. The simulations have shown that the proposed mobility architecture can effectively reduce packet loss and handover latency by combining the merits of Fast Mobile IP and SIP. Thus, the proposed architecture can offer efficient and reliable handover mobility support for the IP-based next generation wireless systems.

Table of Contents

List of Figures

List of Tables

Glossary

This section defines some of the commonly used terms and abbreviations that appear throughout this document.

Access Point (AP) An entity that bridges information between the wireless medium and the distribution medium on behalf of its associated stations. Access points are only used in wireless LANs infrastructure.

Access Router (AR) A router that lies on the periphery of a network and provides mobility services to visiting mobile nodes.

Binding Update (BU) A message that supplies the new care-of address for a mobile node. The binding update contains the mobile node's home address, new care-of address and registration lifetime.

Care-of Address (CoA) A temporary IP address allocated to a mobile node when it is visiting a foreign network.

Correspondent Node (CN) A node that sends or receives a packet to a mobile node, it may be another mobile node or a conventional fixed node.

Duplicate Address Detection (DAD) The uniqueness of the address is verified though Duplicate Address Detection mechanism.

Dynamic Host Configuration Protocol (DHCP) It is a set of rules used by a communication device (such as a computer, router or networking adapter) to allow the device to request and obtain an Internet address from a server which has a list of addresses available for assignment.

Foreign Agent (FA) A mobility agent on a foreign network that can assist a mobile node in receiving datagrams delivered to the care-of address.

Fast Handover for Mobile IP (FMIP) The Fast Handover protocol is one of the enhancements of Mobile IP that allows a mobile node to detect its next point of attachment before the handover mechanism.

Home Agent (HA) An IPv4 or IPv6 router that resides on a mobile node's home network. The home agent forwards a mobile node's traffic to its current point of attachment while it is away from its home network.

Internet Protocol (IP) A network layer technology that uses packet-switching techniques to transmit data.

Mobile IP (MIP) A network layer protocol that allows a mobile node to migrate through different IP networks. Two versions of Mobile IP currently exist: Mobile IPv4 (MIPv4) and Mobile IPv6 (MIPv6).

Mobile Node (MN) A network node that is able to communicate to the Internet while moving through different networks.

Quality of Service (QoS) A feature that prioritises and guarantees bandwidth for selected applications to achieve an optimal service performance.

Real-time Applications Streaming video and VoIP are considered as real-time multimedia transmission. These applications are highly intolerant to delay or packet loss.

Real-Time Transport Protocol (RTP) The protocol is designed to provide end-to-end network transport functions for applications transmitting real-time data such as audio and video.

Session Initiation Protocol (SIP) An application layer protocol that initiates user session for multimedia applications such as video and voice.

Transmission Control Protocol (TCP) The transport-layer protocol is used in conjunction with Internet Protocol (IP) to provide a reliable, connection-orientated service to the invoking application.

User Agent (UA) A user agent is a client application used with a particular network protocol.

User Datagram Protocol (UDP) The transport-layer protocol provides an unreliable, connectionless service to the invoking application.

Voice over IP (VoIP) Also known as “IP telephony”, it is a two-way transmission of voice information over a packet switched TCP/IP network.

Wireless Local Area Network (WLAN) 802.11b A wireless local area networking family of standards developed by the IEEE that emulates conventional Ethernet links. Different implementation of the standards exist namely, 802.11 a/b/g, for a range of data rates.

Chapter 1

1 Introduction

The ever-increasing growth of user demand for mobile communication services and the emergence of new broadband technologies in the wireless and mobile communication domain introduce the challenges that the Next Generation Networks needs to overcome. The success and growth of the Internet has influenced the evolution of mobile communication systems. Access to integrated data, voice and multimedia services through IP (Internet Protocol) connectivity is now provided to mobile devices such as Laptops, PDA and most recently mobile phones.

Today's communication technologies are designed to be user-centric and now form an integral part of everyone's life. The integration of different access network technologies in next generation wireless systems leads to the concept of heterogeneity in user terminals, media sessions and access networks. A heterogeneous pocket-sized phone will provide wireless Internet access at low cost, high data rates and ubiquity i.e. broadband network services will be available to the user anywhere, anytime and through different access technologies. IP Multimedia services are expected to bring diversification to the available services by supplementing the traditional voice-orientated services with real- time services centered on data, video and IP telephony.

Seamless IP handover for next generation IP networks is one of the key challenges in mobility management when switching from, e.g., a Universal Mobile Telecommunication System (UMTS) to wireless local area networks (WLAN) radio communication or between two WLAN access points of different IP addresses. Quality of Service of real-time services provided by the networks will greatly depend on the ability to minimize the impact of handover. The mobile host should be capable of roaming through different access technologies while still maintaining any of its ongoing calls. These developments in technologies require a converged Internet with efficient mobility protocol support. Currently, there are two approaches to support mobility of services in the IP core network: Mobile IP supports mobility across the network layer and Session Initiation Protocol (SIP) through the application layer. Both of these protocols suffer from different types of drawbacks that impact on the media flow during the handover mechanism.

1.1 IP mobility in Next Generation Wireless Networks

The focus is placed on comparing and solving a terminal mobility framework of real-time services during intra-domain handovers that incorporate Network layer Mobile IP proposals and application layer SIP.

The following requirements are identified to be discussed in the implementation of the proposed mobility framework in the project:

- Mobility Management

Mobility management is emerging into several mobility types in terms of terminal, session, personal, service, ad hoc and mode mobility. Terminal mobility includes: handover management and location management. Terminal Handover management allows a mobile device to keep an ongoing call as the device changes points of attachment (from radio channel, base station, networks, etc). Location Management enables the incoming calls delivery for idle mobile hosts. The management of terminal mobility at different levels of the system architecture require different solutions as illustrated using the Internet protocol stack. The network model is used to describe a hierarchical networking protocol stack identifying the information required from each layer. Each layer offers types of services required to the higher layer, while the complexities and intricacies of the services are not discussed.

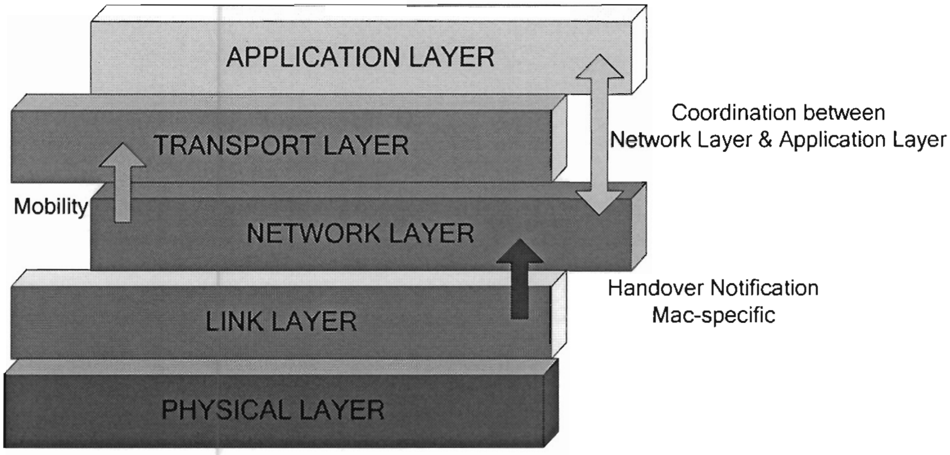

Figure 1.1: TCP/IP model

- Network Layer – Mobile IP

From the protocol stack perspective, the network layer is the lowest possible layer on which the convergence of heterogeneous wireless systems for an all-IP environment can be developed. Network-layer mobility management protocols together with Link-layer schemes are used to support terminal mobility. Mobile IP has been designed to support host mobility on the internet from the network layer. Currently developed with the Internet Engineering Task Force (IETF) Mobile IP working group, Mobile IP aims to support network layer mobility management for inter-domain mobility. A mobile host can migrate through different IP networks in a Mobile IP environment, ensuring that all mobility functionality is transparent to the upper layers. The aim of the transparency mechanism is to allow to view all upper layer sessions (e.g. TCP) in spite of a network layer change. One of the main challenges in Mobile IP is its poor performance support during handover. A Mobile IP handover can lead to performance degradation in terms of handover latencies and packet loss. These may have adverse effects on real-time applications as these applications would perform unacceptably with high amount of latency and packet loss.

- Application Layer – Session Initiation Protocol

Session Initiation Protocol (SIP) is defined by the IETF to establish IP communication for services as Voice over IP (VoIP). Currently, SIP is being deployed in IP Multimedia Subsystems (IMS) of the 3G Partnership Project (3GPP) for call signaling. SIP can provide a framework for Application-Layer mobility management and can be extended to support terminal mobility by augmented signaling. One of the main challenges in SIP mobility mechanism is call disruption if the SIP connection is not well established as the mobile host moves from one point of attachment to another. The call disruption is primarily due to the IP address renewal configuration, bringing a perceivable period of silence in the ongoing call during handover.

In this thesis, an analysis of the mobility protocols that have been developed to support intra and inter domain mobility is presented. Micromobility protocols such as Fast Mobile IPv6 are discussed assuming that Mobile IP is used as the global mobility protocol. The focus is on a proposed integrated mobility management approach to compare the network layer and application layer protocols, and to achieve significant reduction in the handover latency and packet loss to support real-time applications.

1.2 Related work in Mobile IP and SIP

Several schemes have been proposed to handle IP mobility for VoIP applications using the two layer approach: network layer and application layer. As outlined in [1] and [2], the main aim is to compare Mobile IP and SIP-based protocols and to provide mobility management framework that seeks to reduce disruptions during handoff. [3] presents an analysis of Mobile IP micromobility protocols and SIP and proposes a hybrid model scheme that incorporates Cellular IP and SIP. A mobility framework known as Hierarchical Mobile SIP [4] combines Hierarchical Mobile IP and SIP for intra-domain mobility. From an application layer perspective, several description of SIP mobility architectures can be found in [5], [6] and [7] for SIP to provide terminal mobility support to real-time applications. The main motivation in all these schemes is to provide an end-to-end mobility management framework to minimize the adverse effects of the handover performance.

1.3 Thesis Objectives

This study investigates the performance of Fast Mobile IPv6 protocol and SIP protocol to support terminal mobility. Seamless handover is required in an "All-IP" mobility framework and the study specifically focuses on an integrated mobility framework that incorporates the network layer protocol and the application layer protocol. In order to do so, the following issues need to be investigated:

- The requirements of real-time applications are considered to minimize service disruptions in terms of handover latency and packet loss.
- Handoff execution procedures are described based on the link layer trigger.
- A theoretical study to compare the performance of Fast Mobile IPv6 and SIP is presented to investigate their benefits and drawbacks.
- A proposed architectural design integrates both protocols to work in a way to complement each protocol features based on the kind of application.
- An evaluation framework is used through a network mobility software for simulations to quantitatively analyze the proposed design.
- A theoretical investigation is carried out on a mobility decision table to determine which type of applications the proposed mobility framework is suitable for.
- The evaluation framework is designed to analyze the existing protocols framework to the proposed design.
- Lastly, two types of real-time applications i.e. VoIP and video streaming will serve as examples for the network requirements.
- These requirements will be a benchmark to evaluate the proposed integrated mobility framework.

1.4 Scope and Limitations

This study focuses specifically on mobility management protocols Fast Mobile IPv6 and SIP that operate on different network layers. Research will be restricted to network and application layer mechanisms.

Fast Mobile IPv6 is one of the extensions of Mobile IPv6 to provide fast handoff support in order to eliminate inherent Mobile IP delays. Numerous researches are carried out on the scalability of the protocol. Although the performance of the protocol is still being studied, the protocol provides a large margin of independence to be deployed on different visited networks to support end-to-end mobility. However, the main area of research in this study is based on the anticipated change in network in the Fast Mobile IPv6 proactive handoff management.

SIP was originally proposed by the IETF as a multimedia session initiation protocol and is now being adopted in 3GPP IMS platform. SIP supports different types of mobility and additionally can also support terminal mobility. An introduction of mobility awareness is provided at the application layer. The focus on SIP in this study will be placed on terminal mobility to improve the performance of real-time applications during handover.

The real-time applications, for example, VoIP are simulated in the network simulator and thus only audio codec is characterized. The simulation scenarios involve a single mobile node that participated in a single point-to-point VoIP call setup. The scope of the research is basically based at the micromobility level i.e. handover between Wireless LANs are considered and no changes are required to the IP stack. Vertical handovers, for example between WLANs and UMTS, are not considered.

Quality of Service, Security and Authorization, Authentication and Accounting (AAA) issues remain on the periphery of this study.

1.5 Thesis Outline

The remainder of this document is organized as follows:

Chapter 2 and Chapter 3 provide a complete background of the mobility in terms of real-time applications, related work on FMIPv6 and SIP mobility and handover management. The network requirements needed to effectively support VoIP traffic are presented. An outline of Fast Mobile IPv6 and SIP is presented to describe the important differences in each protocol.

This will provide a foundation for the proposed mobility architecture in the study. Handover procedures in both protocols are also considered.

Chapter 4 includes analysis information from the concept discussed in the previous chapter for the proposed mobility framework. The proposed design integrates mobility support with the existing protocols to minimize handover latency and packet loss within overlapped networks. This section focuses specifically for advanced handover management for the different types of application that can be supported within the mobility framework.

In Chapter 5, the implementation network mobility software is presented. The software is modified to include Fast Mobile IP and SIP implementation. Additional modifications are further implemented to support the proposed mobility framework. Two types of real-time traffic are characterized within the network simulator to evaluate the requirements of the traffic in terms of handover latency and packet loss.

Chapter 6 presents the results obtained from the simulation scenarios performed. The parameters metrics that have been investigated are handover latency and packet loss. The proposed design is compared with existing protocols using the parameter metrics. Lastly, an overall analysis of the proposed design to support real-time communication is outlined.

Chapter 7 presents a set of conclusions based on these evaluations. Some important aspects of the protocols used in the study are also included.

Chapter 8 lists some recommendations that were observed during the course of the project. Some of these recommendations lead to related areas of research to this study and future work that can be further developed in the mobility framework.

Chapter 2

2 VoIP for Next Generation Network

2.1 Real-time applications

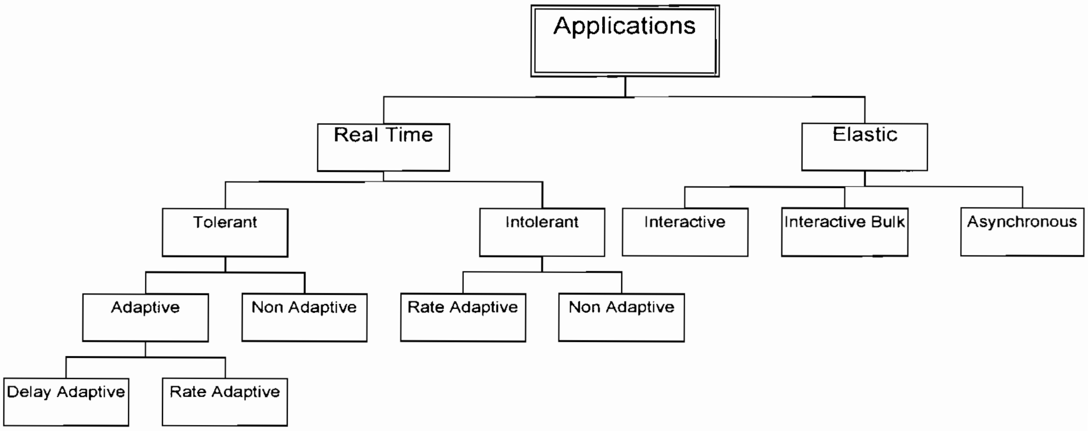

Figure 2.1: Application Classification

Applications provided through the Internet are classified according to their data requirements as outlined by Peterson and Davies [8]. The service model in Figure 2.1 explains how applications can also be classified based on their Quality of Service requirements.

The applications are sub-divided to Elastic and Real-time applications. Elastic applications mostly required best-effort services requirement, which is typically the service level currently available via the Internet. Real-time applications are more delay sensitive and predictable handover mechanisms may be used to obtain efficient handover and minimize latencies as further discussed in the project. Voice over IP (VoIP) is a real-time application which will analyzed in the project.

Human-to-human interactive voice communication has been the dominant mode of human communication, starting from the mouth of the talker until the acoustic signals reaches the ear of the listener. This communication system has been traditionally supported over circuitswitched network such as Public Switched Telephone Networks (PSTN). This section presents an overview of the use of Voice over IP Application (VoIP) systems which are currently being increasingly deployed in packet-switched networks.

2.2 Overview of VoIP

The basic components of the VoIP systems are shown in Figure 2.2 and are subdivided into four components: signaling, encoding, transportation and gateway control.

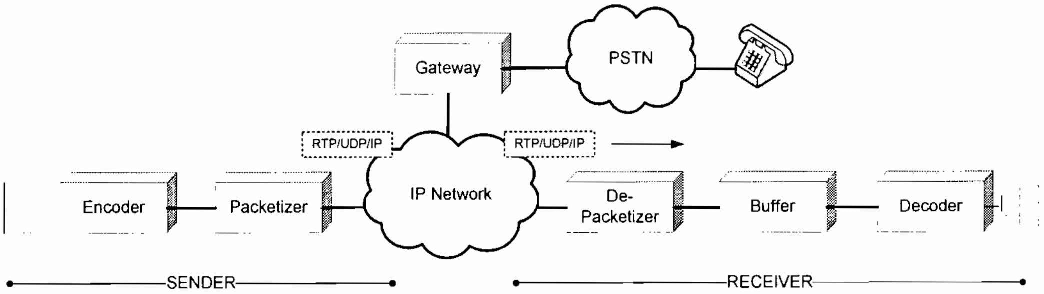

Figure 2.2: VoIP Transmission Path

Currently, there are the two call control standards for IP telephony: the Internet Engineering Task Force (IETF) standard Session Initiation Protocol (SIP) [9] and the International Telecommunication Union (ITU) H.323 [10]. First, the digitized human speech is encoded through an audio codec and the input analogue voice signal is mapped into an appropriate digital format called a frame. The digitized voice frame is concatenated into one packet and transmitted over the IP network. Real-Time Transport Protocol (RTP) [11], User Datagram Protocol (UDP) [12] and Internet Protocol (IP) packets headers are added to the speech segments. The IP network is responsible for transmitting packets from the sender to the receiver. When the incoming packets are received at the receiver side, these packets are converted back into an analogue acoustic output signal. Gateways are also attached to the system to interoperate with different networks such as PSTN.

Quality of service of VoIP application must be catered for when transmitting voice over a data network. Due to the interactive nature of voice application, the main criteria to be catered for VoIP application are as follows:

- Delay
- Jitter
- Packet Loss

Delay

End-to-end delay in VoIP application sums the delays experienced at every stage in the communication path starting from the encoding processes to network transmission and buffering. Table 2.1 below outlines how different delays values are perceived on voice quality.

Table 2.1: Delays in VoIP application

One way delay	Effect
< 100-150ms	Delay, but not detectable
150-250 ms	Slight noticeable delay, Acceptable quality
Over 250-300 ms	Unacceptable delay

From the table, the desired end-to-end delay for a VoIP application must be kept below 150ms. Within the range of 150ms and 250ms, delays do affect the voice quality but at this stage, the output of the voice quality is still acceptable to the users. Over the range of 300ms, the effects of delays make it impossible for the proper voice quality to be generated.

Jitter

Jitter are encountered as the short-term fluctuations within the packet rate of arrival. The number of jitters may affect the delay variation of the VoIP application. Ideally, the jitter variation should be kept below the range of 40 ms but up 70ms the jitter variations may still be acceptable. A de-jitter buffer may be used at the receiver end-system to alleviate some amount of jitters to the overall end-to-end delay.

Packet Loss

VoIP requires low packet loss for the transmitted interactive voice application to be of acceptable quality. A low packet loss ratio is important for the transport of compressed speech. A number of factors can affect the pattern of packet loss, for example, when voice packets may be dropped or lost in the transmission network and cause degradation on the output voice quality.

Furthermore, as outlined in [13] other sources of information loss and packet loss occurs in groups, these will have serious effects on degrading the speech quality.

2.3 VoIP Attributes

VoIP are analogue PCM voice signals, which are encoded and compressed into low bit, rate packet streams by codecs. Different types of codecs are commonly used to generate constant bit-rate audio frames as shown in Table 2.2:

Table 2.2: Commonly Used Codecs

Codec	GSM 6.10	G.711	G.723.1	G.726-32	G.729
Bit Rate (Kbps)	13.2	64	5.3/6.3	32	8
Framing Interval(ms)	20	20	30	20	10
Payload (bytes)	33	160	20/24	80	10
Packets/sec	50	50	33	50	50*

For all codecs except G.729, Packets/sec = 1 / (Framing interval). For G.729, two frames are combined into one packet so that Packets/sec = 1/ (2 Framing interval)

A typical VoIP frame at the IP layer consists of 40-byte IP/UDP/RTP headers followed by a relatively small payload ranging from 10 to 30 depending on the codec used as shown above. The time frame between to adjacent frames approximates to 20 ms corresponding to a rate of 50 packets per second along the VoIP stream. So the efficiency at the IP for VoIP application is already less than 50% [15].

Chapter 3

3 Mobility

3.1 Network Layer Mobility Approach

The Internet Protocol (IP) is the protocol at the network layer that connects networks to networks to route packets from the Internet according to their IP addresses. Mobile IP is a proposed standard protocol that provides a relatively scalable mobility mechanism that allows a mobile node to continue and maintain its IP connections as it migrates to different IP subnets [15]. Developed within the Internet Engineering Task Force (IETF) Mobile IP working group, Mobile IP protocol is designed to allow a moving device, for example a mobile node, to have two IP addresses: a fixed home address and a care-of address that changes at each new point of attachment.

3.1.1 Mobile IP Protocol

Mobile IP is designed into the following branches to support IP mobility: Mobile IPv4, which operates within the IPv4 framework, and Mobile IPv6 which is within the IPv6 framework. The following sections will investigate and outline the main differences between Mobile IPv4 and Mobile IPv6.

3.1.1.1 Overview of Mobile IPv4

Mobile IP supports the mobility of IP nodes in a mobile environment by allocating IP addresses at the network layer. The mobile environment is composed of two types of network: a home network and a foreign network. The scheme uses at least two types of IP addresses: a fixed address that represents the home address of the mobile node and a care-of address (CoA) that changes as the mobile node attaches itself to an IP subnet. The Mobile IP protocol consists of the following entities:

- Mobile Node (MN): A Mobile Node is an IP host that changes its point of attachment from one IP network to another.
- Home Agent (HA): A Home Agent is a router on the mobile node's home network. It stores the current IP address of the mobile node when the mobile node is away from its home network. It intercepts and tunnels any packets destined to the mobile node current location.
- Foreign Agent (FA): A Foreign Agent is a router on the foreign or visited network. It receives the tunneled packets from the HA and forwards them to the mobile node which is in the foreign network. Home and foreign agents are known as mobility agents.
- Correspondent Node (CN): A CN can be either a fixed or mobile node. The CN participates in communication with the mobile node.

To support IP mobility, more than one IP address is allocated to a mobile node. The primary IP address assigned to a MN is known as the home address which represents the identity of the node as its home network. When the MN moves to a foreign network, a transient IP address known as the care-of address (CoA) is generated.

Mobile IP consists of three major operations: Agent discovery, Registration and Tunnelling [16]. Agent discovery is the process of advertising the ability of the mobility agents for services required on each link. A mobility agent periodically uses advertisement to allow the mobile node to find its agents. A MN will use agent discovery mechanisms to detect the new mobility agents when the mobile node migrates to a new IP network. Registration allows a mobile node to register to its home agent and also provides registration service for the mobility agents. Tunnelling is to allow the mobility agents to forward packets destined for the mobile node to be routed properly.

The home agent and the foreign agent cooperate to deliver IP packets destined to the mobile node in the foreign network. Therefore, before the forwarding of IP packets, a mobile node must configure and register its CoA. There are two modes a mobile node can acquire a CoA: a foreign agent CoA or through external mechanisms such as Dynamic Host Configuration Protocol (DHCP) [17]. In the case of a foreign agent CoA, the foreign agent supplies its own IP address to be used as CoA the mobile node. The type of address generated from the external mechanism such as DHCP is known as a co-located CoA and is independent from the foreign agent's IP address.

In Mobile IPv4 [18], the Home Agent maintains a binding cache between a MN's home address and its currently assigned care-of address. When a mobile node has acquired a new CoA, it notifies its home address about its new address. The MN registers and the CoA binding is updated as the MN moves to a different IP network to ensure that the packets are forwarded to the correct foreign network. The binding has a lifetime mechanism associated to it so that if a MN does not renew its binding, it expires and is deleted.

When a correspondent node wants to send packets to a mobile node, it will first need to send the packets to the MN's home address. The IP packets are routed using standard IP routing mechanisms to the home network. In the home network, these packets will be intercepted and will be tunneled by the HA to the MN's current point of attachment using the CoA. Tunnelling is done through IP-in-IP encapsulation i.e. the original IP packet destined to the MN is encapsulated in a new IP packet with the MN's current CoA as a destination address as shown in Table 2.3 for IPv6. This is known as triangular routing i.e. packets need to be routed to the HA first and then to the mobile node when it is away from its home network.

Table 3.1: IPv6 in IPv6 Encapsulation

IPv6 Header (Outer)	IPv6 Header (Inner)	Transport Header	Payload
Source Address: Home Agent	Source Address: Correspondent Node	TCP/UDP	Data
Dest. Address: Care of Address	Dest. Address: Home Address		
← 40 bytes →	← 40 bytes →		

Once the tunnelled packet reaches the foreign network, in the case of the foreign agent CoA mode, the foreign agents receives, extracts and delivers the original packet to the MN. In the co-located CoA mode, the mobile node receives the tunnelled packets and then decapsulates them.

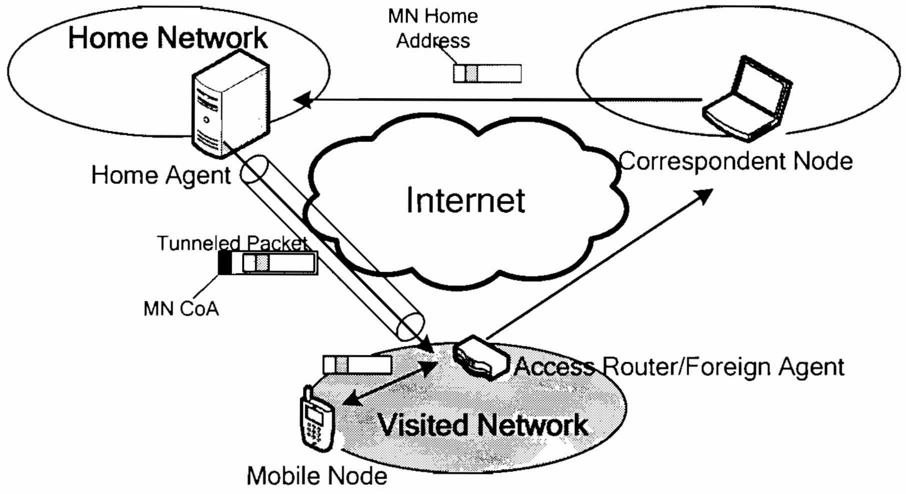

Figure 3.1: Mobile IP system overview

As described in Figure 3.1, mobility support in IPv4 is defined by a Home Agent and a Foreign Agent, which both assist the MN to attain mobility. Packets from the MN need to be tunnelled between the Foreign Agent and the Home Agent to pass the packets along to the Mobile node and a Correspondent node. Mobile IPv4 has some limitations for a wide-scale deployment in the NGN environment. One of the main limitations is that all traffic from the mobile node needs to be tunnelled through the HA thus increasing the overall traffic signaling load as well as the time latency induced for the mobile traffic. Mobile IPv4 is described in depth in [16].

3.1.1.2 Overview of Mobile IPv6

Mobile IPv6 [19] is an extension of Mobile IPv4, by replacing IPv4 implementations with IPv6 mechanisms to support mobility. MIPv6 works as a network layer protocol to provide node-to-node connectivity to the upper layers of the stack. The main relevant differences from IPv4 to IPv6 are outlined below as well as how Mobile IPv6 integrates within the new mechanisms to handle mobility and the routing of packets between the MN's involved.

IPv6 Addresses

IPv6 addresses offer larger address space in terms of 128 bits long whereas IPv4 addresses are only 32 bits long. Several types of IP address are defined in the IPv6 address space: part of the address space is allocated to the IPv4 addresses to assist interworking of IPv4 and IPv6 networks. The address types include loopback and global unicast. In IPv6, the 128-bit address consists of a 64-bit routing prefix, which is used for routing the packets to the required network and a 64-bit interface identifier, which specifies a specific node on the network. Therefore, the IPv6 address configuration in MIPv6 allows to identify the node, its location on the network, or even both cases.

IPv6 Address Autoconfiguration

There are two methods that the mobile node acquires CoA in Mobile IPv6 when it is away from the home network [20]: IPv6 Stateless configuration or stateful auto-configuration.

Firstly, IPv6 Stateless configuration allows an IPv6 host to automatically generate an IPv6 address without the need of external servers. In order to do so, the node configures its link-local address (for home network) based on the network's interface link layer address and its link-local prefix. Once the link-layer address is generated and assigned to the mobile node, it will then generate and autoconfigure a global address by including the network prefix.

On the other hand, stateful address configuration uses an external server to generate a global IPv6 address. The server maintains and updates a database of the available addresses in respect to the nodes to which the addresses are assigned. A very good example of the stateful address configuration mechanism is DHCPv6.

IPv6 Header Extensions

For Mobile IPv6, registration and tunnelling operate similarly to Mobile IPv4. The new extensions headers in IPv6 i.e. the Destination Options and Routing headers are related to IPv6 mobility. The Destination Options provides certain options within the packet that can only be processed once the packet has reached the destination host. The Routing header specifies the type router required to forward the packet to its destination.

In IPv6, like the home agent, the IPv6 Correspondent Nodes are also able maintain a local binding cache that stores and updates the mapping home addresses to care-of addresses of the recent mobile node entry. Therefore, this mechanism allows to bypass the home network because the packets are sent directly to the MN's CoA.

In Mobile IPv6, binding updates between the HA and the CN are used to inform HA and CNs about the changes in the point of attachment of the MN. The binding updates are sent as IPv6 Destination option headers in the IPv6 packet. The CN uses IPv6 Routing Header extensions to send packets to the mobile as it carries the MN's home address. In this case, IP-in-IP encapsulation is not used to tunnel the packets, as the packets destined for the MN are addressed using MN's CoA. Therefore, Mobile IPv6 supports route optimization i.e. enables a correspondent node to send packets directly to a mobile node, thus alleviating the triangular routing of Mobile IPv4 [19].

3.1.2 Neighbor Discovery mechanism

Neighbour Discovery is an IPv6 protocol that defines a set of IPv6 mechanisms that allows a node to acquire information about the other nodes present on the same link. These mechanisms determine whether the neighbouring node is still reachable through Neighbour Discovery [21].

In IPv6 network, the discovery of a care-of address is still required and is configured by using stateless Address Autoconfiguration and Neighbor Discovery as mentioned above. Thus the need for foreign agents are not required to support IPv6 mobility and their functionability can be replaced by IPv6 routers called access routers (ARs). Within the ARs, the IPv6 protocols are integrated to support visiting mobile nodes along the link. Through Neighbor Discovery mechanisms, link layer addresses and the reachability of neighbouring nodes can be confirmed.

Within an IPv6 environment, Neighbor Discovery is used by a mobile node to be able to discover and refresh the links to an existing one. The MN can detect the presence of an access router through periodic router advertisements. Unsolicited periodic router advertisements are sent by the routers at a minimum allowed interval between 0.03 and 0.07 seconds, 50 ms on average [19]. Based on the router advertisement it receives from the default router settings, the MN will configure a care-of address. Neighbour Unreachability Detection (NUD) can also confirm to the MN that an access router is bi-directionally reachable in the case where the up-link and downlink properties may differ.

3.2 Handover System Overview

Before IP mobility support is described, it is important to consider the different types of handovers mechanisms. In an IP network a mobile node (MN) is usually connected via access entities such as WLAN access points (layer 2) or Mobile IP agent (layer 3). When the mobile node migrates from the coverage of one of the access entity point to another, a handover process is performed.

Handovers that mostly occur between access entities of the same network technology are termed *Horizontal Handoffs*. Handovers that occur between different network configuration e.g. WLAN to UTMS are referred to as *Vertical Handoffs* and pose a significant challenge. In addition, mechanisms such as authentication/authorization, billing, location management to handle seamless handovers need to be considered in a heterogeneous multi-access network as shown in Figure 3.2.

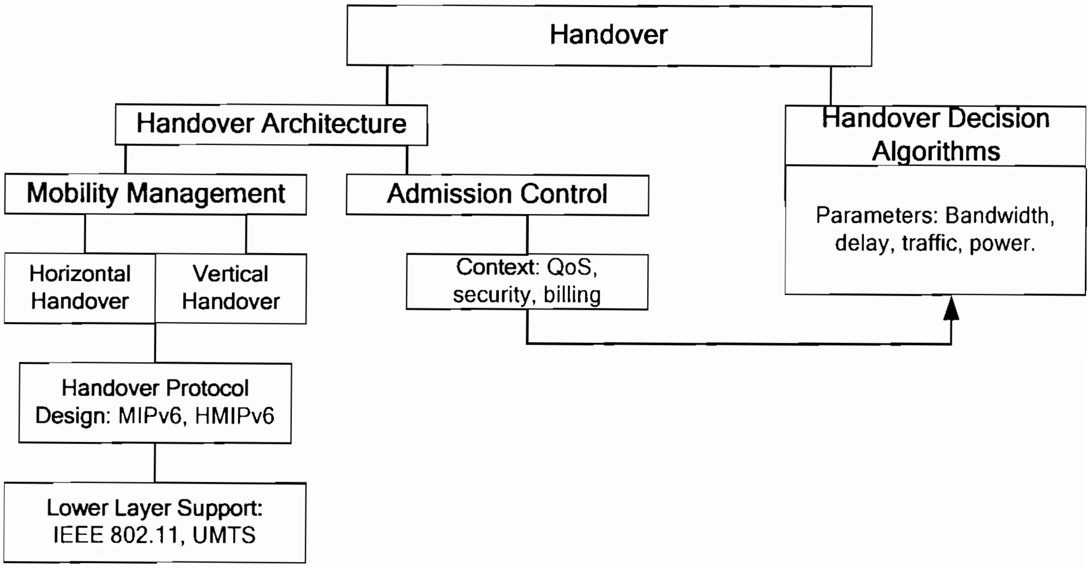

Figure 3.2: Handover Architecture

Handovers can be classified into two categories: *hard* and *soft.* In the case of a hard handoff, a MN can only communicate through one access entity at any given stage in the handover process. On the other hand, a soft handover performed by the MN has the possibility to access the network through more than one access entity simultaneously only if the coverage areas of the access technologies overlap.

Handover Latency

Handover latency is defined as the time that elapses between the last packet/frame, which has been received via a previous access entity, and the first packet/frame that has arrived along a new access entity after a handover [22]. The term *fast* or *low* latency handoff aims at minimizing the time interval during which a mobile node does not send or receive data information. The aim of a *smooth* handoff is to minimize loss of data even though significant latencies will be incurred through buffers. The goal of this mobility study is to achieve a *seamless* handoff that is both fast and smooth handoffs.

In the case of the study, overlap regions with WLAN access points will be used within their access range to prevent the terminal from losing connectivity.

3.2.1 Mobile IP Handoff Overview

As discussed in the previous chapter, a mobile node acquires a new CoA as a result of moving from its current access router or mobility agent in an IP network. Mobile IP handoff mechanism is divided in the three main stages:

- Link Layer Handoff
- Movement Detection
- Registration

3.2.1.1 Link Layer Handoff

Link Layer Handoff involves Link-Layer 2 (L2) mechanisms and precedes Mobile IP handoff. This mechanism is especially important when considering wireless access technologies such as WLAN. Alternatively, a handover decision mechanism can solely be based on Layer 3 indications independently of Layer 2 triggers. This may result in greater handover delays and service disruptions. A MN can, for instance, use signal strength or signal quality information in terms of coverage to determine whether to switch to a new care-of address for a better connection. A link-layer handoff within WLAN device is completed when the mobile device has been associated with its new access router and a new link-layer setup connection is established.

Table 3.2: Link Layer Triggers

Link Layer Trigger	Description
Link Up	Indication that the MN has established a connection with an access point. The L3 process may start sending packets as a link has been established.
Link Down	This indicates that the MN has lost a connection with an access point and the link can no longer be used for data transmission.
Link Quality Crosses Threshold	Preconfigured threshold has been assigned for the link quality for a certain period of time to allow the network layer to prepare for a handover.
Link going Down	Link Down event is anticipated, so network layer must initiate the handover procedure.
Link going Up	This indicates that this trigger may be used for cases where the establishment of radio communication lasts long enough to influence network layer decisions such as network detection and selection (e.g. for avoiding the selection of a network).
L2 Handover Start	This trigger indicates that the link layer receives radio signals with better link quality than to the one which it is currently connected. The MN starts a L2 handover to attach to a new access point.

The movement anticipation as discussed in Fast Handover for Mobile IPv6 is based on the Layer 2 triggers [23]. The Layer 2 trigger contains information based on the link layer protocol, below the IPv6 network protocol stack in order to allow the layer 3 (L3) handover. The table 3.2 above shows the main Layer 2 triggers.

When an access router (AR) receives a L2 trigger, it will find a matching entity identification to an IP address, that is, an access point identification, for example, that will allow it to know towhich subnet the access point connects.

Subsequently, the mobile node uses movement detection mechanisms by exchanging information with the neighbouring AR's to associate itself to any new agent and determine whether an IP-level movement has taken place.

3.2.1.2 Movement Detection

Movement detection is performed to detect that a mobile node is away from its current mobility agent/access router by the reception of IP packets through periodic advertisements. These advertisements carry the updated information about the new link environment and provide the mobile node with its IP configuration such as network prefix, CoA. In Mobile IPv4, Agent Discovery mechanisms use extended standard Internet Control Message Protocol (ICMP) to perform movement detection [24]. Router Discovery with ICMP uses agent advertisement and solicitation messages to detect mobility agents.

A mobile node may use any combination of mechanisms and the primary mechanism used for Mobile IPv6 is IPv6 Neighbor Discovery protocol which includes Router Discovery and Neighbor Unreachability Detection (NUD) [21].

3.2.1.3 Router Discovery

Router Discovery is the mechanism used to detect new routers and on-link network prefixes for mobile nodes. A mobile node within the IPv6 environment may send Router solicitation messages or wait for the periodic Router Advertisement messages. The mobile node keeps an entry within its Default Router List (containing its Prefix List for each network prefix) for each router with its network based on the received Router Advertisement messages. The entries in the list have an expiry time associated with the advertisement.

In the case of the mobile node being away from home, it will select and extract one router from its default router list and use a network prefix as obtained from the Prefix list. From this information, the mobile node configures a new care-of address and registers it with its Home Agent, as described in section 2.2. While the mobile node is away from its home network, it has to quickly detect when the default router becomes unreachable so that it can switch to another default router and configure a new CoA. Neighbour Unreachability Detection is used by the mobile node to detect whether the default router has become unreachable. The mobile node also monitors the receipt of any IP packets from its current access router to indicate whether it is still reachable to the MN. As specified in IPv6 Neighbour Discovery, the mobile node can also check if it is still reachable by its default router through the periodic multicast Router advertisement messages sent by the router.

Within some types of network, lower layer information, e.g. from link layer, may allow the mobile node to determine wireless signal strength or signal quality of the available default routers to switch to. Together with the lower layer mechanisms and NUD, the mobile node will be able to connect to the most appropriate router to provide better connection.

When a mobile node moves from one point of attachment to another, upon movement detection, the node will configure a new care-of address according to the new point of attachment and will send a binding update to its home agent. In the case of a Mobile IPv4 node, it can configure its own CoA based on the movement detection information. If external (stateless) mechanism such as DHCP acquires a CoA, it will incur a significant delay between movement detection and registration.

As specified in IPv6 [19], the MN needs to verify the uniqueness of its link-local address on its new link. This is performed through the Duplicate Address Duplication (DAD) [25] as the MN sends one or several neighbour solicitations to its new address. Once the address has been established, the MN may use either stateless or stateful address autoconfiguration to generate its new CoA. The performance of IPv6 Duplicate Address Duplication incurs delays and may hamper the CoA configuration.

From the ongoing Autoconfiguration draft [20], the MN should perform DAD in parallel with the link-local address and global-link address to avoid additional time to handover latency. In addition to the CoA configuration, the handover delay can further be increased by the Binding

Update procedures. Authentication and registration of the CoA is performed in the last stage of the Mobile IP handoff. These stages add further delays in the Mobile IP handoffs.

3.2.2 Micromobility in Mobile IPv6

Mobile IP suffers from several well-known drawbacks leading to the concept of locality: macromobility and micromobility.

When mobile nodes move between networks of small coverage area, for example, within 802.11 wireless LANs, Mobile IP introduces two causes of latency when handoff are performed:

- Movement detection latency: the time required by the MN to detect change of point of attachment. The latency may be high depending on the movement detection mechanisms required.
- Registration latency: during handoff process, registration messages may travel through the Internet to locate its corresponding HA. This process is subject to significant delays and an increased registration signaling load will be experienced on networks with a large number of MNs.

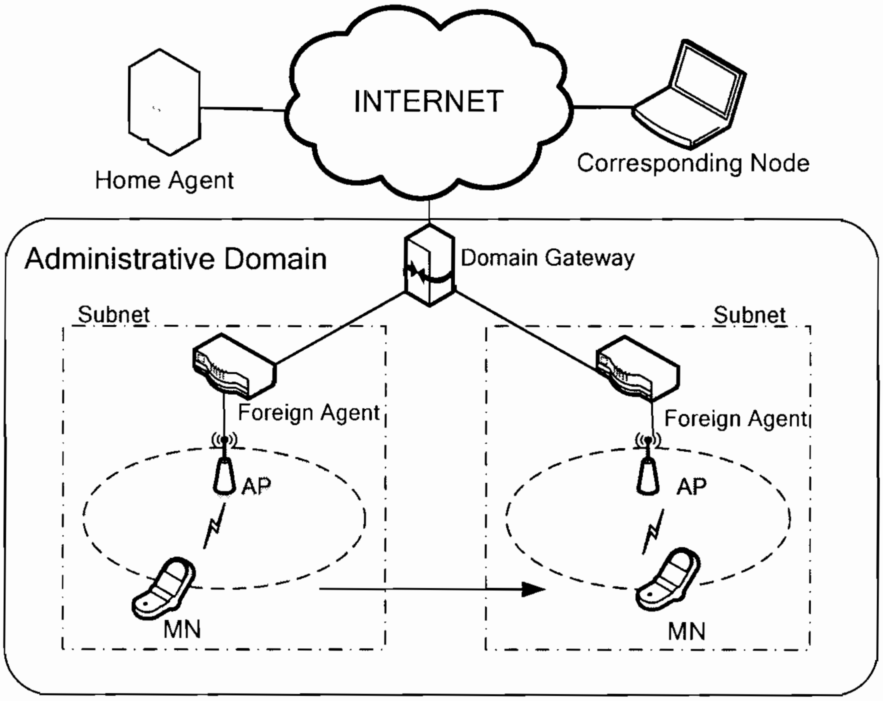

Figure 3.3: Hierarchical Network Infrastructure

The principle of local mobility of a mobile terminal can be achieved by the local movement of the mobile node between its neighbouring networks. The hierarchical access network infrastructure as shown in Figure 3.3 provides an administrative *domain* that is made up of several Mobile IP subnets. Micromobility protocols deal with the localized, intra-subnet movement with a domain, while macromobility deals with the inter-domain mobility.

3.2.2.1 Background on Micromobility protocols

Each mobility proposal as discussed in [26] aims at reducing movement detection latency and to optimize the handover management within a domain. Many different micromobility protocols have been developed such as Hierarchical Mobile IP, Fast handover for Mobile IP, Cellular IP [26] to interoperate with Mobile IP and support mobility management. The focus of the project design is on Fast Handoff protocol which anticipates the handoff process through the radio layer trigger allowing the MN to perform its registration before the handoff actually occurs. A detailed investigation of the other micromobility protocol is beyond the scope of this study.

3.2.3 Fast Handovers for Mobile IPv6

Fast handovers for Mobile IP protocol proposed by the mobile-IP Working Group of the IETF [27] specifies the enhancements to MIPv6 that enable a MN to connect to a new point of attachment more quickly. The protocol aims to reduce service degradation by minimizing the time during which the MN is unable to send or receive IPv6 packets. With the emergence of real-time traffic, it is necessary to ensure rapid handovers to avoid unnecessary latencies (i.e. handover latency) and IP connectivity. In the protocol, the mobile node acquires information about its new access router prior to moving to it. When the mobile node is detected by the new access router, a new link is already established to send and receive application packets. These features are necessary to effectively support real time applications which require smooth continuous data delivery.

Movement detection from link Layer 2 allows the mobile node to know about its new point of attachment before the handoff. Two types of handover mechanisms are handled by the Fast Mobile IPv6 protocol: Tunnel-based and anticipated handover.

Tunnel-Based Handover:

This mechanism depends on a link layer dependence that needs some link layer technologies to trigger the handoff. In this scenario, handover cannot be initiated until the MN has the Layer 2 connectivity to the new access router.

Anticipated Handover:

This mechanism is solely based on network layer information independent of the link layer. Handover can be initiated while the MN still has Layer 2 connectivity to the previous router. Anticipated Handover will only be used in the case of the performance study.

The aim of the protocol is to enable the MN to configure a new care-of address before it gains connectivity to its next access router so that the new care-of address is addressed immediately to establish connection with the new access router.

To implement Fast handover, the following signaling messages are executed between a MN and the access routers. From the point of an upcoming handover, a previous Access Router (PAR) is viewed as the router to which the MN is currently attached and the new Access Router (NAR) is the router to which the MN will be moving.

The signaling messages are as follows and are further described in the sections below:

- Router Solicitation for Proxy (RtSolPr) - MN informs the PAR of a possible connection to a NAR.
- Proxy Router Advertisement (PrRtAdv) – The PAR provides information about the NAR to the MN to facilitate the expedited movement detection.
- Handover Initiate (HI) – Message is exchanged from the PAR to the NAR for the new assigned CoA.
- Handover Acknowledgment (HAck) – Message is exchanged from the NAR to the PAR to confirm the assigned CoA.

- Fast Binding Update (F-BU) – Fast Binding Update is sent from the MN to the PAR.
- Fast Binding Acknowledgement (F-BACK) – Message from PAR or NAR to the MN to indicate that the binding is completed.
- Fast Neighbor Advertisement (F-NA) – The MN sends the F-NA to initiate the flow of awaiting packets when it reaches the NAR.

The following section will describe how the above entities are involved in a Predictive and a Reactive Determined Handover scenario:

3.2.3.1 Predictive Handover

A Predictive Determined Handover is when the MN is responsible for defining and initiating the handover prior to the handover as seen in Figure 3.4.

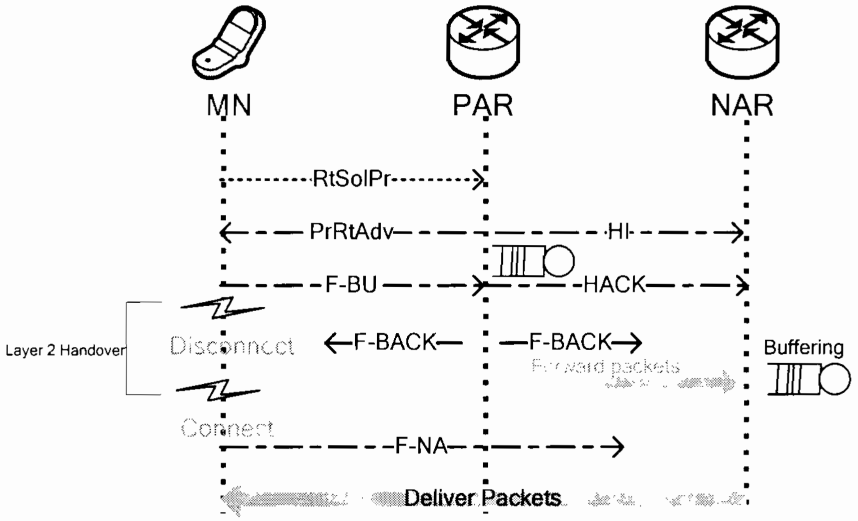

Figure 3.4: Predictive Handover

To initiate the fast handover mechanism, the MN sends a RtSolPr message to the previous access router (PAR) to indicate that a fast handover is required to move to its next point of attachment. The RtSolPr message contains about the Link Layer address or the identifier of its new point of attachment. In response, the MN will then receive a PrRtAdv message from the PAR to inform the MN of its new care-of address that will be used to deliver the packets together with the new access router's (NAR) IP address and Link Layer Address. In addition to the above message, the PAR sends an HI message to the NAR with both the new configured CoA and the old CoA that was used at the PAR. The exchange of information between the routers is to facilitate the forwarding of packets and to minimize the latency perceived by the MN during handover.

The NAR checks whether the new formulated CoA is a valid address to ensure it has no duplicate. If the new CoA is valid, the NAR adds it to neighboring cache entry and responds with a HAck message. In the case that the new CoA is not valid (for example: already in use by another node), the NAR adds a host route pointing to its mobility interface for the old CoA and sends a response to the PAR with a Handover Acknowledgement message i.e. “Handover accepted but new CoA not valid.”

When the MN sends a Fast Binding Update to the PAR to confirm the handover is to take place, a binding is required to forward the packets. The Fast Binding Update is sent depending on the Link Layer conditions at the handover time, and if this is not possible, the Fast Binding update must be sent through after the MN attaches to the new AR.

On receipt of the Fast Binding Update and of the HAck message, the PAR can initiate the forwarding of the packets destined to the MN’s old CoA to either the newly assigned CoA or the NAR. This depends on the Handover Acknowledgement value and to its Link Layer support indications. In such a case, the PAR may delay the routing change until the MN has been disconnected. The PAR acts as a Home Agent with a Home Address being the old CoA and a CoA being either the MN’s new CoA or the new access router CoA.

The MN does not use the new assigned CoA until the Fast Binding Acknowledgement message is sent through a temporary tunnel. The Fast Binding Acknowledgement message can either be received when the MN is still connected to the PAR (in that case MN just needs to inform NAR to provide connectivity) or that the message is sent but not received at PAR (in that case a copy of it will be assigned to MN at the NAR). If the Fast Binding Acknowledgment message is not received at all, this implies that the Fast Binding Update was received by the PAR and the MN must thus resend the Fast Binding Update message to PAR.

As soon as the MN gains connectivity with the NAR, a Fast Neighbor Advertisement message will be sent. This message is to trigger the forwarding of the packets for the MN, assuming that the NAR is aware of the MN or else packets are likely to be dropped. The Fast

Neighbor Advertisement message contains the old and new CoA as well as the link layer address. The NAR will check the link layer address to check if there is a mapping in the Neighbor Cache. If there is no such mapping, the NAR will send a Routing Advertisement to the MN's old CoA with the NAACK option. If the mapping found in either old or new CoA, the NAR will change its cache entry to REACHABLE option and NAACK option is sent with the Routing Advertisement. These options are further described in the IETF draft [27].

3.2.3.2 Reactive Handover

In the case of the Reactive Determined Handover, the PrAdvRt message does not depend on the RtSolPr message as seen in Figure 3.5. This type of handover supports both stateless and stateful CoA configurations.

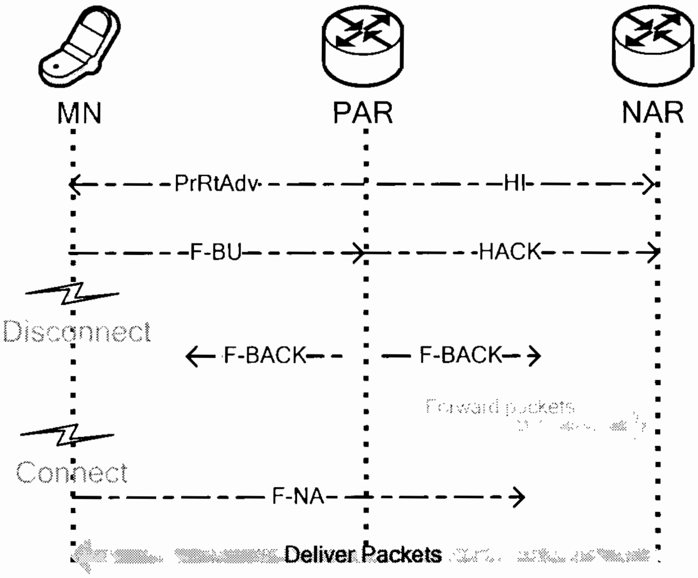

Figure 3.5: Reactive Handover – Stateless

Stateless new CoA Configuration

When the PAR notifies the MN that it is moving to a NAR, it generates a new CoA based on the MN's Interface ID and the NAR's prefix. The CoA is forwarded through using the PrAdvRt message with the Link Layer Address. At the same time, the older sends a Handover Initiate message to indicate the old and newly generated CoA to the NAR.

The rest of the mechanism is the same as discussed in Predictive Fast Handover. As discussed above, the PAR will act as a Home Agent with home address being the MN's old address and newly assigned CoA will be the NAR or MN's new address.

Stateful new CoA Configuration

In this case, the HI/HAck message exchanges precede the PrRtSol message that is sent from the PAR to MN. This mechanism is to correctly retrieve the stateful allocated CoA from the NAR through the PrRtSol message when this information is not available to the PAR. And the rest of the Fast Handover mechanisms are the same as discussed above.

3.3 Application Layer Approach: Session Initiation Protocol

In 1996, the Session Initiation Protocol (SIP) was originally developed in an academic project to the control of unicast and multicast media distribution with a message structure based on SMTP [28]. An increase in interest developed for Internet Telephony at that time and early standardization was concentrated on the use of SIP for Telephony. Numerous standardizations have taken place within different SIP working groups. In the year 2000, the 3GPP (Third generation mobile) also selected SIP to be within the IP communication infrastructure for call signaling.

SIP is an application layer protocol and develops the basis of Internet-centric multimedia communication architecture [6]. It has been standardized by the Internet Engineering Task Force (IETF) to allow a wide variety of sessions to take place between two or more entities over an IP network. A SIP session is defined as any interactive communication that enables converged voice and multimedia services i.e. from a two-way telephone call, multimedia conference session to an instant message exchange. Currently SIP is used in the three main categories: telephony (includes multimedia conferencing), instant messaging and session mobility. In this project, only SIP telephony will be considered and the following sections describe the underlying mechanisms of SIP.

SIP is one of the leading session management protocol that provides support for multimedia services. The Real Time Transport Protocol (RTP) [11] supports the media transport. The Real Time Control Protocol (RTCP) [29] provides a signaling mechanism to provide feedback on the RTP data transfer for Quality of Service management. The Session Description Protocol (SDP) [30] describes the properties of media sessions (e.g. multicast address, bandwidth, media, coding) to the available channels. The Session Announcement Protocol (SAP) controls the periodic announcement of established sessions. SIP furthermore uses of UDP and TCP to convey signaling messages. Figure 3.6 summarizes the protocols related to SIP.

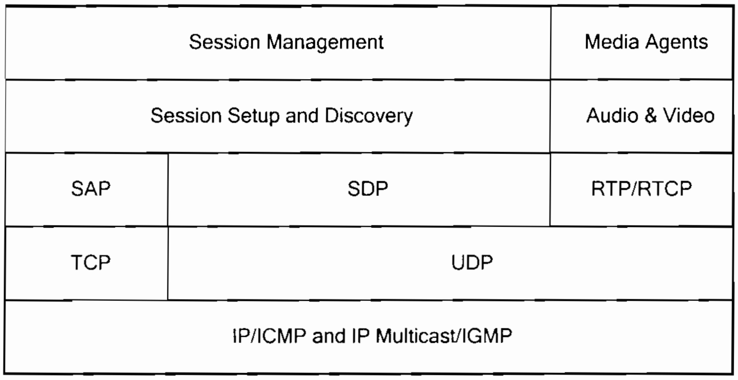

Figure 3.6: SIP Related Protocols

SIP usually runs over User Datagram Protocol (UDP) but may also run over the Transport Control Protocol (TCP) and the Stream Control Transmission Protocol (SCTP). This section will be discussed further in this chapter.

3.3.1 SIP Addressing

SIP typically identifies its users by means of email-like addresses. A user in the SIP environment is addressed as a User Agent (UA) with the following email-like address "user@host": where "user" identifies the user name and "host" is the domain name [5]. The concept behind this addressing form is that they are generated independently from the current location of the UA.

3.3.2 SIP Entities

SIP defines the following entities to establish a SIP connection on hosts and nodes in a network: User Agents, Proxy Servers, Redirect Servers and Registrar Servers.

- User Agent: it consists of two different parts, the User Agent Client (UAC) and the User Agent Server (UAS). The UAC is the entity that generates SIP requests and receives the responses that follow from the request. The UAC is the entity that is responsible for sending the responses to the SIP requests.

- Proxy Server: An intermediary entity used to establish the calls between the UAs. There are two types of proxy servers: the stateful proxy and the stateless proxy.
  - The stateful proxy maintains the transaction state during the processing of the SIP requests. This type of proxy permits the forking of requests, i.e., sending copies of the request with different URLs to different destinations. The forking mechanism is useful if the proxy server does not know the final destination of the request. The mechanism can operate the search of request either in parallel or through sequential search.
  - On the other hand, the stateless proxy does not support the keeping of the transactions states and is mostly used in SIP backbones. It simply forwards the requests to another server, without taking to consideration the reliability of the transaction.
- Redirect Server: it is a UAS that generates responses to the requests it receives and returns a response to the client by indicating an alternative set of addresses to forward the request. In contrast to the proxy server, it has a lower state overhead with more processing messages.
- Registrar Server: A server that stores and saves the information of the accepted requests to provide a location service and to generate address translation within its own domain. This server implements a redirect and a proxy server with a built-in registrar. The server can act in the two modes i.e. as a proxy or redirect server depending on the request configuration.

3.3.3 SIP Messages

SIP messages are textual and the header of each SIP message contains specifications (e.g. message type, protocol version, destination address) about the network. The payload of a SIP message consists of the SDP specifications. Within a routable infrastructure in an IP environment, the main signaling messages as shown in Table 3.3 are required to establish a session and to negotiate for the compatible media type required for the conversation.

Table 3.3: SIP signaling

Signaling Messages	
INVITE	SIP UA sends an INVITE request. The message body of this request contains the Session Description Protocol (SDP) of the available media channels of the UA.
ACK	When both parties have the SDP descriptions and that the media channels can be established, the caller UA acknowledges the successful receipt of the response.
BYE	To terminate the call, either party sends a BYE request.
Re-INVITE	If one of the parties wants to change the media type, it can send a re-INVITE message to establish a new call with the SDP corresponding to the new media type.
OPTIONS	OPTIONS message allows query to the SIP server about the user's service capabilities.
CANCEL	This message cancels any pending requests.
REGISTER	A message request that allows registering to a SIP server.

3.3.4 SIP Mobility Support

SIP supports different types of mobility mechanisms at the application level, ranging from high-level mobility i.e. Service mobility, Personal mobility and Session Mobility to Terminal mobility [5]. Some of the mobility mechanisms are provided by the protocol while others can be implemented through SIP extensions.

Service Mobility

Service Mobility allows the user to maintain access to its subscribed and personalized services consistently even when it is connected to a foreign network service provider. SIP can be used in such mobility by updating the home server the user identifies itself to, for example, by correlating its domain name to the domain server. The SIP user agent also has the ability to upload its timestamped configuration information and the server will update the current version.

Personal Mobility

Personal mobility maps a logical address to the user to allow addressing a single user at different points of access. SIP forking proxies are used to reach the user at any of the devices such as PDA, PSTN phone through the same user name.

Session Mobility

Session Mobility allows the user to maintain an ongoing session while changing terminals i.e. from a mobile phone to laptop within a personal network area (PAN). Session Mobility can be supported by SIP, into different mechanisms. The simplest approach is that the end systems that are streaming a real-time application convey their IP addresses and the ports from the primary end system to the other using a new INVITE request. The other mechanisms such as third-party call control and the mechanism implemented through the REFER method are discussed in [28], [31].

3.3.5 SIP Mobility – Terminal Mobility

Terminal Mobility

To support terminal mobility at the application level, it is necessary that the network support IP mobility [15] to be able to maintain connectivity whilst the user is changing points of attachment i.e. access points, base station terminals. Terminal mobility is divided into three stages as follows [6]:

- Pre-Call mobility – Mobility before the call: this requires that the mobile node receives its new address before making or accepting a call. Therefore each time a new IP address change is detected by the terminal, a new register is created in the SIP Registrar server and updates the new parameters. When a correspondent node requires a connection with the mobile node, the CN will first send an INVITE request through its proxy to the "home" SIP server to request the MN's current location. In the case that the address of the mobile node is not obtained, the proxy server can then multicast the INVITE request to a larger scope. Once the location address is obtained, the CN sends an INVITE request directly to MN as illustrated in Figure 3.7. The other signaling mechanisms are as discussed in the previous chapter.

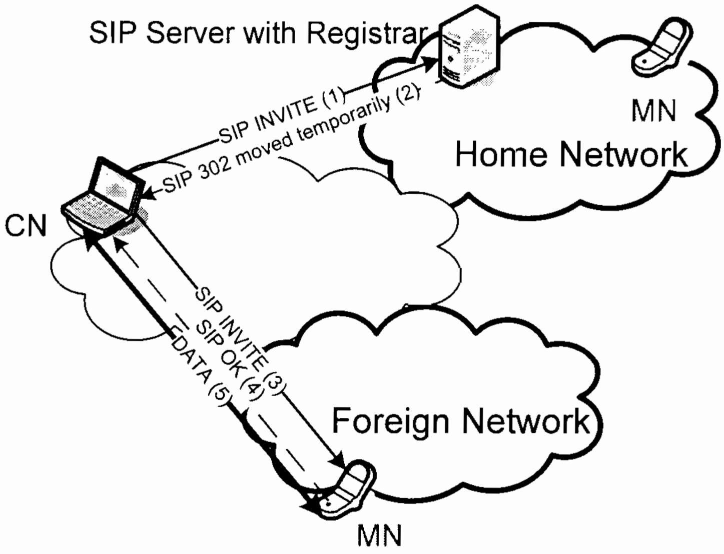

Figure 3.7: SIP-based Pre-call Mobility

- Mid-Call Mobility – Mobility throughout a call: if a mobile node moves during the session as shown in Figure 3.8, the terminal will detect a network address change (This is achieved through a DHCP server or a variant of it) and will send a new INVITE message (Re-INVITE) with updated SDP to the correspondent node without going through intermediate SIP proxies. The INVITE request will inform the remote user of the change in the session parameters with the new IP address to forward the packets correctly. The new invitation also updates the current ongoing session description, in the case of a real-time session this is achieved through a RTP translator. The SIP registrar server creates and updates a new register for the new discovered session parameters. The significant drawbacks which will be further discussed in the next chapters on the SIP-based mobility mechanism are the disruptions caused on the call-setup and the absence of mobility management support for long-term TCP connections.

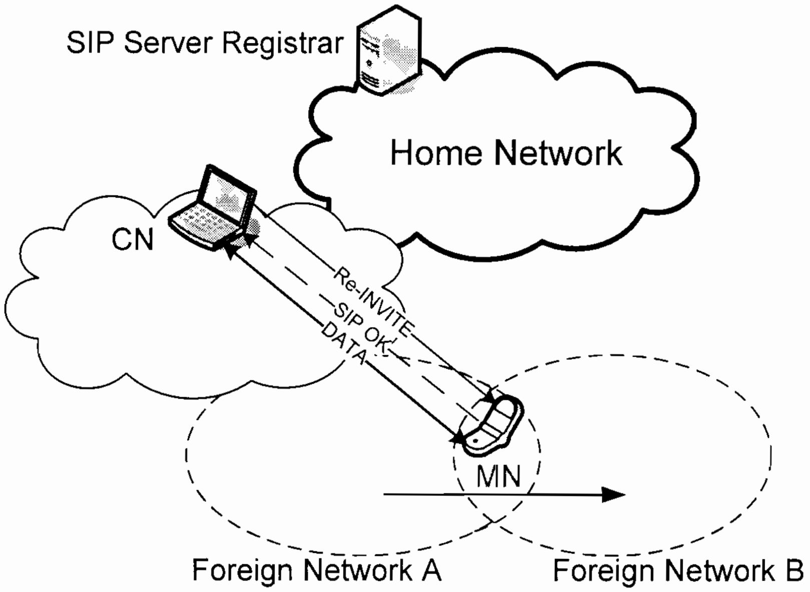

Figure 3.8: SIP-based Mid-Call Mobility during handoff

- Error Recovery: Error recovery is an automatic mechanism for error recovery in the case of the address connection breaking. This mechanism assigns different IP addresses to the mobile terminals and registers the new addresses in the Registrar server [5].

3.3.6 SIP-Based Handovers Mechanisms

There are several ways to achieve fast handovers through SIP mechanisms as outlined in [32]. The following subsections will summarize the SIP-Based Fast Handovers Mechanisms.

RTP Translator

As mentioned in the section of terminal mobility, when a mobile node moves from one domain to another, and obtains a new IP address, it sends a Re-INVITE to the correspondent node. This mechanism is to ensure that the new traffic is forwarded to the mobile node's new destination. In the event that the SIP Re-INVITE message being delayed, the transient traffic will still be forwarded to the old destination. An RTP translator is designed at each domain of the mobile node to provide support to forward the transient data. As the mobile node moves from one subnet to another, and the new address IP is obtained, it sends a Register message to its SIP Registrar (located within the mobile node's home network). The SIP server checks and updates its database, and sends a message to the RTP translator. The RTP translator acquires the new IP address of the mobile node and forwards the transient traffic. The RTP translator has a time out mechanism limited by the multicast address which is the duration to forward the packets. There are several de-activating mechanisms as outlined in [32] that can be triggered as soon as the correspondent node stops sending the packets to the mobile's node home network.

SIP Outbound Proxy

The SIP outbound proxy is another technique to support the fast-handoff movement of a SIP mobile node. SIP INVITE request traverses the SIP outbound proxy (i.e. within visited network). The proxy accesses the Session Description Protocol information that contains the mobile host's media and port configurations. Therefore this approach is to the advantage of the RTP or NAT (Network Address Translator) as it simplifies their configuration. The INVITE request information is stored until a new Re-INVITE message is released.

Back-to-Back SIP User Agent (B2BUA)

The Back-to-Back SIP mechanism consists of two SIP user agents: one user agent receives a SIP requests and maps the SDP parameters and the other agent re-issues the request.

As the mobile node moves to new domain, it sends both a SIP INVITE and Re-INVITE message to the B2BUA. A session to deliver the transient data to the CN (via RTP or NAT translator) is established between the B2BUA and the node. The connection continues until the mobile node moves to another domain.

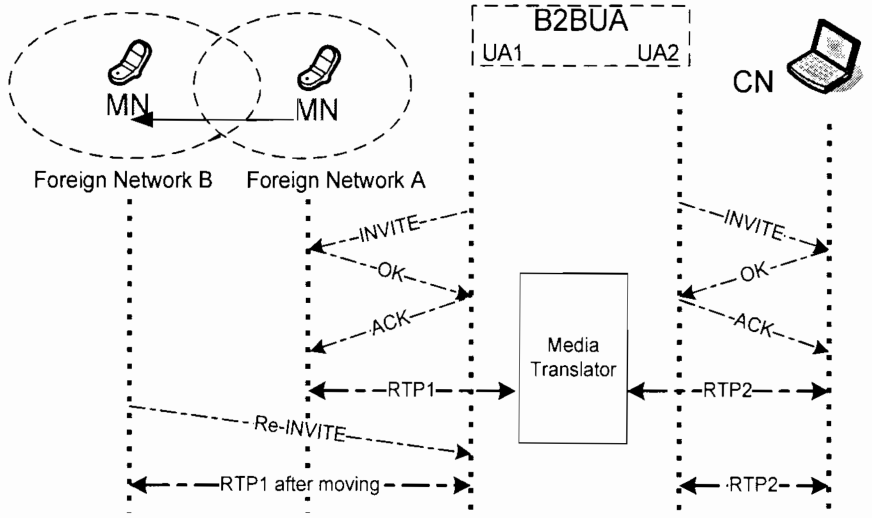

Figure 3.9: B2B flow mechanism

Figure 3.9 shows the flow mechanism used in this approach. In this mechanism, the SIP request first needs to be addressed to the B2BUA. If not, end-to-end encryption may prevent the B2BUA from setting up the call.

3.4 Comparison of Mobile IP and SIP

As outlined in the previous chapters, both protocols support different types of drawbacks in terms of handover mechanisms to support real-time services within the mobility framework.

3.4.1 Problems related to MIP

A well-known problem within Mobile IP is known as the Triangular Routing. Packets from the CN to a MN which is in a visited network have to be routed from the CN to the HA. The HA then encapsulates the packets in a Mobile IP tunnel, sends them to a foreign agent which in turns forwards the packets to the mobile node. Triangular routing increases the traffic load on the network. Although the Routing Optimization mechanism [33] alleviates the use of triangular routing mechanisms by sending the binding updates directly to inform the current location of the MN to the actual host, the following drawbacks are incurred:

- Changes in the IP stack of the CN are required in the route optimization mechanism, in order to allow the encapsulation of the IP packets and to store the current care-of addresses of the FA or the MN
- Extra delays are placed on the CN as only the HA may send binding updates to it. These additional delays may be unacceptable to delay sensitive traffics. After receiving the updates, the CN will only then find out where to send the packets and during which the previous FA must forward the packets to the current location. In the case of Mobile IPv6 networks, deployment of FA is unnecessary and their functionality is performed through IPv6 access router.
- As mentioned above, the MN must rely on its FA to forward packets until the CN has the binding update set up. This may incur extra delay on the network infrastructure. The HA and foreign agents are potential bottlenecks as they accommodate large amount of traffics and users.
- When a MN uses external stateless mechanisms to configure its co-located CoA in IPv4 network, the mechanism imposes delays, for example using DHCP will require query to local DHCP sever and messaging overhead signaling introducing additional delays. On the other hand, IPv6 Address Autoconfiguration allows a MN to efficiently configure its CoA, but its main drawback is that of Duplicate

Address Detection (DAD) that is performed on all global IP addresses. The DAD mechanism is performed both on the link-local addresses and global addresses respectively. This proves to be an efficient process for the MN since several broadcasts are issued for a reply.

Mobile IPv6 contains a number of facilities compared to Mobile IPv4 that improves mobile management in the IP network. While Mobile IP comes inherently with wireless network architectures and provides transparency to the user location, but it comes at a heavy cost with new network entity deployment, sub-optimal routing and latency in terms of handoff and overheads. The main challenge for Mobile IP is its poor performance during handover. To overcome some of these problems, micromobility protocols have been proposed. Fast Handover for IPv6 is the focus of the project although it is not best suited for all scenarios. These issues provide an avenue for further study. At this point of time, Fast Mobile IP handover can lead to network degradation and significant interruptions in the traffic and packet routing may be observed.

3.4.2 Problems related to Session Initiation Protocol

From the application layer perspective, the SIP-based mobility approach offers several advantages and has already been adopted by the Third Generation Partnership Project (3GPP) in its IP Multimedia Subsystem (IMS) on IP network. Like all protocols, it continues to suffer from significant drawbacks as outlined below:

- SIP cannot support long-term TCP connections as the end points of a TCP connection cannot be kept constant within the SIP mobility framework. On the other hand, the transparent mobility in Mobile IP keeps the long-term TCP connections valid during the movement of the mobile node.
- Disruption delays may be caused when the moving MN is in an overlapped region of access points.
- External Address Configuration is integral in SIP-mobility, as the MN always needs to acquire an IP address, for example, via DHCP server. Large amounts of delays are introduced and signaling messages also add up to the delays incurred.

- Since SIP is an application layer protocol, the SIP-based messages may not be served with the highest priority in terms of their associated components and this may introduce further additional delays.

The table below summarizes the main comparison between each protocol and will be further discussed in the next chapters.

Table 3.4: Comparison between Mobile IP Mobility and SIP Mobility

	IP	SIP
Layer	Network Layer	Application Layer
Address	IP address	SIP url
Update	Binding Update	Re-Invite message
Registration	Home Agent	Proxy server
Services	TCP/UDP	RTP/UDP

To conclude, the main shortcomings of each protocol that are used in the project are discussed above and will be taken in consideration the architectural design. Each protocol contributes high handover latencies depending on the different protocol operations in overlapped networks. When considering transporting delay sensitive applications over a network, the end-to-end delay, delay variation (jitters) and packet loss may highly affect the traffic flow. To overcome delay and packet loss problems, a Fast Mobile IP framework is proposed to allow the anticipated handover mechanisms and operate in an optimized routing mechanism. The proposed framework operates by acquiring a new IP address each time a mobile device moves from one link to another through external address configuration, for example, DHCP. The proposed design will reside mostly as an application layer protocol with UDP/RTP/IP traffic. The basic concept of the design is to complement each protocol's performance, i.e., Fast Mobile IP and SIP features respectively based on the kind ofapplication.

Chapter 4

4 Mobility Design

The main goal in a network mobility framework is to minimize handover delays as much as possible, i.e., while moving from one point of attachment to another within a wireless heterogeneous network. Real-time applications such as VoIP applications are highly delay-sensitive and delays, jitter and packet losses will affect the nature of the voice traffic. To provide a complete mobility management framework for real-time applications, it is necessary to combine both network layer protocol FMIPv6 and application layer protocol SIP, in a way to complement each protocol features based on the kind of application. This chapter focuses on merging the traditional protocol schemes to form an integrated protocol scheme to support the handover procedures in overlapping networks.

A general architecture is proposed that aims to build up a mobility framework that will cater for real-time applications in the Next Generation IP Networks (NGN).

4.1 Multi-Layer Approach

A Multi-layer approach explores the contributions from different types of layers to provide an optimized extended mobility framework as compared to a single-layer approach. Hybrid architectures and cross-layer signaling will provide the support across the layers for an optimized mobility framework.

The architectural design of the proposed mobility framework in this project aims to provide IP-based handoff management from network layer Fast Mobile IP and SIP at the application layer. In that way, it will allow intrinsic connections between low-level and high-level mobility. A case scenario that entails collaboration between the two layers could be as follows: Within a heterogeneous network, while a mobile phone is switching its session from a cellular network to a WLAN connection, session and terminal mobility, i.e. from a UMTS to WLAN can happen simultaneously as illustrated within the mobility framework Figure 4.1.

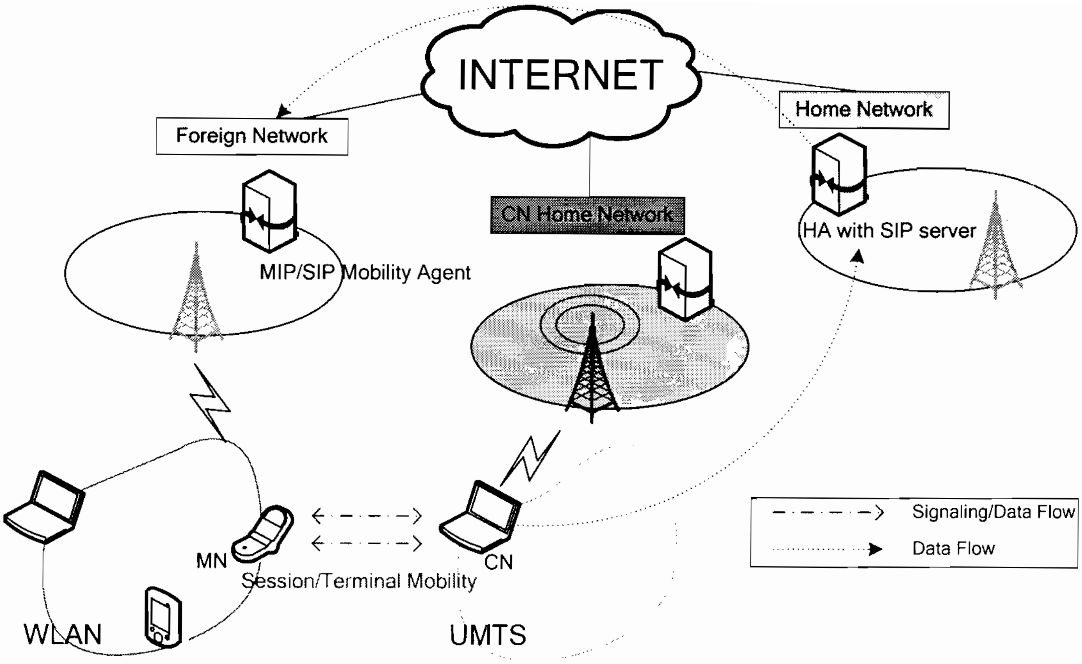

Figure 4.1: Example of MIP/SIP Network Model

4.2 IP-Based Handover

Mobility-incurred issues can best be adapted to through the monitoring of the physical and link layer channel conditions. In the case of an overlay network environment, the conditions reported can be used to determine an imminent vertical handoff to an underlying or overlay network. Link layer uses different Medium Access Control (MAC) mechanisms to trigger the handover and report about the impending handover to the network layer in advance. This allows the network layer to prepare for handover by registering with the target point of attachment before layer 2 handover is complete leading to shorter handover delays. Expediting an IP-based handoff to the network layer can improve handover performance.

The aim of Fast Handovers for Mobile IPv6 protocol specification is to enable the MN to configure a new care-of address before it moves to a new access router. In that way, the new care-of address will allow the MN to connect to the new access router immediately layer 2 (L2) handoff is complete, with minimal interruption to packet flow. The layer 2 triggers that happen before the handover contain the identity of the access routers present in the overlay networks. The trigger consists of beacon signals from the access routers in the vicinity of the MN.. In our implementation of fast handover over 802.11 WLAN, the handover is controlled by the MN based on the reception of beacons. The MN initiates the process of acquiring the new CoA as soon as handover trigger is received and processed. This is done in a way that ensures that no duplicate or invalid addresses are generated. In this design model, both mobile and network controlled handovers are possible.

4.2.1 Address Configuration

After completing the layer 2 handoff, address configuration as discussed in chapter 2 may follow stateful i.e. through DHCP or stateless address reconfiguration procedures. In any of the case, DAD is needed to verify the uniqueness of the address and the process consumes a lot to time to the whole handover procedure. In the design model, DAD is deactivated assuming that the probability of duplicating addresses is very low.

4.2.2 Registration within Home Agent and SIP Registrar

The purpose of the registration in Mobile IP is to inform the MN's Home Agent of the new care-of address through a Binding Update (BU) message. In the same way, the MN can also inform the CN about the new care-of address to allow the CN to appropriately forward/send the packets destined to the MN. In the case of TCP or non-SIP applications, the connections can be maintained without a disruption.

An extension of the Home Agent specification is proposed in the design model in order to co-locate the mechanism of the SIP registrar. This is intended to minimise the delay that might occur when HA is notifying the SIP registrar of the new CoA. For the purpose of the project, it is necessary that during a SIP session re-establishment, the CN is informed of the MN's new IP address so that it can communicate directly with the MN. In order to do so, a binding mechanism between the new CoA of the MN and the user level identifier is required to update the current location of the MN. Once the current location is updated, the SIP proxy and SIP Redirect server database can be updated. The Domain Name System (DNS) records and helps in finding SIP proxies responsible for routing the SIP messages to the destination domain.

4.3 Functional Design based on Fast Handovers for Mobile IPv6 and SIP

In this section, a description of the SIP-based mobility management supporting anticipated handover at the IP layer is described in next generation heterogeneous wireless networks.

In a wireless overlay network, handovers may be anticipated based on link-layer triggers that indicate the coverage status of the new network. As described above, the Fast handover for Mobile IPv6 protocol supports the predictive handover mechanism. This mechanism will enable the MN to know to which access router it will next move to and to configure a new care-of address before regaining IP connectivity to its next access router.

The design also supports the MN initiated handover as the MN is aware of the current active network interfaces. When a link-layer trigger is received and processed during an ongoing session before the MN moves to its next point of attachment, a new CoA is formulated to associate itself to the new access router and start the IP-level movement at layer 3. The new formulated care-of address is auto-configured to be the new CoA address of the MN. This can either be done through the IPv6 stateless address configuration mechanism [20] which allows reducing address configuration delays or through DHCP servers. The MN then proceeds with pre-registration with the HA before the handover actually takes place. The notification delay to the SIP registrar servers becomes negligible as it forms part of the HA network.

4.3.1 SIP Session re-establishment

After acquiring a new IP address before handover, the MN as a SIP client initiates the handover procedure by sending a re-INVITE (1) message to the CN as shown in Figure 4.2. The SIP re-INVITE message initiates the registration within the SIP registrar at the home network of the MN and carries the updated SDP parameters to the CN. As a result, call parameters are re-negotiated on an end-to-end basis [34] with the SIP proxy server and SIP Redirect server as an intermediate to support soft handover. In this scheme, end-to-end negotiation protocol [35] is implemented within the SIP proxy together with SDPng (SDP extensions) for Quality of Service coordination. The session re-establishment allows the CN to redirect all its ongoing media streams and signaling messages directly to the MN's current IP address as it attaches to the new point of attachment.

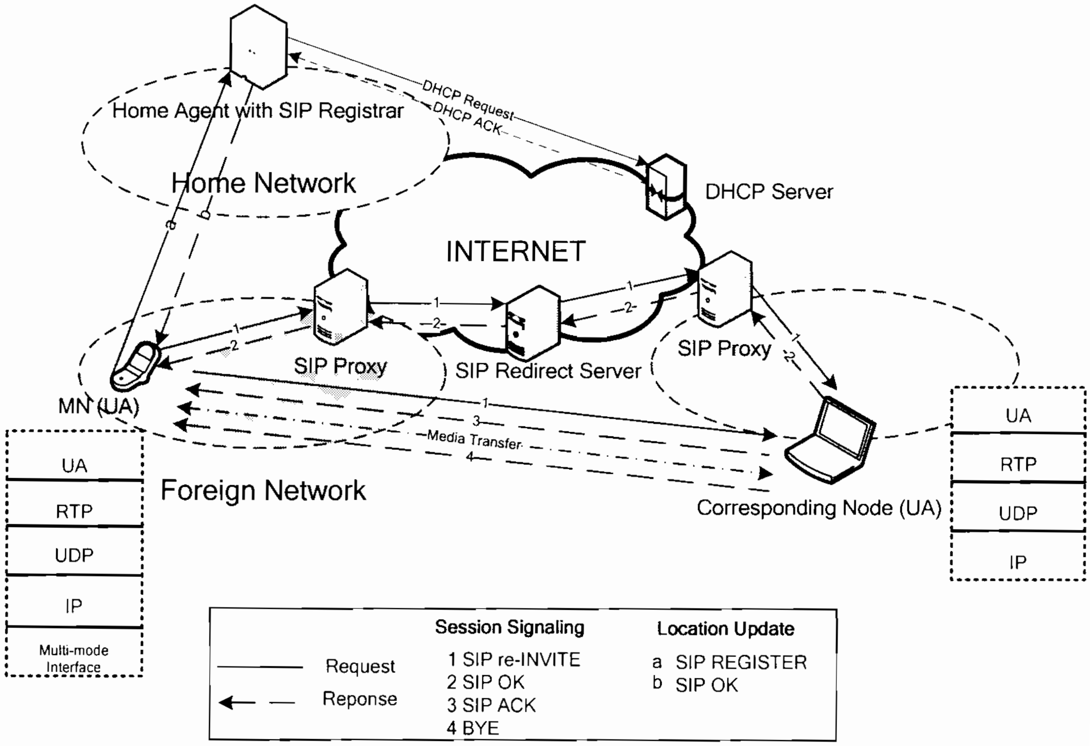

Figure 4.2: Signaling Procedures of SIP within integrated architecture

The re-INVITE (1) message contains the new IP address and the updated contact field where the MN will in future receive SIP messages. If the CN responds with a SIP OK (2) message, agreeing to re-INVITE request, the MN will in turn respond with an ACK (3) to complete the SIP messaging before data transfer. For multimedia applications, it is necessary to decrease delays, packet loss as much as possible, and the integrated schemes aims at avoiding triangular routing and any kind of encapsulation mechanism during the ongoing calls.

4.4 Handover Policies for different scenarios

Different types of mobility scenarios need to be considered in the architecture according to the pre-defined handover policies from the mobility policy table [36]. The mobility policy table is based at the MN and depends on the mobility scenarios of the node. Other parameters such as service providers, network type and Quality of Service (QoS) parameters specific to the application transfer are also considered by the mobility policy table.

As illustrated in Figure 4.3, as the link layer trigger is received by the MN, it will be able to determine which protocol scheme i.e. SIP, FMIPv6 or proposed FMIPv6 and SIP mechanism will provide best support in terms of handover delays to support the type of application during handover. In the case of a terminal handover that is investigated in the project, differentiation between the types of applications needs to be considered. The classification of the applications is categorized as real-time/non-real-time applications.

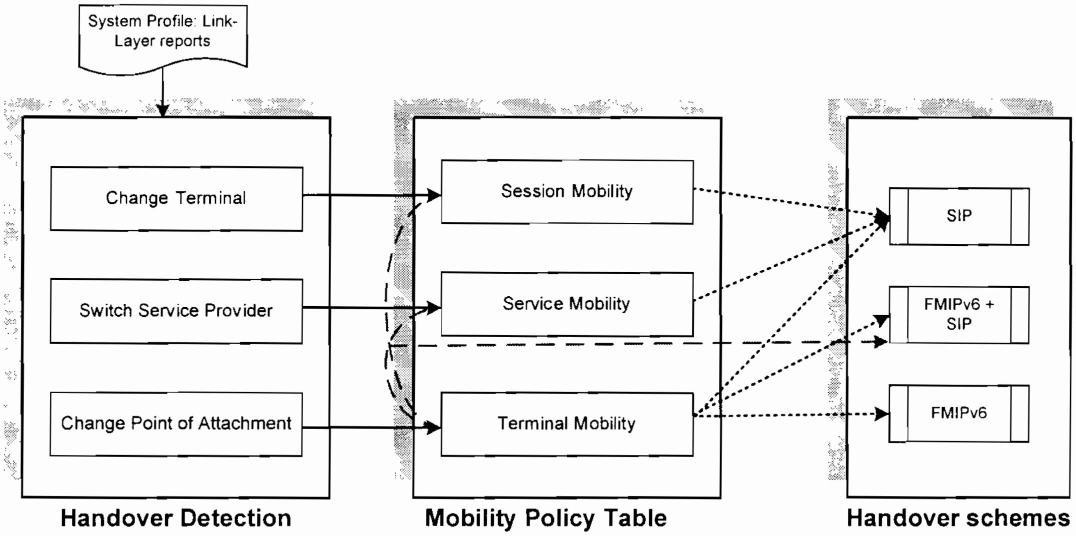

Figure 4.3: Handover Policies

4.4.1 Case 1: Non Real-Time: TCP Application

The significant drawback in the proposed architectural scheme is the absence of supporting long-term TCP connections within SIP mobility. TCP application is considered as a transparent application and Mobile IP can support such applications. In that case, when a TCP header is detected, packet delivering will be performed through Mobile IP mechanism. A tunnelled route in Mobile IP will be initiated for these sessions before the MN moves to the new network. All the packets from the old access router are tunnelled via the Home Agent (in the case of IPv4) of the MN with the likely results of increased delay and traffic load.

4.4.2 Case 2: Real-Time: RTP over UDP Applications

Application layer protocol SIP can support UDP packets without needing to know how the underlying access network transports the packets. When a terminal handoff is taking place, the integrated scheme of FMIPv6 and SIP will provide best support to the real-time applications which are highly delay sensitive as discussed above.

4.5 Message Flow within architecture framework

The proposed architecture has both FMIPv6 and SIP to be invoked simultaneously to provide an integrated handover mobility scheme as explained in the following message flows.

4.5.1 FTP Transfer from CN to MN – FMIPv6

This timeline will occur based on the mobility policy table from the MN, i.e. for the case of a TCP data transfer.

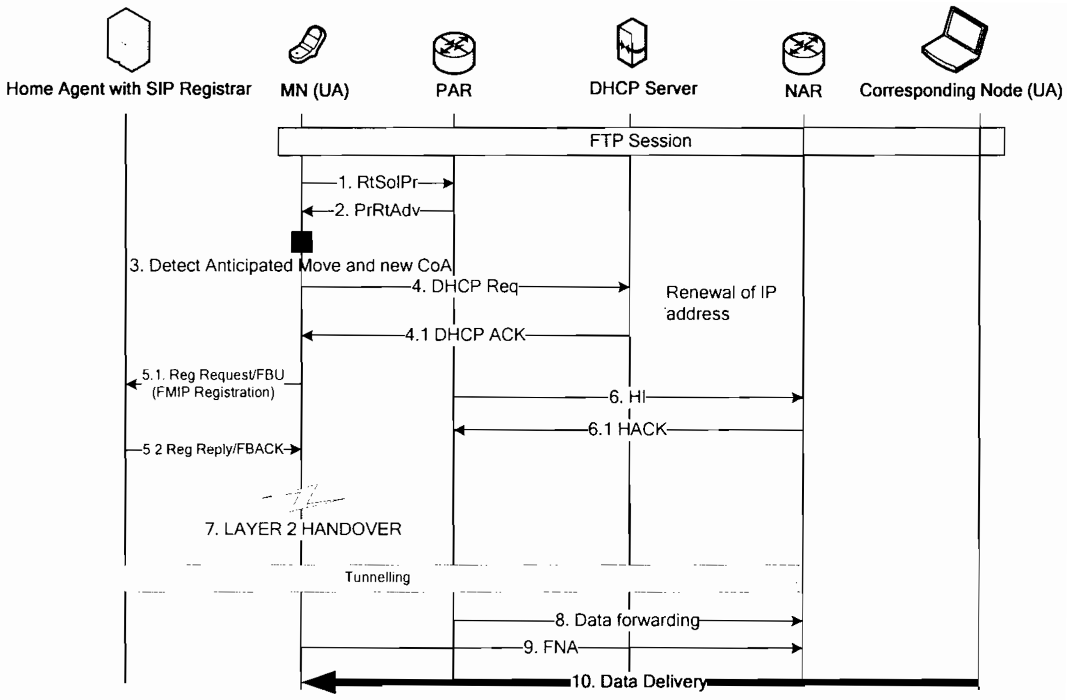

Figure 4.4: Handover Signaling Flowing using FMIPv6

In Figure 4.4, the handoff delays associated will only be related to Fast Mobile IPv6 protocol.

4.5.2 SIP Mid-Call Mobility – RTP over UDP applications

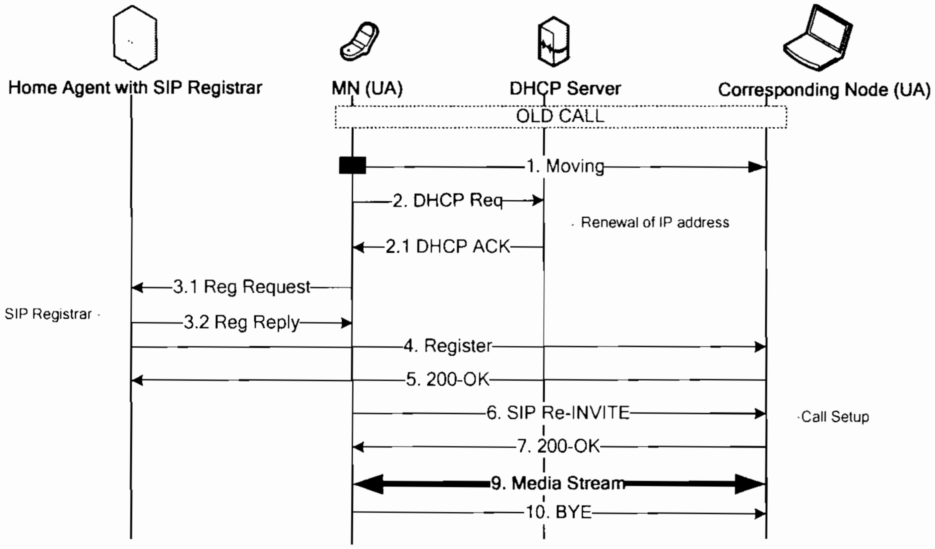

Figure 4.5: SIP Message Flow during mid-call mobility

Based on the mobility policy table for a session or terminal mobility setup, the SIP message flow will be triggered. In this case, the anticipated handover mechanism cannot be used as the SIP signaling mechanism cannot detect movement detection as shown in Figure 4.5. Note that SIP protocol sends a request to configure new care-of address only when it is out of the overlapped network, thus bringing significant delays to the data transfer.

4.5.3 Streaming VoIP traffic from CN to MN – FMIPv6 + SIP

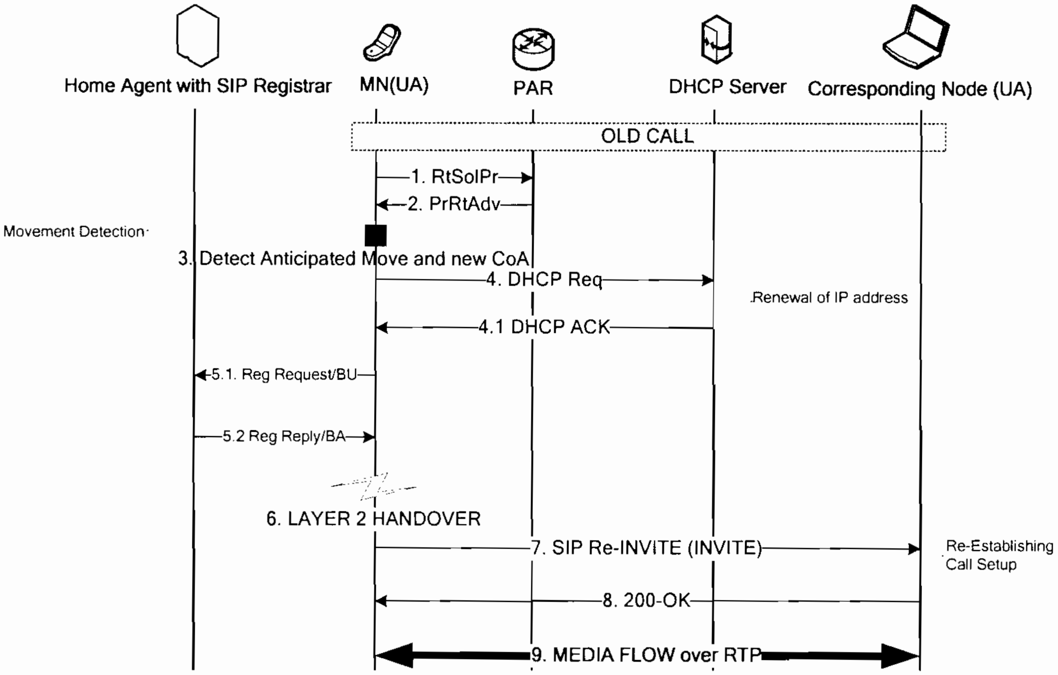

Figure 4.6: Handover Signaling Flowing using FMIPv6 + SIP

The design as discussed above aims at reducing the signaling load by integrating the redundant messages from both protocols for complete message registration for ongoing calls. The transfer of the media flow is done through SIP.

4.6 Delays incurred within architecture framework

Different delay time parameters involved in the design affect the rate of packet losses and handover delays. To determine the mobility performance of the design, FMIPv6 and SIP signaling delays are considered in Figure 4.7.

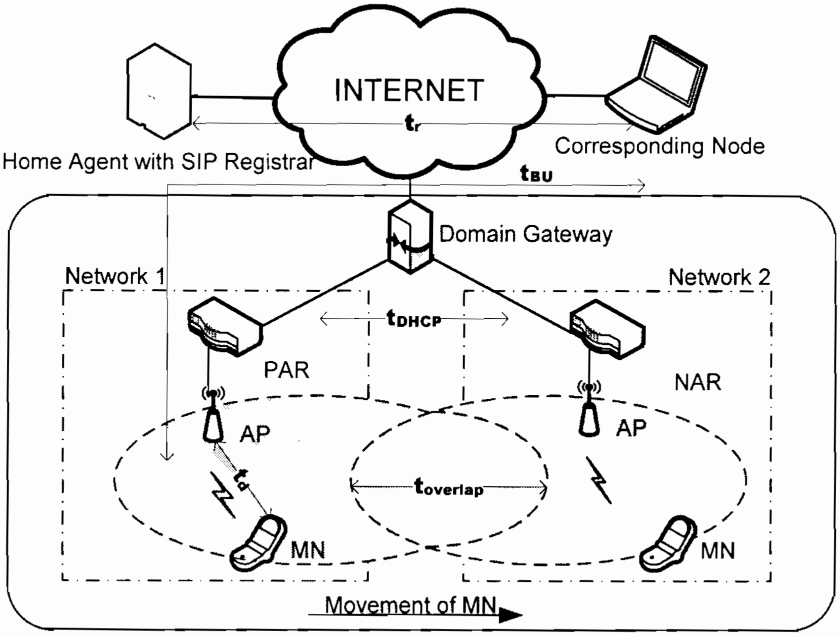

Figure 4.7: Delays considered within architecture

Detection Time (t_d): is the time from when a MN is under the coverage of wireless access network to the instant that pro-actively detects movement detection through the Router Solicitation for Proxy (RtSolPr) of a possible new access router and link-layer L2 triggers. This delay is attributed to the low layer performance of the network. The anticipated handover through FMIPv6 mechanism is used to help minimize the impact of the handover latency.

Overlap Time ($t_{overlap}$): is the time when a MN is located in the overlapped area between the adjacent access networks.

DHCP Time (t_{DHCP}): is the time to configure new IP address and to configure the binding update. In the design approach, Duplicate Address Detection is not considered in this disruption time. Some empirical results indicate the renewal of IP address of around 160ms.

Registration Time (t_r)**:** This is the time required to send the binding update to the HA to update the SIP registrar. Note that MIPv6 sends a binding update to the home agent and the CN, and to receive the first packet, assuming a binding acknowledgement from the home agent was received. SIP re-INVITE message and SIP re-Registration message are used for the binding update schemes in the integrated architecture. Specifically, the re-INVITE message is used during soft handover to announce the newly acquired IPv6 address.

Binding Update Time (t_{BU}): is time for the binding update delay at the Home Agent with the SIP registrar.

End-to-end delay between MN and CN ($t_{MN \Rightarrow CN}$): is the time for the round trip transmission delay of signaling and packet delivery.

As defined in chapter 2, handover latency from a MN perspective is defined as the amount of time to initiate disconnection from the old access point of attachment to receive the first packet from the new network access point. Thus, from the discussion above, the total handover latency (T_H) across the integrated architecture can be given as follows:

$$\mathbf{T}_H = \mathbf{t}_d + \mathbf{t}_{overlap} + \mathbf{t}_{DHCP} + \mathbf{t}_{BU} + \mathbf{t}_r$$

The handover equation suggests the trade-off involved in optimizing the detection time during vertical handovers. Signaling and delays introduced by AAA (Authentication, Authorization and Accounting) procedures strongly depend on the protocols chosen which are ignored in the integrated architecture.

The end-to-end delay in the proposed integrated architecture may include signaling propagation delay, queuing and processing delays incurred within the network components. Extra delay can be caused through message loss, congestions in wired networks and fading in wireless networks. SIP signaling delays accounts during the SIP signaling messages used to establish and maintain the call set up with SIP servers and during the re-negotiation of session parameters

That is within the mobility architecture, end-to-end delay may change due to the network conditions i.e. propagation characteristics of time-varying channels, signaling traffic loads and alterations in the movement of the MN or CN. Delay between the MN and the CN is factored by the different parameters as outlined above.

4.7 Handover Signaling Load

Network signaling traffic load incurred for a terminal handover within the integrated scheme includes IP address distribution, binding updates and home registration with addition to the signaling messages inherited from the traditional schemes. In the real dynamic world for mobility management, these network signaling traffic loads may not affect the system since the integrated scheme combines redundant messages of the traditional schemes. The additional messages used in the proposed scheme aim at supporting more reliable and precise handover procedures for real-time communication. A detailed investigation of the delays and signaling loads affect the integrated scheme is discussed in the next chapter through simulations.

Chapter 5

5 Experiment Implementation in NS-2

This chapter describes the underlying simulation platform used in the Network Simulator version 2 (ns-2)[37] to support the proposed integrated architecture (as discussed in previous chapters). This evaluation framework supports Fast Mobile handovers and SIP signaling messages. Real-time traffic and VoIP application will be supported to illustrate and compare the proposed architecture to the existing schemes.

Developed by the VINT research group, ns-2 is a discrete event object-orientated simulator, written in C++, with OTcl (Object Tool Command Language) as a network interpreter. The architecture of the simulator allows simulation routines at a software level implementation in C++ while OTcl provides the interface to describe the topology of the simulated network.

5.1 Evaluation Framework Overview

The protocols extensions that have been presented in ns-2 are described in this section. The base ns distribution in the experiment setup is ns-allinone-2.27, supporting the ns wireless extension for basic Mobile IP protocol operations.

5.1.1 Fast Mobile IPv6 protocol implementation

Fast Hierarchical Mobile IPv6 protocol extension released by Hsieh [38] is adapted for the case of the simulation to support Fast Handover for Mobile IPv6 proposal.

The ns node entity in ns-2 has been modified as outlined in [39] to support the use of the encapsulator/decapsulator modules for the hybrid wireless-wired nodes. These modules by default ensure the function of the binding update and the tunnel mechanisms i.e. IP in IP encapsulation for the nodes type setup (wired, wireless and hybrid wireless-wired network setup) of the base stations entity. The signaling messages of the protocol i.e. *RtSolPr, PrRtAdv, HI, H-Ack, F-BU, F-Back and F-NA* are also supported in the protocol entity and in the access points' configurations.

The movement scenario to support the Fast Mobile IP protocol is to allow the mobile node to move linearly at a constant speed between two access network coverage areas. A priority handover algorithm scheme is implemented within the ns wireless extension to allow the MN to switch from its PAR to the NAR to minimize unexpected interferences and maintain consistency during handover. This is done by setting the beacon priority of the NAR to be higher than the PAR, so that during the handover occurs in the overlapped region, the MN will receive the first beacon messages from the NAR. In the simulation setup, L2 handover is only based on the periodic beacons from the ARs.

5.1.2 SIP Implementation

The SIP simulation framework is based on the NIST SIP patch [40] with the updated ns-2.27 SIP patch [41]. The SIP protocol model supports the SIP protocol concept as described in [9].

The protocol model has the main SIP entities: *User Agent*, *Proxy Server*, *Redirect Server*, *Registrar* and *DNS*. The main signaling methods supported within the protocol entity are *REGISTER*, *INVITE*, *200-OK*, *ACK* and *BYE*. For the purpose of the simulation, Re-INVITE message is used under the same concept as the INVITE message to support direct communication during the handover. For the simulation, the SIP Registrar is modified and adapted to be co-located within the Home Agent of Fast Mobile IPv6. The SIP patch originally supported only wired network model and in the case of the simulation it is modified to support both hybrid wired-wireless and wireless network models. The cmu-trace files in ns-2 have been modified and recompiled to support the wireless model (Refer to Appendix C).

For the media transport within the SIP protocol NIST model, RTP [11] performs the functions immediately above IP and UDP in the protocol stack.

The media flow, for example, VoIP offers a number of features but at the expense of a significant amount of overheads as shown in Figure 5.1. In the case of IP, IPv4 header is 20 bytes and IPv6 header is 32 bytes. Therefore the total amount of overhead from RTP + UDP + IP is 40 bytes (IPv4) and 52 bytes (IPv6). As described in [13], the amount of overhead increases the bandwidth requirement with an overhead of 80% for IPv4 and 86% for IPv6 for voice application. RTP can be supported by RTCP to provide additional support for RTP.

Once the voice application has been established across an IP network, the codec, in this case, PCM coding scheme G.711, generates an RTP/UDP/IP header for each voice packet in the streams of the voice packets. In the case of the simulation, the overheads are added up to the voice over IP packet size.

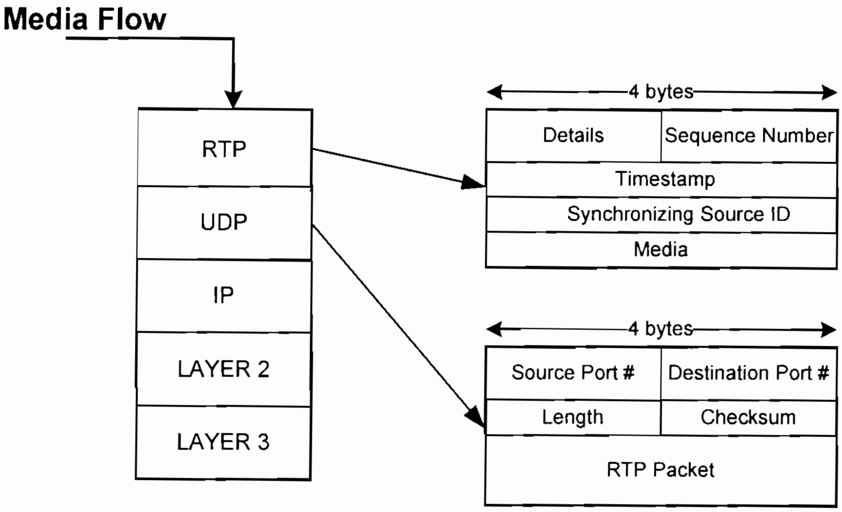

Figure 5.1: Media Flow over IP

5.2 802.11b WLAN in ns-2 Simulation Environment

The CMU Monarch extension [42] in ns-2 supports the wireless extensions that enable the simulations to have mobile nodes that can communicate via wireless interfaces configurations. The infrastructure mode to support 802.11b corresponds to the Orinoco WLAN parameters (i.e. 2.4MHz Lucent Orinoco WaveLAN DSSS radio interface with a range of 170m at 11 Mbps).The OTcL code is implemented as follows:

```
#Initialize the SharedMedia interface with parameters to make
Phy/WirelessPhy set CPThresh_10.0
Phy/WirelessPhy set CSThresh_5.011872e-12
```

```
Phy/WirelessPhy set RXThresh_1.02054e-10
Phy/WirelessPhy set Rb_11*1e6
Phy/WirelessPhy set Pt_0.031622777
Phy/WirelessPhy set freq_2.472e9
Phy/WirelessPhy set L_1.0

# Setting the bandwidth to 11 Mbps
Mac/802_11 set dataRate_11Mb
Mac/802_11 set basicRate_11Mb
```

Each cell consisting of an overlapped region has effective radius coverage of 40m within the 170m propagation range. An overlapping region of 20m is considered between the adjacent access routers. Each access point operates on different frequency bands using the NOAH routing protocol to support the wireless link model.

5.3 Routing Protocol NOAH

The routing protocol model NOAH [43] is an ns-2 extension provided through a patch to allow delivery of the packets in a wireless environment without Mobile IP. In contrast to other routing agents, i.e. DSDV, DSR, NOAH is a wireless routing agent that only supports direct communication between base stations/access points and the mobile nodes.

The nodes within the simulator are modelled to support the routing agent without holding constraints on switching capacity and message processing speed.

5.4 FMIPv6 + SIP Protocol Implementation Simulation Model

All the simulations are performed using the network topology as shown in network simulation topology Figure 5.2. The following simulation environment consists of a correspondent node (CN) streaming real-time traffic (i.e. streaming VoIP and video packets) with RTP over UDP to a mobile node (MN), Home Agent (HA) with co-located Registrar and SIP Redirect servers. In the case of a small scale simulation environment, it is not necessary to include a DNS. The SIP redirect server connected to the CN is given the url of deeya.crg.za and the SIP redirect server connected to the MN is given the url of lou.yahoo.uk.

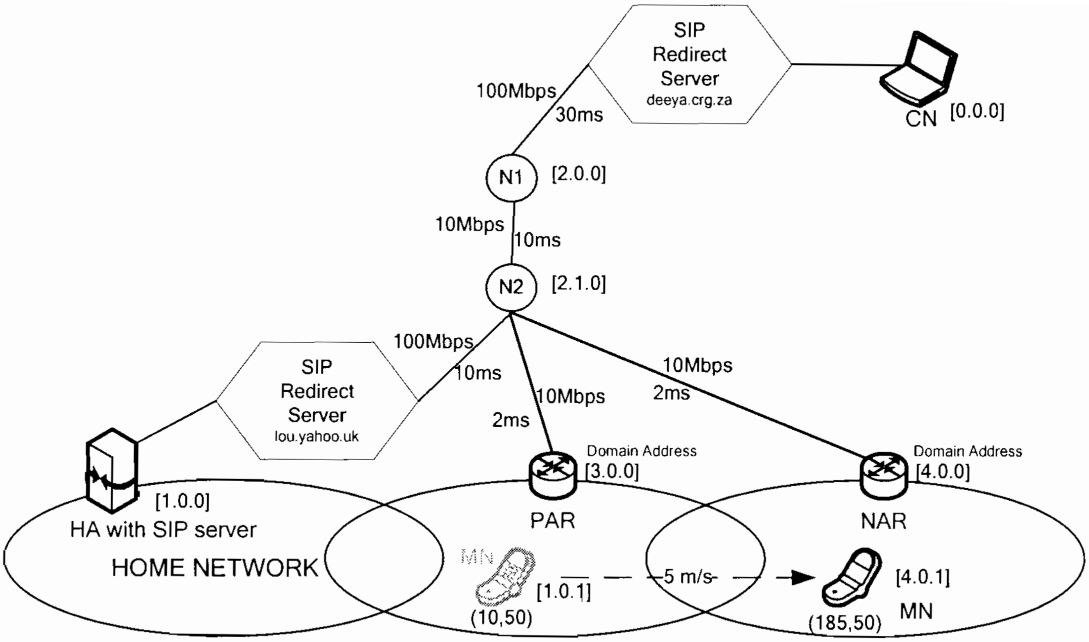

Figure 5.2: The simulation network topology

5.4.1 Traffic considered in simulation model

The CN is a CBR source, transmitting packets in a RTP over UDP medium. The MN acts a sink, by receiving the packets from the CN at a constant inter-arrival rate. Two types of packet are characterized: VoIP packet and real-time packet (streaming video application).

VoIP Packet:

VoIP connection can be modelled by a stream of packets with a fixed packet size and transmission rate. The CN produces a fixed length of packets of 200 bytes that is made up of a payload of 160 bytes and a header (RTP+UDP+IP) of 40 bytes. A typical PCM voice-coding scheme G.711 is emulated with a packet data rate of 64kbps corresponding to 20ms frames.

Streaming video Packet:

The CN produces a fixed length of packet size of 210 bytes to characterize streaming audio across the network simulator model that includes the 40 bytes header payload. The data rate for the transmission is set to 384Kbps. The high data rate is set in this case, assuming, that different real-time traffic can stream simultaneously.

In terms of bandwidth and link delay between the two intermediate wired nodes (N1, N2) and the access routers (PAR, NAR) are configured to 10Mbps and 10ms respectively. Between the access routers and the mobile node, these parameters are set to 1Mbps and 10ms. And between the HA and the CN, these parameters are set to 100Mbps and 10ms and 30ms link delay as shown in the simulator model.

5.4.2 Queuing Model

The link between the HA and the CN to the intermediate wired nodes use the RED (Random Early Detection) queue [44] to start detecting congestion before any buffer overflows for TCP traffic. All the other links that are within the SIP redirect servers and from the intermediate nodes to the ARs utilize the droptail (FIFO) queuing strategy in the network topology.

5.4.3 Simulation Scenarios

The simulation mobility scenario is based on the design proposed in chapter 3. The design supports the anticipated handover mechanisms of Fast Mobile IPv6 to allow the configuration of the care-of address before handover. In such a way, that SIP will then allow direct communication between the CN and the MN through establishing the re-INVITE message before the MN switches to the NAR.

5.4.4 Movement scenarios

The movement model for the simulation scenarios allows the MN to move linearly between the two access networks. The MN starts to move towards the NAR from PAR at 10s from simulation time, at a speed of 5m/s. The signaling messages are exchanged to establish new ongoing session and the real-time traffic through Constant Bit Rate (CBR) is generated over the SIP mobility framework. The MN reaches the NAR when the simulation time is at 45s.

5.5 Handover mechanisms in network topology

The following lists describe the mechanisms implemented to support the integrated scheme:

In our simulation model, ns-2 cannot support the L2 triggers and the L2 handover is therefore treated through the reception of periodic beacon messages. The beacons are captured in the registration agent (re_gagent) and after the set number of beacons has been received, so that the MN can switch and handoff to the new access router. The L2 handover time is fixed to 20ms in the network topology for fairness in movement detection. Address resolution time is also fixed due to the network limitation of ns-2.

Initially, the MN is attached to its HA and its care-of address will be associated with a domain address [0.0.1]. This corresponds to the hierarchical IP address of the HA. The CoA will associate with the hierarchical address since the simulation is based on a small-scale network, and thus the implementation of a DHCP server is not required. As the MN node moves out of the coverage of the PAR as illustrated in the network topology, it will configure a new CoA for every associated visiting network with IP addresses ranging from [3.0.1] to [4.0.1]. The new IP address is updated in the SIP registrar and the old IP address is deleted.

The experiments were categorized depending mainly on the protocols required based on the mobility policy table and type of application. For Fast Mobile IPv6 set up, L2 handover will only take place to the new access router only when the binding updates are sent through. In our proposed FMIPv6+SIP model, L2 handover will initiate the re-INVITE message in SIP after auto configuration of the IP address.

In the integrated model, the encapsulation/decapsulation scheme and tunneling in Mobile IP are deactivated, since SIP takes over the packet delivery mechanisms.

Finally, in all the simulations that are considered in the project, only one MN is considered to establish a constant single connection at a time. The results obtained from the simulation topology will be compared the performance of the existing protocols i.e. FMIPv6 and SIP protocols to analyze the effects of the integrated mobility model.

Chapter 6

6 Experimental Results and Analysis

This chapter evaluates the performance of the proposed integrated mobility scheme within ns-2 framework. The results of the performance of the proposed integrated scheme are presented and compared to the existing protocols' architectures i.e. FMIP and SIP. The main motivation for the optimization of the proposed scheme is to reduce the delays incurred by the existing protocols during handover. The mobility framework was tested for two types of applications: emulation of streaming video and VoIP application. Of importance to the simulation is the signaling that takes place when the mobile node is going through the handover process. The results show that the proposed integrated mobility framework performs better when compared to architectures employing FMIP or SIP only. The implication of the results is that using our integrated mobility framework, mobile nodes experience smooth handover which do not adversely affect real-time application currently running on these nodes.

The performance metrics investigated were as follows: handover latency, packet loss and throughput for two types of traffic (VoIP and streaming video). The impact of mobility on the applications running on the mobile node (MN) in terms of packet loss and handover latency is evaluated during the handover process. The speed of the node movement on the handover disruption time is also shown.

6.1 Handover Latency and Packet Loss

Handover latency and packet loss are the mobility performance metrics considered in the simulated wireless environment. The handover latency, in this case, is measured as the time interval between the last packet stream received through the PAR path and the first packet stream received through the NAR path. The packet loss is the amount of packets drop within the wireless links while streaming VoIP traffic during handover. As discussed in Chapter 3, the measurement of these parameters was performed between the two overlapped networks.

Trace files format were generated when running the OTcL codes and AWK (in ns-2) was used on the trace files to retrieve the required data.

Figure 6.1 illustrates the handover signaling disruption timeline of the protocols discussed in the experimental setup. The handover disruption times in FMIPv6 and the integrated scheme depends largely on the availability of the handover related information from lower layers to the IP layer. In pure SIP, the disruption time is higher because it has no mechanism to indicate eminent handover. The handover disruption time for FMIP and the integrated mobility framework does not differ much because both use the same handover detection mechanism to indicate eminent handover. The timeline only shows important messages exchanged between the MN and the AR, HA, and SIP agents.

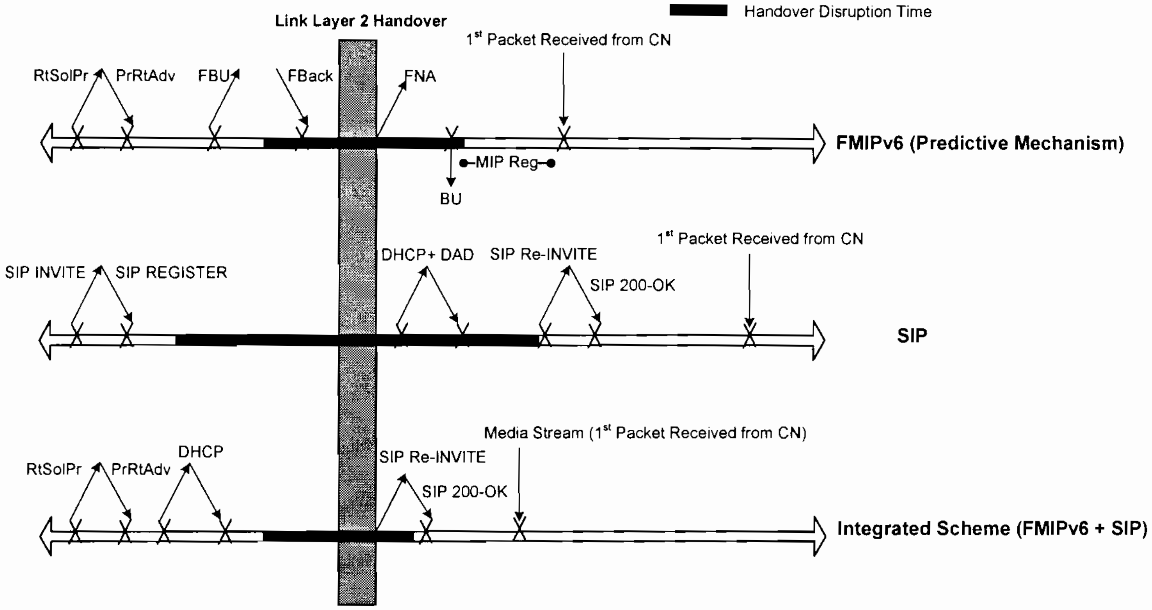

Figure 6.1: Handover Signaling Flow for FMIP, SIP & FMIP+SIP

Table 6.1 tabulates the different parameters measured in the handover disruption time interval. FMIP handover performance was measured over UDP application at a constant bit rate (CBR) while SIP and the integrated mobility model performance were measured with UDP application over RTP.

Table 6.1: Comparisons of Parameters of Different Protocols schemes.

	FMIPv6 (Predictive Mechanism)	SIP (Terminal Mobility Mechanism)	Integrated Mobility Model FMIPv6 + SIP
Packet Lost	16	13	6
Average Handover Latency (ms)	100.96	-	110.75
Average Throughput (Kbytes/s)	63.92	58.54	61.98

In terms of packet loss as shown in Table 6.1, the integrated model shows a 37% decrease in packet loss compared to the FMIPv6 predictive mechanism. The integrated model shows an improved performance because of SIP taking over the re-establishment of media flow after FMIP movement detection mechanism. The significant high packet loss in FMIPv6 as compared to the integrated scheme could be attributed to the time ambiguity problem in FMIP implementation in ns-2. FMIPv6 mechanism performs IP care-of address configuration and prepares for the tunnelling before the handover between the two ARs. During that time, the MN cannot receive any packets from the new router before the layer 2 handover takes place. The integrated model experiences a higher handover latency than FMIP even though it uses the same prediction mechanism as FMIP because it uses SIP immediately after layer 2 handover to re-establish data flow which takes longer to converge. Although the integrated scheme had higher handover latency, it experienced a 37% decrease in packet loss. This disparity can only be attributed to the implementation of FMIP in the simulation. A much-refined implementation would have resulted in lower packet loss than the integrated scheme. In terms of average throughput, the protocol mechanisms achieve to approximately the same system performance after handover.

6.2 Results of Handover Latency

Each data point on all graphs shown below corresponds to an average of 20 independent handover simulation events. The handover associated signaling latency is measured against the distance travelled by the MN with reference to the CN. The graphs do not include the delay incurred during DAD execution.

The following graphs illustrate the signaling delay comparison of the proposed integrated scheme to traditional schemes (FMIP and SIP).

6.2.1 Handover Delay Comparison between SIP and Integrated Mobility framework

Figure 6.2 shows the comparison of handover signaling latency between the integrated scheme and SIP.

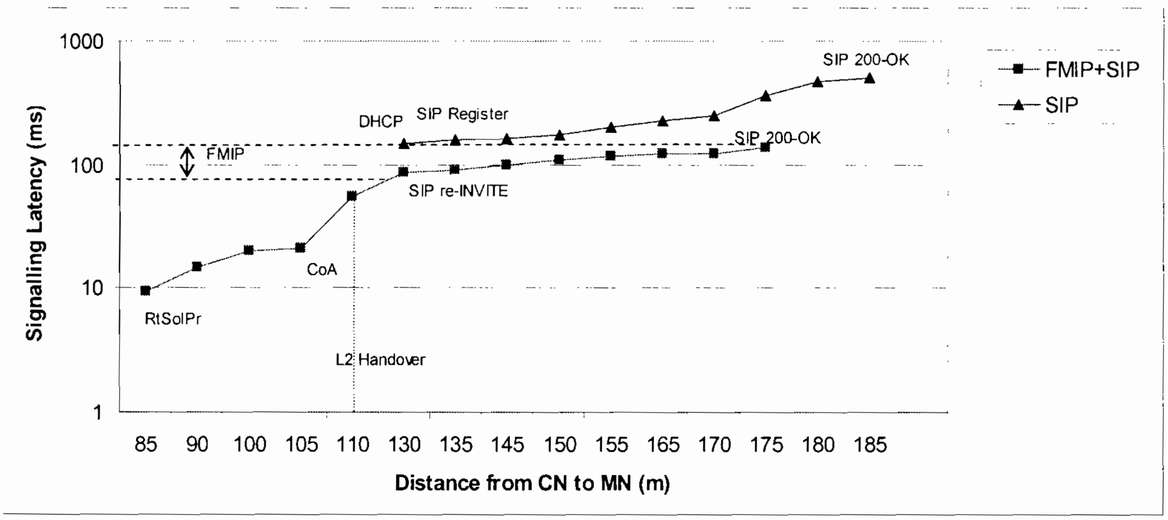

Figure 6.2: Handover Signaling Delay for VoIP in SIP & FMIP+SIP Mobility Framework

The integrated scheme (FMIP + SIP) shows a marked improvement in performance in terms of handover signaling latency compared to SIP. The integrated scheme shows a 42% reduction in handover delay over SIP. This improvement is attributed to the FMIP handover detection mechanism. In the integrated scheme FMIP is used for movement detection and SIP is used to re-establish the session between the MN and the CN. The CoA is configured before L2 handover enabling the MN to send a SIP re-invite message to the CN immediately after L2 handover is complete. In comparison, the SIP scheme has to wait until L2 handover is complete before it can get the CoA from the DHCP server and then send a SIP re-invite message to the CN. The handover signaling in the integrated schemes converges faster than in the SIP scheme owing to the absence of movement detection in the latter. Therefore, the 42% performance improvement in handover delay is attributed to FMIP as shown on Figure 6.2. In the figure, all the signaling after L2 handover are SIP signaling messages to re-establish media flow and all signaling before L2 handover are FMIP related message for movement detection. The 42% handover delay improvement also accounts for the low packet loss the MN experiences during handover as compared to SIP, refer to Table 6.1. The main signaling messages exchanged between the MN and CN are as labelled on the figure.

6.2.2 Handover Delay Comparison between FMIP and Integrated Mobility framework

In this section, FMIPv6 performance is compared to the proposed integrated scheme. FMIPv6 simulation model characterizes VoIP application at a constant bit rate (CBR) with UDP and the integrated scheme supports real-time communication over UDP with RTP. At 8.6 s, movement detection mechanism in FMIP is triggered resulting in the MN sending the router solicitation message (RtSolPr). This initiates the FMIP associated signaling to prepare for eminent L2 handover. The MN is then assigned the CoA before L2 handover. In the FMIP scheme, the MN continues with registration with the HA and the CN by sending BUs. Whereas in the integrated scheme, the MN waits for L2 handover to be completed before re-establishing the media flow through SIP. From the Figure 6.3, there is marginal difference in performance in terms of handover signaling delay because both schemes use the same movement detection mechanism. FMIP scheme converges faster than the integrated scheme, which has to go through SIP signaling to re-establish media flow. Therefore, FMIP re-establishes packet flow faster than in the integrated scheme even though we cannot account for the high number of packet loss in FMIP. As mentioned in section 6.1, this disparity can only be attributed to the implementation of FMIP in ns-2.

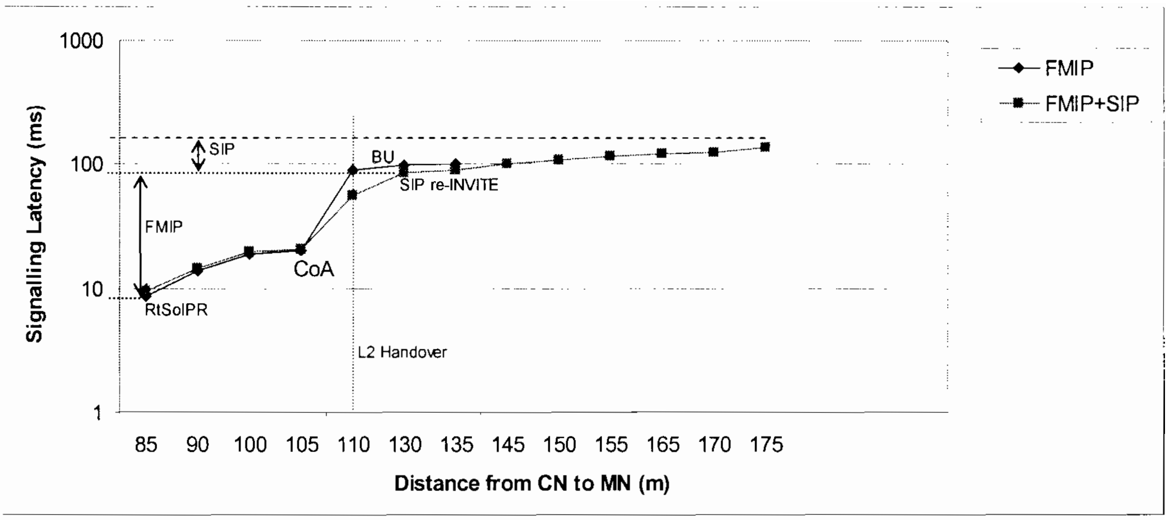

Figure 6.3: Handover Signaling Delay for VoIP in FMIP & FMIP+SIP Mobility Framework

6.3 Impact of moving speed

In this section, the impact of movement speed of the MN on handover disruption time is investigated. The MN's speed is varied from 2m/s up to 30 m/s. From Figure 6.4, both FMIP and the proposed integrated scheme (FMIP + SIP) are severely affected by the increase in speed although the proposed scheme shows marginal improvement in performance. The result can be attributed to the fact that both schemes employ same handoff detection mechanism of detecting the new access router well in advance of the actual handover.

With the moving speed of the MN increasing, the detection time is reduced and thus preparation for the anticipated handover process cannot be completed in time of the handoff. Though the disruption time is a function of the handoff detection mechanism used, handoff preparation time is protocol dependent and remains constant as long as the same protocol is used, in this case FMIP. This dependence explains the marginal difference in performance of the two schemes. Movement speed also affects packet loss due to handoff. The increase in MN's speed increases the possibility of packets being forwarded to the outdated path and thus increasing the probability of packet loss.

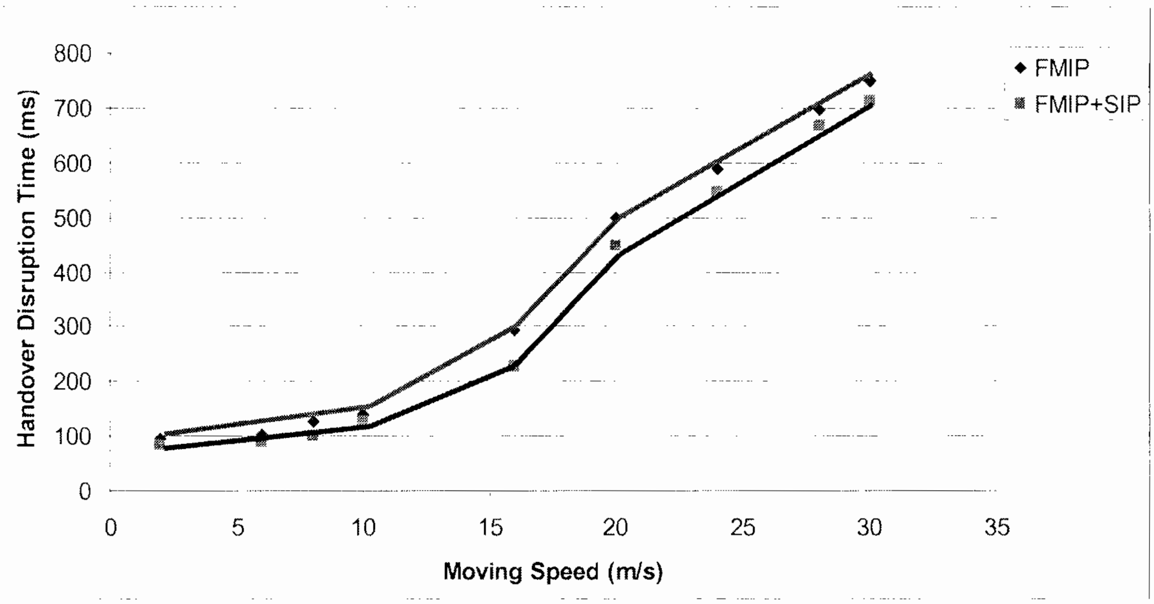

Figure 6.4: Moving Speed of Mobile node vs. Handover Latency

6.4 Impact of WLAN Range

Figure 6.5 shows the effect of the different WLAN ranges between the PAR and the NAR on the handover disruption time.

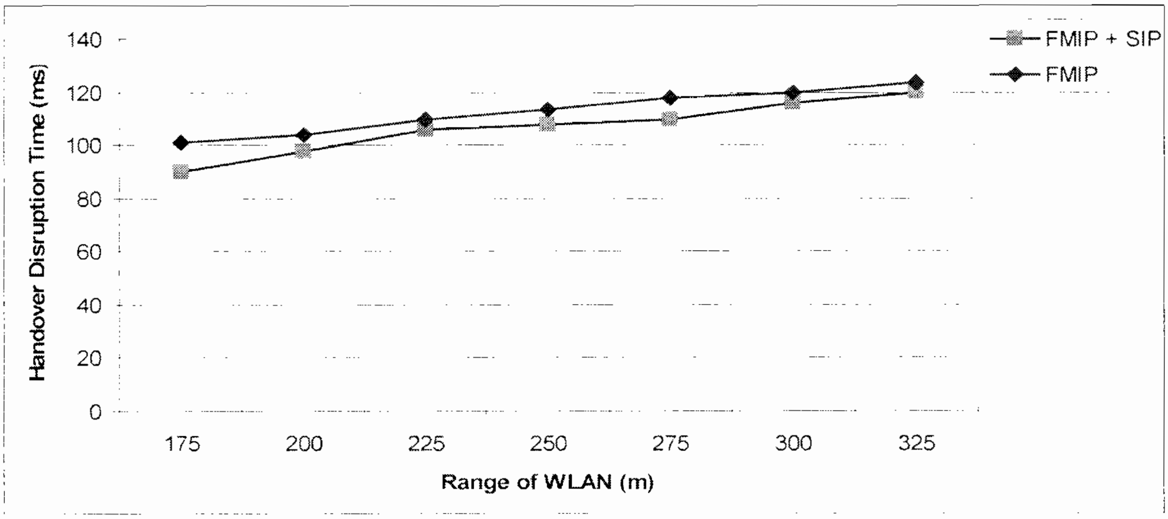

Figure 6.5: Range of WLAN vs. Handover Latency

Figure 6.5 shows that range of APs have no effect on the average handoff delay. This is because handoff only takes place in the overlap region between the two APs. Since the handoff detection mechanism employed in FMIP uses signal strength from the beacons received from APs in the vicinity of the MN, the effect of range between the two APs has minimal effect. This is because the MN node will only initiate handover if and when the beacon it receives from another AP other than the current is stronger. This takes place in the overlap region and, thus it is the extent of the overlap that affects the handover rather than the range of APs. From a micromobility perspective, the integrated proposed scheme and FMIPv6 relatively suffer from the same average handover delay as the WLAN AP range changes. The figure shows that range has marginal effect on handover latency in both schemes and therefore, no significant change in handover latency was observed. On a small scale network, WLAN configurations do not affect the overall latency delay for the protocol schemes.

6.5 Overall Comparison of the different protocol schemes

The following graph illustrates that handover signaling delays for the existing and proposed schemes. The proposed integrated mobility model shows an overall reduction in handover signaling latency compared to pure SIP schemes for any type real-time traffic investigated in the project. The difference between the integrated scheme and FMIP is marginal because both employ the same movement detection mechanism, but they differ in the method they use to re-establish the media flow. In pure SIP the CoA is only assigned to the MN after L2 handover which delays the re-establishment of the media flow further. However, in the integrated scheme, the CoA is assigned before L2 handover thereby enabling SIP to re-establish the media flow faster after handover. This is shown in Figure 6.6 where SIP signaling is delayed until well after L2 handover is complete.

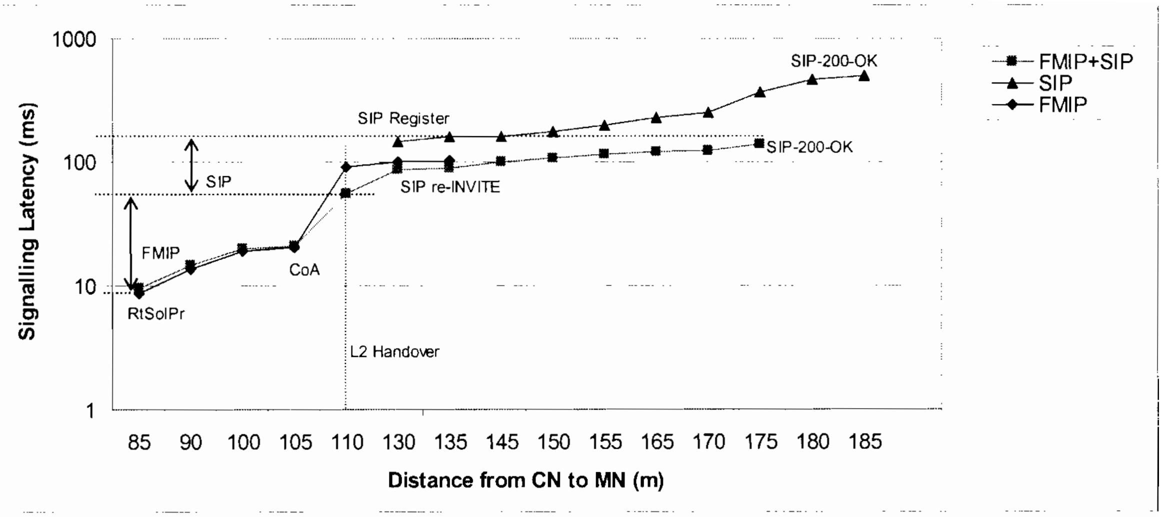

Figure 6.6: Handover Signaling Delay

The performance of real time applications was not adversely affected in the proposed integrated scheme as compared to pure SIP scheme due to shorter disruption time and lower packet loss. The integrated scheme offered smooth handover resulting in lower packet loss with minimal effect on the VoIP application.

Chapter 7

7 Conclusions

The main objective of this study is the investigation to minimize the handover latency and packet loss in overlapped network and a mobility architecture integrating two IP-based mobility protocols to improve the network performance. An integrated Fast Mobile IPv6 and SIP handover management mobility architecture is proposed that exploits both the complementary capabilities of each protocol and aims at reducing their functionality redundancies. The basic idea in the mobility framework is to support various mobility scenarios by making use of FMIPv6 and SIP procedures in a jointly optimized way to improve performance.

Real-time application requirements have been described to propose an efficient mobility framework that can minimize service disruptions of these applications during handover. Both Fast Mobile IPv6 and SIP have been described and compared on the basis of terminal mobility. The proactive approach for Fast Mobile IPv6 has been discussed in terms of its performance for fast and seamless handover. Therefore, a mobile device can anticipate a change of network as well as acquiring information on the new visiting network before the handover mechanism.

The SIP-based mobility approach has been considered and has offered several advantages over Mobile IP based solutions. The main drawbacks in SIP are the acquisition of care-of address through DHCP that may cause call disruptions during handoff and absence of mobility management for long-term TCP connections. In the study, IPv6 address configuration and removal of DAD have been considered to minimize the disruption time.

To reduce system redundancies and signaling loads, several functionalities of Fast Mobile IPv6 and SIP have been integrated to optimize the architecture. SIP Registrar has been co-located with the Mobile IP Home Agent to facilitate location management. Therefore any care-of address that is generated will be mapped to an SIP address as a co-located care-of address. Advanced handover management based on the type of applications has been set up to support various mobility scenarios. By using this approach, in the case of real-time traffic, the

Home Agent does not tunnel the packets on behalf of the mobile node but SIP takes over the packet delivery through RTP.

A software simulation framework, based on the mobility protocols described above has been implemented using network simulator NS-2. Simulation topology has been set up to analyze and compare the performance of FMIPv6, SIP and the integrated mobility framework. This analysis has been performed by characterizing two types of traffic VoIP and streaming video and the numerical analysis has been based on the metrics of handover latency and packet loss. Several modifications have been applied to support the protocols due to the limitations of the software simulator. The following conclusions can be drawn from the experiments covered in the thesis:

- Layer 2 handover is triggered based on periodic beacons received from each access router and based on the priority basis for handover. The link layer delays are introduced to illustrate the impact of the layer 2 trigger latency.
- The simulations metric are handover latency and packet loss. Comparison between the different protocols is considered and from the results, it can be concluded that the proposed mobility framework can reduce packet loss and handover latency. This has been achieved by compensating for the shortcomings of Fast Mobile IPv6 and SIP.
- Movement speed of a mobile node has also been investigated and has significant effects on FMIPv6 and the proposed scheme as the movement detection is done in advance. With higher moving speed, the probability of forwarded packets being dropped is higher as the handover latency will increase accordingly.
- The proposed scheme is more reliable as it minimizes the duplication of signaling messages especially in terms of sending out binding update, registration or re-INVITE messages. Thus, this minimizes the effects incurring extra overheads and signaling loads from the additional messages for the network.

From the simulation results, the proposed mobility architecture can offer powerful mobility support in terms of seamless handover to mobile devices for IP-based next generation networks. The basic idea of the mobility framework has been to jointly optimize the capabilities of Network layer protocol FMIPv6 and Application layer protocol SIP. Thus, the architecture offers flexibility to be adapted in future network developments to support real-time applications effectively under the “Always Best Connected Concept”.

Chapter 8

8 Recommendations

This study encompasses a broad spectrum of networking concepts by integrating two major IP mobility management architectures, Fast Mobile and SIP. A list of recommendations that can be associated to some future work are outlined below:

- The proposed architecture has been tested on a network simulator by using only one access technology i.e. 802.11b WLAN configuration. Other WLAN technologies could also have been used such as HIPERLAN to test the handover performance for the proposed mobility framework. Advanced mobility management can be configured on the architecture to support inter-domain handover, for example for a UMTS network to a WLAN network.
- The simulation framework only deals with terminal mobility. Future work can evaluate the advantages of the proposed architecture on different types of mobility schemes, for example a mobile device can at the same time switch from session mobility to terminal mobility.
- Quality of Service must be guaranteed from an end-to-end basis from the access technology up to the application layer. Quality of Service must provide adaptation support to real-time traffic during the re-establishment of these services. QoS context transfer information can be forwarded from the previous access router to the new access router to allow message exchanges and thus reinitiate QoS along the new data path immediately.
- Authentication, Authorization and Accounting (AAA) mechanisms have not been considered in the mobility framework. It is essential to develop an appropriate scheme to support such mechanisms to allow a mobile node to migrate from one access technology to another.
- Security is very important in a mobility framework. Future work must be devoted to security mechanisms for both Fast Mobile IP and SIP for the integrated architecture.

Therefore, network functions i.e. security and billing invoked during the handover mechanisms should be designed and adapted to real-time applications.

- A realistic evaluation framework must have been used to test the mobility framework further. Real-time applications can be streamed along the evaluation framework and their in-depth effects should have been investigated in terms of the output of the voice or the video quality.
- A larger number network with a number of mobile nodes could increase the scalability and feasibility of the project experiments.
- Cross-Layer interaction should also be considered to offer a unified platform for the mobility protocols and to synchronize their handover mechanisms with the link layer.

References

[1] Jin-Woo Jung et al. Performance Evaluation of Two Layered Mobility Management usingMobile IP and Session Initiation Protocol. Proceedings of IEEE GLOBECOM, December 2003.

[2] T.T Kwon et al. Mobility Management for VoIP service: Mobile IP vs SIP. IEEE Wireless Communications Magazine, Vol. 9, No. 5, October 2002, pages 66-75.

[3] N. Akhtar et al. SIP-based End System Mobility Solution for All-IP Infrastructures. Proceedings of IST Mobile and Wireless Communications Summit, June 2003.

[4] F. Chahbour, N. Nouali and K. Zeraoulia. Fast Handoff for Hierarchical Mobile SIP Networks. Transactions on Engineering, Computing and Technology, Vol. 5, April 2005.

[5] H. Schulzrinne and E. Wedlund. Mobility Support using SIP. Proceedings of second ACM International Workshop on Wireless Mobile Multimedia, August 1999.

[6] H. Schulzrinne and E. Wedlund. Application-Layer Mobility using SIP. Proceedings of ACM Mobile Computing and Communications Review, July 2000.

[7] N. Banerjee, K. Basu and S.K. Das. Hand-off Delay Analysis in SIP-based Mobility Management in Wireless Networks. Proceedings of IEEE Integrated Parallel and Distributed Processing Symposium, 2003.

[8] L. Peterson and B. Davie. Computer Networks – A Systems Approach. Second Edition: Morgan Kaufman Publishers, 2000, pages 489-494.

[9] J. Rosenberg et al. Session Initiation Protocol. IEFT, RFC 3261, June 2002.

[10] H. Schulzrinne and J. Rosenberg. A Comparison of SIP and H.323 for Internet Telephony. Proceedings of Network and Operating System Support for Digital Audio and Video (NOSSDAV), Cambridge, England, July 1998.

[11] H. Schulzrinne et al. RTP: A transport protocol for real-time application. IETF, RFC 1889, January 1996.

[12] J. Postel. User Datagram Protocol. IETF, RFC 768, August 1980.

[13] D.J Wright. Voice over Packet Networks. Wiley Series in Communications Technology, 2001.

[14] W.Wang, S.C Liew, Q. Pang and V.O.K Li. A Multiplex-Multicast Scheme that Improves System Capacity of Voice-over-IP on Wireless LAN by 100%. Proceedings of the IEEE Symposium on Computers and Communications, Egypt, 2004.

[15] C. Perkins (Editor). IP Mobility Support. IETF, RFC 2002, October 1996.

[16] C. Perkins. Mobile IP. IEEE Communications Magazine, Vol.35, No.5, August 1997, pages 84-99.

[17] C. Perkins and J. Tangirala. DHCP for Mobile Networking with TCP/IP. Proceedings

of IEEE International Symposium on Systems and Communications, June 1995.

[18] C. Perkins. IP Mobility Support for IPv4. IETF, RFC 3344, August 2002.

[19] D. Johnson, C.Perkins and J.Arkki. Mobility Support in IPv6. IETF, RFC 3775, June 2004.

[20] S. Thomson, T. Narten and T. Jinmei. IPv6 Stateless Address Autoconfiguration. IEFT Internet Draft, draft-ietf-ipv6-rfc2462bis-07.txt, December 2004.

[21] T. Narten, S. Thomson and T. Jinmei. Neighbor Discovery for IP Version 6 (IPv6). Internet Draft, draft-ietf-ipv6-rfc2461bis-02.txt, February 2005.

[22] R. Chakravorty et al. On Inter-network Handover Performance using Mobile IPv6. [Online] Available: http://www.cl.cam.ac.uk/users/rc277/handovers.pdf.

[23] J. Kempf et al. Supported Optimized Handover for IP Mobility - Requirements for Underlying Systems.IETF Internet Draft, draft-manyfolks-l2-mobilereq-01.txt, November 2000

[24] S.E Deering (Editor). ICMP Router Discovery Messages. IETF, RFC 1256, September 1991.

[25] S. Daniel Park. Minimized Duplicate Address Detection. IETF Internet Draft, draft-park-dna-mdad-00, September 2003.

[26] A. T. Campbell et al. Comparison of IP Micro-Mobility Protocols. IEEE Wireless Communications Magazine, Vol. 9, No. 1, February 2002.

[27] R. Koodli (Editor). Fast Handovers for Mobile IPv6. IETF Internet Draft, draft-ietf-mishop-fast-mipv6-03.txt, October 2004.

[28] M. Handley et al. SIP: Session Initiation Protocol. IEFT, RFC 2543, March 1999.

[29] H. Schulzrinne. Real Time Control Protocol. IETF, RFC 3550, April 1999.

[30] M. Handley and V. Jacobson. SDP: session description protocol. IETF, RFC 2327, April 1998.

[31] R. Sparks. SIP call control. Internet Engineering Task Force Work in progress, February 2001.

[32] A. Dutta et al. Fast handover Schemes for Application Layer Mobility Management. Proceedings of the 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2004), 2004.

[33] D. B. Johnson and C. E. Perkins. Route Optimization in Mobile IP. IETF Internet Draft, draft-ietf-mobileip-optim-05.txt (work in progress), February 2000.

[34] D.S. Nursimloo and H.A. Chan. Integrating Fast MobileIPv6 and SIP in 4G Network for Real-time Mobility. Proceedings of Malaysia International Conference on Communications and IEEE International Conference on Networks Kuala Lumpur, Malaysia, 16-18 November 2005, pages 917-922.

[35] Teodora Guenkova-Luy et al. End-to-End Quality of Service Coordination for Mobile Multimedia Applications. IEEE Communications Magazine, Vol. 22, No 5, June 2004.

[36] Q. Wang, M.A Abu-Rgheff and A. Akram. Design and Evaluation of an Integrated Mobile IP and SIP Framework for Advanced Handoff Management, EEE Communication society, 2004.

[37] The UCB/LBNL/VINT Network Simulator–ns (version2). [Online] Available:http://www.isi.edu/nsnam/ns. Accessed 2006.

[38] Robert Hsieh. FHMIPv6 Extension for ns-2. [Online] Available: http://mobqos.ee.unsw.edu.au/robert/. Accessed 2006.

[39] Robert Hsieh and Aruna Seneviratne. A Comparison of Mechanisms for Improving Mobile IP handoff latency for End-to-End TCP. Proceedings of MobiCom'03, California, September 14-19, 2003.

[40] NIST–SIP home page. [Online] Available: http://www-x.antd.nist.gov/proj/iptel. Accessed 2006.

[41] Rui Prior. NIST SIP patch for ns-2 ported to ns-2.27. [Online] Available: http://www.ncc.up.pt/~rprior/ns/index-en.html. Accessed 2006.

[42] The Monarch Project. [Online] Available: http://www.monarch.cs.rice.edu/cmu-ns.html. Accessed 2006.

[43] J. Widmer. Extensions to the *ns* Network Simulator NOAH. [Online] Available: http://www.icsi.berkeley.edu/~widmer/mnav/ns-extension/. Accessed 2006.

[44] B.Braden et al. Recommendations on Queue Management and Congestion Avoidance in the Internet. IEFT, RFC 2309, April 1998.

[45] Marc Greis. Tutorial for the Network Simulator ns. [Online] Available: http://www.isi.edu/nsnam/ns/ns-documentation.html. Accessed 2005.

[46] NS2-Manual. [Online] Available:http://www.isi.edu/nsnam/ns/doc/index.html. Accessed 2005.

Appendix A: NS-2 Implementation

NS is an object orientated simulator, written in C++, with an OTcL interpreted as a front-end [45]. The C++ and OTcL hierarchy are closely related to each other.

The following appendix describes the OTcL files written to simulate the network topology in ns-2. The link characteristics, namely bandwidth and propagation delays are measured in Megabits/s and milliseconds. Each simulation run lasts for 80 seconds of the MN's linear movement between the PAR and NAR to stabilize the results. Each data point for the data collection corresponds to an average of 20 independent events.

A.1: Wireless Configuration

This section of the code explains the physical wireless setup parameters in ns-2. The network components like Link Layer (LL), Interface Queue (IfQ) and MAC layer for the wireless channels (type of antenna, radio-propagation model) to transmit and receive signals are as follows:

```
=======================================================================
# Define options
# =======================================================================

set opt(chan)   Channel/WirelessChannel       ;# channel type
set opt(prop)   Propagation/TwoRayGround      ;# radio-propagation model
set opt(netif)  Phy/WirelessPhy               ;# network interface type
set opt(mac)    Mac/802_11                    ;# MAC type
set opt(ifq)    Queue/DropTail/PriQueue       ;# interface queue type
set opt(ll)     LL                            ;# link layer type
set opt(ant)    Antenna/OmniAntenna           ;# antenna model
set opt(ifqlen)         50                    ;# max packet in ifq
set opt(nn)             1                     ;# number of mobilenodes
set opt(adhocRouting)   NOAH                  ;# routing protocol
set opt(cp)     ""                            ;# cp file not used
set opt(sc)     ""                            ;# node movement file.
set opt(x)      670                           ;# x coordinate of grid
set opt(y)      670                           ;# y coordinate of grid
set opt(seed)   0.0                           ;# random seed
set opt(stop)   80                            ;# time to stop simulation

set opt(ftp1-start)      10.000000000

=======================================================================
```

A.2: WLAN Configuration

The following code addresses the 802.11b WLAN configurations in the simulation setup. This section is further discussed in Appendix B.

```
#set Antenna

Antenna/OmniAntenna set X_ 0
Antenna/OmniAntenna set Y_ 0
Antenna/OmniAntenna set Z_ 1.0
Antenna/OmniAntenna set Gt_ 1.0
Antenna/OmniAntenna set Gr_ 1.0

#Initialize the parameters#
# Initialize the SharedMedia interface with parameters to make
# it work like the 2.4GHz Lucent Orinoco WaveLAN DSSS radio interface (11
Mb/s and 170 m)
Phy/WirelessPhy set CPThresh_ 10.0
Phy/WirelessPhy set CSThresh_ 5.011872e-12
Phy/WirelessPhy set RXThresh_ 1.02054e-10
Phy/WirelessPhy set Rb_ 11*1e6
Phy/WirelessPhy set Pt_ 0.031622777
Phy/WirelessPhy set freq_ 2.472e9
Phy/WirelessPhy set L_ 1.0

# Setting the bandwidth to 11 Mbps
Mac/802_11 set dataRate_ 11Mb
Mac/802_11 set basicRate_ 11Mb
```

A.3: Topology Setup

This section defines the start of the simulation and the topology used. The mobile node moves within a topology of 670mX670m, i.e., a topography object with x and y coordinates of the boundary (x=670, y=670).

```
# create simulator instance
set ns_   [new Simulator]

$ns_ use-newtrace
set tracefd  [open normalsip-out.tr w]
set namtrace [open normalsip-out.nam w]
$ns_ trace-all $tracefd
$ns_ namtrace-all-wireless $namtrace $opt(x) $opt(y)

# Create topography object
set topo   [new Topography]

# define topology
$topo load_flatgrid $opt(x) $opt(y)
```

A.4: Hierarchical Addressing

The simulation setup uses hierarchical routing in order to route packets between the wireless and wired domains [45]. In the wireless topology, the packets are routed using NOAH routing protocol. Within the hierarchical topology structure, domains and clusters are defined on how the nodes are connected to each through the links and are associated to the care-of address of the mobile node.

```
# set up for hierarchical routing
$ns_ node-config -addressType hierarchical

AddrParams set domain_num_ 5           ;# number of domains
lappend cluster_num 1 1 2 1 1          ;# number of clusters in each domain
AddrParams set cluster_num_ $cluster_num
lappend eilastlevel 1 3 1 1 2 1            ;# number of nodes in each
cluster
AddrParams set nodes_num_ $eilastlevel ;# of each domain
```

A.5: Routers Configuration

This section explains the hierarchical address of each node. The mobility scenario consists of a Home Agent (HA), and two access routers that acts as Foreign Agents (FA). The HA and FA are access points similar to the base-station nodes as further discussed in ns-manual.

```
# create God
#   3 for PAR and NAR
create-god [expr $opt(nn) + 3]

set n0 [$ns_ node 0.0.0]
set CN [$ns_ node 1.0.0]
set n2 [$ns_ node 1.0.1]
set n3 [$ns_ node 1.0.2]
set n4 [$ns_ node 2.0.0]
set n5 [$ns_ node 2.1.0]

# NOAH nodes (wireless+wired) => HA, PAR, NAR
# MN is a special node (i.e. a NOAH node with wiredrouting turned off)

set chan_ [new Channel/WirelessChannel]
$ns_ node-config  -mobileIP ON \
                  -adhocRouting NOAH \
                  -llType LL \
                  -macType Mac/802_11 \
                  -ifqType Queue/DropTail/PriQueue \
                  -ifqLen 50 \
                  -antType Antenna/OmniAntenna \
                  -propType Propagation/TwoRayGround \
                  -phyType Phy/WirelessPhy \
                  -channel $chan_ \
                   -topoInstance $topo \
                  -wiredRouting ON \
```

```
                   -agentTrace ON \
                  -routerTrace OFF \
                  -macTrace ON
#PAR + HA
set HA [$ns_ node 3.1.0]
[$HA set regagent_] priority 3
set PAR [$ns_ node 3.0.0]
[$PAR set regagent_] priority 4
#NAR
set NAR [$ns_ node 4.0.0]
[$NAR set regagent_] priority 5

# create a mobilenode that would be moving between PAR and NAR.
# note address of MH indicates its in the same domain as HA.
$ns_ node-config -wiredRouting OFF

set MH [$ns_ node 3.0.1]
set node_(0) $MH
set HAaddress [AddrParams addr2id [$HA node-addr]]
[$MH set regagent_] set home_agent_ $HAaddress

# Position (fixed) for base-station nodes (HA & FA)

$HA set X_ 20.000000000000
$HA set Y_ 50.000000000000
$PAR set X_ 5.000000000000
$PAR set Y_ 50.000000000000
$PAR set Z_ 0.000000000000
$NAR set X_ 200.000000000000
$NAR set Y_ 50.000000000000
$NAR set Z_ 0.000000000000
$CN set X_ 185.000000000000
$CN set Y_ 70.000000000000

# movement of the MH
$MH set Z_ 0.000000000000
$MH set Y_ 50.000000000000
$MH set X_ 25.000000000000

# MH starts to move towards FA
$ns_ at 10.000000000000 "$MH setdest 185.000000000000 50.000000000000
5.000000000000"

#SETUP LINK# Connect the links
$ns_ duplex-link $n0 $n1 100Mb 10ms DropTail
$ns_ duplex-link $CN $n4 100Mb 10ms DropTail
$ns_ duplex-link $n2 $CN 100Mb 10ms DropTail
$ns_ duplex-link $n3 $CN 100Mb 10ms DropTail
$ns_ duplex-link $n4 $n5 100Mb 10ms DropTail
$ns_ duplex-link $n4 $HA 1Mb 2ms DropTail
$ns_ duplex-link $n5 $PAR 1Mb 2ms DropTail
$ns_ duplex-link $n5 $FA 1Mb 2ms DropTail

#Setting Up SIP servers
set serverid [$n0 id]
set serverid1 [$n3 id]

# agents
set sipA [new Agent/SIP deeya crg.ee.uct]
$sipA set packetSize_ 1000
$sipA set print_ 1   ;# if set 1 , display message flow on screen
$sipA set Server_ $serverid
$sipA set Lifetime_ 120
$sipA set Mode_ 1
$ns_ attach-agent $CN $sipA
```

```
set sipB [new Agent/SIP lou lou.yahoo.uk]
$sipB set packetSize_ 1000
$sipB set print_ 1
$sipB set Server_ $serverid
$sipB set Mode_ 1
$sipB set Lifetime_ 120
$ns_ attach-agent $MH $sipB

set sipC [new Agent/SIPRedirect lou.yahoo.uk]
$sipC set packetSize_ 1000
$sipC set print_ 1
$ns_ attach-agent $n0 $sipC
set sipD [new Agent/SIPRedirect crg.ee.uct]
$sipD set packetSize_ 1000
$sipD set print_ 0
$ns_ attach-agent $n3 $sipD


#Setup a RTP traffic over SIP connection
set st [new Application/SIPTraffic]
#set CBR [new Application/Traffic/CBR]
$st attach-agent $sipA
$st set type_ CBR
#Packet size 421 bytes => Real-Time Application
#Packet size VoIP 120 bytes => VoIP Application - G.711 codec
#Add RTP(8bytes)+UDP(12bytes)+IP Header(20bytes) => 40 bytes # Deeya
$st set packetSize_ 160
$st set rate_ 64Kb
$st set random_ false

#finish procedure
proc finish {} {
      global ns_ tf
      $ns_ flush-trace
      #close $tf
      puts "Running NS-simulation"
#     exec nam out_sip_test.nam &
      exit 0
      }

$ns_ at 10.0 "$sipA register $serverid"   ;# register its location redirect
server
$ns_ at 11.2 "$sipB register $serverid1"  ;# register its location with
redirect server
$ns_ at 15.0 "$st start lou lou.yahoo.uk"
$ns_ at 20.0 "$st send"
$ns_ at 140.0 "$st stop"
$ns_ at 160.0 "finish"


puts "Starting Simulation..."

$ns_ run
```

The Tcl scripts create output files in the form of trace files used as data sets. To extract the required data, AWK have been applied to the trace files. The AWK files used can be found in the accompanying cd-rom.

Appendix B: WLAN Configuration in NS-2

B.1: WLAN Channel Simulation

This chapter describes how to simulate an IEEE 802.11b based WLAN channel in ns-2. Within ns-2.27 framework, WirelessPhy module is used to simulate a wireless channel. WirelessPhy uses a propagation model found in ~*ns*/ns-allinone-2.27/ns-2.2/indep-utils to estimate the signal strength required of a frame at the receiver level. Mac/802_11 module is used to simulate the function of the MAC layer. The following thresholds from the receiver to determine whether a frame is corrupted are as follows:

CSThresh_: Carrier Sense Threshold. The WirelessPhy module from the receiver side uses this threshold to determine whether a frame can be detected. If the signal strength of a frame is weaker that the CS-Thresh value, it is discarded by WirelessPhy module.

RxThresh_: Radio Threshold. It is used by the WirelessPhy module to determine whether a frame has been correctly received. If the signal strength is stronger than the RxThresh_ value, the frame is tagged as correct. Otherwise, it will be discarded by Mac/802.11 if the frame is corrupted.

CPThresh_: Collision Threshold. This threshold applies when two frames are received simultaneously. If the signal strength ratio is larger than the CPThresh value, the frame with the stronger signal strength will be received and the other frame will be ignored. Otherwise, the frames will be discarded.

Environmental noise has not been discussed in the project since high frequency is being used by WLAN and environment noise is normally very small in small scale networks.

B.2: Schematic of Mobile Node

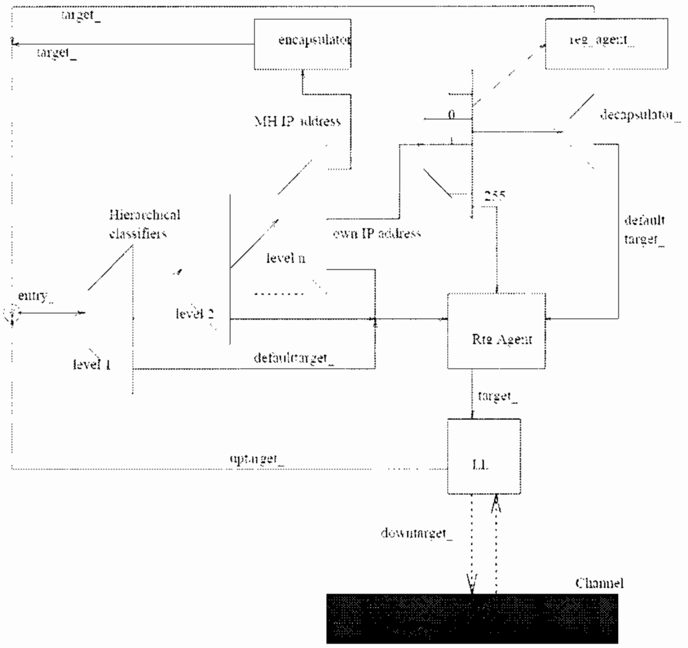

Figure B1: Schematic of Mobile Node [46]

Figure B1 illustrates the schematic of a mobile node configuration in ns-2. The mobile node has a regagent_ module that receives and responds to beacons by sending out solicitations to HA or FAs. The care-of address of the mobile nodes changes as it associates itself to the address of the base station in foreign domains. A priority handover mechanism of Fast Mobile IP is set up so that the mobile node attaches to the NAR as it reaches the visiting network.

Initially, the MN is in the range of the HA and receives all the packets directly from its CoA and in the case of hierarchical addressing, it will be the address of the HA. As the MN moves out of range from its HA, it will then associate itself to the foreign agent and the CoA changes from its HA to that of the FA. The HA sets up the encapsulator module and all packets are tunneled for the MN towards the FA. The packets are always routed towards the current CoA. The FA decapsulates the packets and forwards them to the mobile node [46]. Most traffic flow occurs in a hybrid simulation setup of wired and wireless connections.

Appendix C: Protocol Implementation in NS-2

This chapter captured the important issues on how to configure Fast Mobile IPv6 support compilation procedures and SIP support from the NIST module.

C.1: Fast Mobile IP with Fast Handover mechanism

This section describes the implementation of the Fast-handover protocol suit based on Hsieh's implementation. The following methods and procedures for Mobile IP extensions that has been modified to support the Fast-handover protocol suite for ns-2 configurations are described in: ns/mip.{h,cc}, mip-reg.{h,cc}, packet.h. Fasthandover.cc and fasthandover.h are the main code of the FMIP implementation. The following library tcl files are modified to support the protocol implementation: ns-agent.tcl, ns-default.tcl, ns-lib.tcl, ns-mip.tcl, ns-node.tcl, ns-packet.tcl. The protocol implementation consists of some IPv6 and IPv4 functionalities.

Throughout our integrated protocol implementation, in mip-reg.cc, the MAP_MODE and FAST_MAP_HANDOVER scheme was disabled to ensure the priority handoff for the nodes. The encapsulator/decapsulator module that enables tunnel mechanisms was disabled in the case of SIP media delivery.

The following procedures explain how fast handover simulations occur in the implemented module:

After the MN has moved to the new Access Network, the fast handover is triggered by the first encounter to an advertisement (adv) beacon.

1. The MN sends a RTSOLPR message to the PAR.
2. The PAR then sends HI to NAR.
3. The NAR then sends up the necessary encap/decap functionalities and returns the HI message with a HACK to the PAR.
4. With the receipt of the HACK, PAR sets up the necessary encap/decap functions.
5. A reg () process is performed that syntically performs the F-BU and F-Back function to PAR.

6. After receiving the MIPT_REG_REPLY, it then sends the F-NA message to PAR.
7. The receive verifier will be updated to accept any packets from NAR.

The simulation setup supports all types of node in ns, namely, wired and hybrid wired-wireless node.

C.2: Session Initiation Protocol Implementation

This section describes the SIP implementation based on NIST ns-2 SIP module [40]. The NIST SIP module originally supports the wired ns node. The following codes have been modified to support SIP functions in ns-2 shell: app.h, agent.h, agent.cc, packet.h. SIP messages signaling are implemented in sip.{h,cc}. The registrars and SIP redirect servers implementations can be found in sip_server.{h,cc}. A traffic generator module to support SIP media stream over RTP can be found in sip_tg.{h,cc}. The following library tcl files are modified to support the protocol implementation: ns-default.tcl and ns-packet.tcl.

To support the implemented SIP module for the wireless extensions, the cmu-trace.cc has been further modified in terms of receiving and sending SIP signaling messages and packets over a wireless link. Figure C1 is a flow chart of the inheritance of the cmu-trace module.

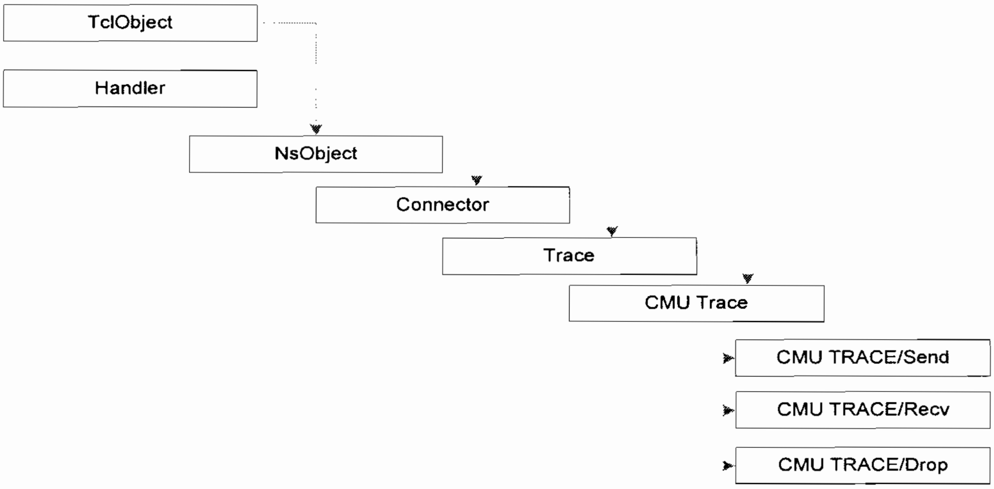

Figure C.1: Flow Inheritance of cmu-trace

Appendix D: Accompanying CD-ROM

The following information may be found on the CD-ROM that includes the simulation software and the related materials:

- Simulation Software

A complete installation of the ns-allinone-2.27 can be found in the "ns-2.27" directory. All source code and shell scripts that have been developed are included.

- Research Articles and Related Papers

Electronic copies of some research papers listed in the References section of this text can be found in the "Research Paper" directory.

- Thesis Document

The thesis can be found in a pdf format in the "Thesis directory".