A Cross-Layer Mobility Management Framework for Next-Generation Wireless Roaming

Mohamed Abdalla Abdelatif

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
August 2007

As the candidate's supervisor, I have approved this dissertation for submission.

Name: Professor H. Antony Chan

Signed: ______________________

Date: ________________________

Declaration

I hereby declare that: (1) the above thesis is my own unaided work, both in conception and execution, and that apart from the normal guidance of my supervisor, I have received no assistance apart from that stated below; (2) except as stated below, neither the substance or any part of the thesis has been submitted in the past, or is being, or is to be submitted for a degree in the University or any other University.

I am now presenting the thesis for examination for the Degree of (PhD) / (MSc in Electrical Engineering). I also grant the University free license to reproduce the above thesis in whole or in part, for the purpose of research.

Mohamed Abdalla ABDELATIF

Name

Date

To my family.

Abstract

In order to manage mobility in next-generation networks for 4G services protocols will have to contend with problems of link heterogeneity, providing seamless handoff, satisfying QoS, and enhancing user experience. This thesis proposes a mobility management framework that aims to provide a framework for advanced mobility algorithms that allows the challenges of next-generation roaming to be met. The framework features tools that gather context and content information, guarantee low-level QoS, provide security, and offer link and handoff management. The framework aims to be scalable and reliable for all-IP heterogeneous wireless networks whilst conforming to 4G service requirements. The framework is designed as a cross layer function that utilizes an IEEE802.21 Media Independent Handover stack. Simulation experiments were run to compare Session Initiation Protocol (SIP), Mobile IP (MIP) and the proposed solution to target the ideal IEEE802.21 user. The performance metrics registered during the simulation runs were the load incurred on the IEEE802.21 Media Independent Handover Function and the protocol stack reactivity of the protocols. The results obtained from the experiments confirm the proposed framework as the ideal IEEE802.21 user suited for Media Independent Handovers. The framework is seen to be an important step in guaranteeing advanced mobility management for 4G networks.

Table of Contents

List of Figures

Glossary

3GPP (3rd Generation Partnership Project): A predominately European standardization body in charge of defining standards for future and evolving networks.

4G: 4th generation; implies a futuristic concept and includes Beyond 3G (B3G).

4G networks: All-IP networks that provide an access, control and management, and service plane for network communications.

Access Router (AR): A front-end router that provides the first point of connection to the network, also used synonymously with Access Point (AP).

Authentication, Authorization, and Accounting (AAA): A security protocol that ensures the secure and accountable association of two network clients, used in this case for IP networks predominately.

All-IP: A 4G notion of global networks that are built on the IP protocol that provides a platform for Internet services.

Bit Error Rate (BER): The number of erroneous bits in a data link frame.

Committed Information Rate (CIR): Committed Information Rate is a value that indicates the service level agreement between a network and client for throughput during a session.

Correspondent Node (CN): The end-recipient or end-sender in a communication between itself and the MN (mobile node). Could either be a client or server with roaming capabilities.

Handoff: Refers to changing the Point of Attachment (PoA) of a UE from network to another to ensure service continuity. Handover is also synonymously used with handoff.

Home Agent (HA): A network element defined in the Mobile IP protocol operation as a router in a mobile node's home network that acts a proxy for the node when it roams away from the home network.

IEEE (Institute of Electrical and Electronic Engineering): An American-based institute whose 802 working groups are charged with defining standards for wireless networks.

GGSN (Gateway GPRS Support Node): The gateway router that bridges between a UMTS cellular network backbone and a packet data network, usually the Internet.

IMS (Internet Multimedia Subsystem): An architecture for provisioning, managing and controlling high-grade services for the Internet.

IP (Internet Protocol): The standard protocol for controlling and managing networks that interconnect to form the Internet by housing routing, addressing and other features in its suite.

ISO (International Standards Organization): A body that verifies standards, referenced here for defining the ISO protocol stack.

IWF (Interworking Function): Functional network entities that bridge across networks that differ technological by providing an anchor point for mobility for roaming clients.

Mobility management: A set of procedures that allow roaming devices to attach to a different network to access network services. It consists of handoff management and location management.

Mobility management protocol (MMP): A protocol that is dedicated to handling mobility management functions with provisions for handoff management and in some cases location management.

Mobile IP (MIP): Mobile IP is a part of the network layer IP suite and handles terminal mobility through address redirection.

Multihoming: A techniques used to assign multiple IP address to a single terminal for mobility.

Next-generation networks: See 4G networks.

Next-Generation Wireless Networks (NGWN): Advanced 4G networks that rely on wireless access radio technologies.

Packet Data Unit (PDU): A unit of information that is encapsulated and sent upwards or downwards in layer-by-layer fashion within the protocol stack in order to either relay information between the layers or to pack or unpack data payloads for transmission.

Policy Decision Function (PDF): A functional entity in IMS that manages and provisions QoS for the IMS traffic plane.

Point-of-Attachment (PoA): The physical data-link between a client and an access subnet that is a low-level association for channel communication.

QoS (Quality of Service): A term that encompasses the throughput and delay variables of a high-level transmission for assurances to the application and user.

OSI (Open Systems Interconnection): A basic standard reference model for defining and delineating the functionalities of protocols by modularizing networks into 5 main functional layers in a protocol stack.

Radio Access Networks (RAN): Usually last mile connectivity access networks that use radio interfaces for client communication.

Received Signal Strength (RSS): Represents the power of the received transmission at the client.

System Architecture Evolution (SAE): The latest, at time of writing, 3GPP definition of an upgraded evolved wireless terrestrial network.

Session Initiation Protocol (SIP): Session Initiation Protocol is an application layer mobility management protocol that can handle session and terminal mobility, and is primarily used for real-time UDP sessions such as VoIP.

Single-layer protocols: Protocols that reside in a single layer of the protocol stack that do not have access to cross layer information. Otherwise referred to mono-layer in this document.

SIP/MIP: Refers to an integrated hybrid solution that combines SIP and MIP for mobility management with several implementations available in literature.

SRNC: Serving Radio Network Controller. Refers to the UMTS network element that aggregates several Node_Bs. Here the Serving RNC governs a single Node_B domain.

UE (User Equipment): Refers to the user and the terminal client device used for network access. Used interchangeably with MT (mobile terminal), MH (mobile host) and MN (mobile node).

XLM (Cross Layer Manager): The main architectural feature of the proposed mobility solution. The term is synonymously used to refer to the functional entity and the whole proposed solution at the same time depending on the context of the sentence.

Chapter 1

Chapter 1 Introduction

Seamless network roaming and ubiquitous network coverage are the basis of global roaming which is one of key drivers of Next-Generation Wireless Networks (NGWN). Users, regardless of their location or access device, will expect access to a uniform set of recognizable services. This omnipresent access realized through ubiquitous network coverage for metropolitan areas at least.

The majority of communication services are seeing deployment on the Web, which is an economical platform for reaching users globally. As a result, all ubiquitous access networks will have a direct or indirect connection to the Internet as a backbone core network. The access networks, defined as IP-CAN (Internet Protocol-Connectivity Access Networks), natively transport IP packets for services deployed on the Internet Multimedia Subsystem (IMS). As users migrate across the IP-CANs, they expect handoff between the adjacent networks to be transparent. This seamless handoff will be the task of mobility management protocols that will have to ensure service continuity across heterogeneous IP-CANs. Transferring a running service session to a different IP-CAN network context is a major problem in NGWNs because the access networks are inherently different in their radio technologies.

The IP-CAN access networks are heterogeneous for a couple of reasons. One is that capital investment in existing infrastructure will mean that network owners will not want to forgo their technology and invest in other expensive ventures. The other reason is that different network requirements drive different network specifications. Practically, a single network cannot cater for all different user needs or provide all services. Issues such as cost, coverage, data rate and ease of deployment dictate differentiated technology specifications [1]. Because of these two reasons, existing networks will have to amalgamate to form NGWNs.

The access networks connected to the core Internet network have different configuration, association, QoS, and security mechanisms. Such incompatibility among networks introduces high complexity and overhead when a mobile node switches its Point of Attachment (PoA) during handoff. Although Interworking Functions (IWF) hardwired into the network infrastructure allow for context transfer amongst networks, it is more feasible for the IWFs to interwork only with the common IP-based core network in NGN and rarely with the adjacent networks. The IP plane, which is the platform for service delivery and application deployment, is the convergence point for all heterogeneous access networks. As a result, all stack layers above and including layer-3 are common across all access networks. This eases service, personal and profile mobility when moving across IP-CANs since there is a common network layer. However, this does nothing to terminal mobility since the data-link and physical layers are still different across IP-CANs.

For session mobility across IP-CANs, layer-3 handoff is necessary to maintain a locatable address for end-to-end IP communication. However, this does not guarantee that a terminal will be able to connect to a given IP-CAN through the network layer. Network-layer mobility management protocols cannot uniformly handoff a roaming terminal across different access links without adequate layer-2 support [2] for terminal mobility. Mobility between cells of the same radio technology is easily achievable by layer-2 handoff management mechanisms that are inbuilt in the data-link layer for a defined technology. The wider macro-mobility (mobility across different domains) requires vertical handoff [3], which should be able to handle context transfers between different technologies. What is needed is a mobility management protocol that simultaneously handles network layer macromobility and data-link layer terminal mobility. Mobile IP is the envisioned macro-mobility protocol for IP-CANs, and upgrades into MIPv6 when IPv4 migrates to IPv6.

MIPv6 tackles mobility through the redirection of a globally re-routable and globally identifiable address (IPv4 or IPv6 address). Since all mobile nodes expect to have an IPv6 address in NGWN, macro-mobility across IP-CANs is easier using MIPv6 by further allocating the roaming device a new IP address (care-of-address) when it migrates to a new subnet. The care-of-address (CoA), configured for its migrated subnet, is reachable via the node's old IP address (home address).

In addition to MIPv6 there are other mobility management processes residing at higher layers in the protocol stack such as the popular SIP. These mobility management protocols (MMP) -designed to handle terminal, personal, session or service mobility- require a transparent platform for network context switches. Since all the mobility types of personal, session and service mobility are inherently dependent on terminal mobility, heterogeneity in access links impairs mobility at all layers. To counter the problem of heterogeneity in access links, the IEEE802.21 working group drafted a service set to optimize handovers across heterogeneous networks. The IEEE802.21 Media Independent Handover Services (MIHS) [4] abstract links from the higher layers, allowing uniform and transparent handoffs to occur. However, even by using MIHS most mobility management protocols (MMP) lack the decision-making capabilities required in the advanced mobility scenarios prevalent in NGWN.

NGWN advanced mobility scenarios form when a mobile node is running stringent multimedia or data content in the overlapping areas of several heterogeneous access networks. Here, MMPs need to be aware of content such as the user preferences and the resource requirements of an application running on a service session. They also need to be aware of contextual information such as the capabilities and resources of connected or neighbouring networks. A subset of content/context-aware scenarios is the Always-Best-Connected (ABC) scenario in which multiple access technologies are available to the multimodal mobile node. A multimodal mobile node houses several network interfaces that connect to different access technologies. Because the available networks have different coverage, costs, and bandwidth specifications, an ABC mobility solution will run an application on a network whose specifications best match the application constraints [2]. Converging content to the best-effort network increases the handoff-decision complexity and further impairs a seamless handoff.

IEEE technologies all have host-based mobility management whilst 3GPP compliant networks have mobility agents in the core network. The trend continues in future networks for host-based IEEE802.21 and 3GPP's System Architecture Evolution (SAE) where a mobility management entity controls mobility for roaming user equipment (UE).

MMPs in NGWN require mechanisms for context and content awareness to satisfy 4G service requirements. In addition, the MMPs' movement detection and network registration times need to be low enough to ensure seamless handoff. In this thesis, we evaluate the envisioned MMPs for the future and ascertain their feasibility in an NGWN environment. We then introduce a proposed mobility management framework that can alleviate the problems faced by MMPs in NGWN. We compare the performance of our proposed scheme to that of other related MMPs. The results obtained then form the basis of our conclusions and recommendations.

1.1 Problem Statement

We contend that the envisioned MMPs for future NGWNs, namely pure SIP and MIP and other MMPs presented in Chapter 2, cannot handle mobility in advanced NGN scenarios. We explain the downfalls of current MMPs when put in a NGWN context in Chapter 2.

In general envisioned MMPs cannot handle mobility in the advanced NGWN scenarios due to one or more of the following:

- Handoff across different radio technologies is not supported.
- Seamless handoff is not realized due to the low reactivity of MMPs.
- Network registration with the IP core is not carried out
- The protocols do not include higher layer decision metrics for handoff and do not involve users in the handoff process.

We propose a Cross-Layer Manager Framework (XLM Framework) that resolves these problems currently facing traditional MMPs. The XLM Framework employs cross-layer design and IEEE802.21 MIHS to leverage mobility management in NGWN.

1.2 Methodology

> Meeting the challenges facing mobility management in next-generation all-IP wireless systems requires the use of new methodologies that allow for intelligent and adaptive solutions.

1.2.1 Cross Layer Design

The mobility management framework presented in Chapter 3 uses cross layer design for capturing data that is useful for handoff from across the protocol stack. Cross-layer design and optimizations are a popular solution for allowing the traditional ISO OSI protocol stack to cope with wireless environments [5]. The protocol stack does not perform well under noisy and volatile conditions because of the strict layering of protocols where Packet Data Units (PDU) traverse across two or more layers to reach its destination. Cross-layer design circumvents this by providing a fast path between any pair of layers for cross-stack communication. Various cross-layer design techniques for mobility management are presented in [6]. Mono-layer mobility management techniques cannot provide adequate support for roaming in advanced mobility scenarios. As a result, cross-layer techniques are becoming increasingly favourable for co-ordinating mobility-related information in lower and higher layers [6, 7].

1.2.2 IEEE802.21 Media Independent Handovers

The IEEE802.21 working group is in the process of standardizing a draft release released in June 2006. The document [4] details Media Independent Handovers that allow for generalized and uniform interdomain handoffs between heterogeneous access networks. The standard introduces a Media Independent Handover Function (MIHF) that exports Media Independent Handover Services (MIHS), generalized mobility-related messages that map to technology-specific primitives. Service Access Points (SAP) interface between the access technology modules and the MIHF. Each connected technology module exposes unique vendor primitives to SAP which export generic MIHS.

The MIHS categorize into Media Independent Event Services (MIES), Media Independent Command Services (MICS) and Media Independent Information Services (MIIS). MIH users poll the status of a connected link by subscribing to MIES that map to defined event messages. The MICS allow MIH users to control and configure the connected link for handoffs and subscription messages. The MIIS form from Type-Length-Value (TLV) messages that have low decoding complexity for fast relay of neighbour maps, QoS, and security information. MIIS include fields that are pending definition but are open to vendors and protocols outside the standard should the need arise.

1.2.3 Mobile IP

Mobile IPv6 (MIP) [8] is the de-facto standard for macromobility in all-IP networks. It allows a roaming node with an IPv6 stack to change its Point of Attachment (PoA) through location updates that re-route packets to the node's new subnet. A roaming node re- attaches to a new subnet by configuring its IP address to reflect its new location. A binding update (BU) sent to its home router, known as a home agent (HA), carries a newly configured address known as the Care-of-Address (CoA). The update re-routes packets to the node's new location including setting up a direct route to the correspondent node (CN) through route optimization.

Internet drafts have been drawn for Mobile IPv6 including Dual Stack Mobile IPv6 (DSMIPv6) [9], Proxy MIPv6 [10] and QoS Support in Mobile IPv6 [11]. PIMP could anchor host-based mobility to a network-controlled mobility entity for SAE. In this thesis, QoS supported MIPv6 guarantees resources along the handoff data path.

1.2.4 Session Initiation Protocol

Session Initiation Protocol (SIP) [12] has become popular session mobility protocol realized as the de-facto signalling standard for Internet Multimedia Subsystem (IMS) with its latest specification being RFC 3261. SIP sets up and tears down end-to-end sessions in IP systems by guaranteeing a QoS negotiated path for real-time applications. SIP can act as an agent for session and terminal mobility through INVITE or Re-INVITE messages that establish a connection to the handoff device or domain.

1.3 Contribution

This thesis aims to provide a framework that allows for seamless handoff between heterogeneous networks in all-IP networks including provisions for QoS, security, and handoff stability. Specifically the thesis contributes in the following:

1) To analyse the current MMPs that may handle mobility in NGWN and identify their shortcomings.

2) To propose a Cross-Layer Manager Framework (XLM Framework) that will satisfy the requirements of mobility in NGWN.

3) To provide proof-of-concept experiments to evaluate the feasibility of using cross-layer design for mobility management.

1.4 Thesis Outline

The organization of the subsequent chapters follows:

Chapter 2 describes current and proposed MMPs for next-generation mobility.

Chapter 3 presents the proposed XLM Framework with architectural features and mobility operation.

Chapter 4 shows the results of experiments conducted to evaluate the performance of XLM against that of SIP and MIP.

Chapter 5 draws out conclusions from the previous chapters and the thesis's main findings.

Chapter 6 puts forward recommendations about MMP in NGWN.

Chapter 2

Chapter 2 Current and Proposed Mobility Management Schemes

Mobility management protocols (MMP) developed for roaming nodes ensure session continuity once coverage is lost. Even though wired nodes can change their Point-of-Attachment (PoA) in a network, in this thesis we only consider wireless networks, which are widely popular for roaming. Mobility management consists of handoff management and location management.

Radio access technologies define handoff management in their suite such as the IEEE family, where hosts initiate handoff when the link-layer characteristics degrade below a threshold. This host-based mobility is true for all IEEE802 technologies. Conversely, 3GPP promotes network-controlled mobility, governed by management entities co-located with core routers.

Handoff management consists of movement detection and network registration. Movement detection defines how a mobile host realizes when it is moving, indicated mostly by channel deterioration or loss of coverage. Network registration allows the mobile host to re-attach itself to the network by associating itself with the domain.

Location management usually consists of resuming a call delivery session for sleeping or inactive nodes by updating a node's location through paging. Furthermore location management facilitates the delivery of a call to the node.

The architecture of next-generation hybrid wireless networks presents significant challenges to mobility management protocols. The envisioned architecture for these wireless systems as drawn by the Seamless Multimedia Services Over all IP-based Infrastructures (Evolute) group [13] is in Fig.1.

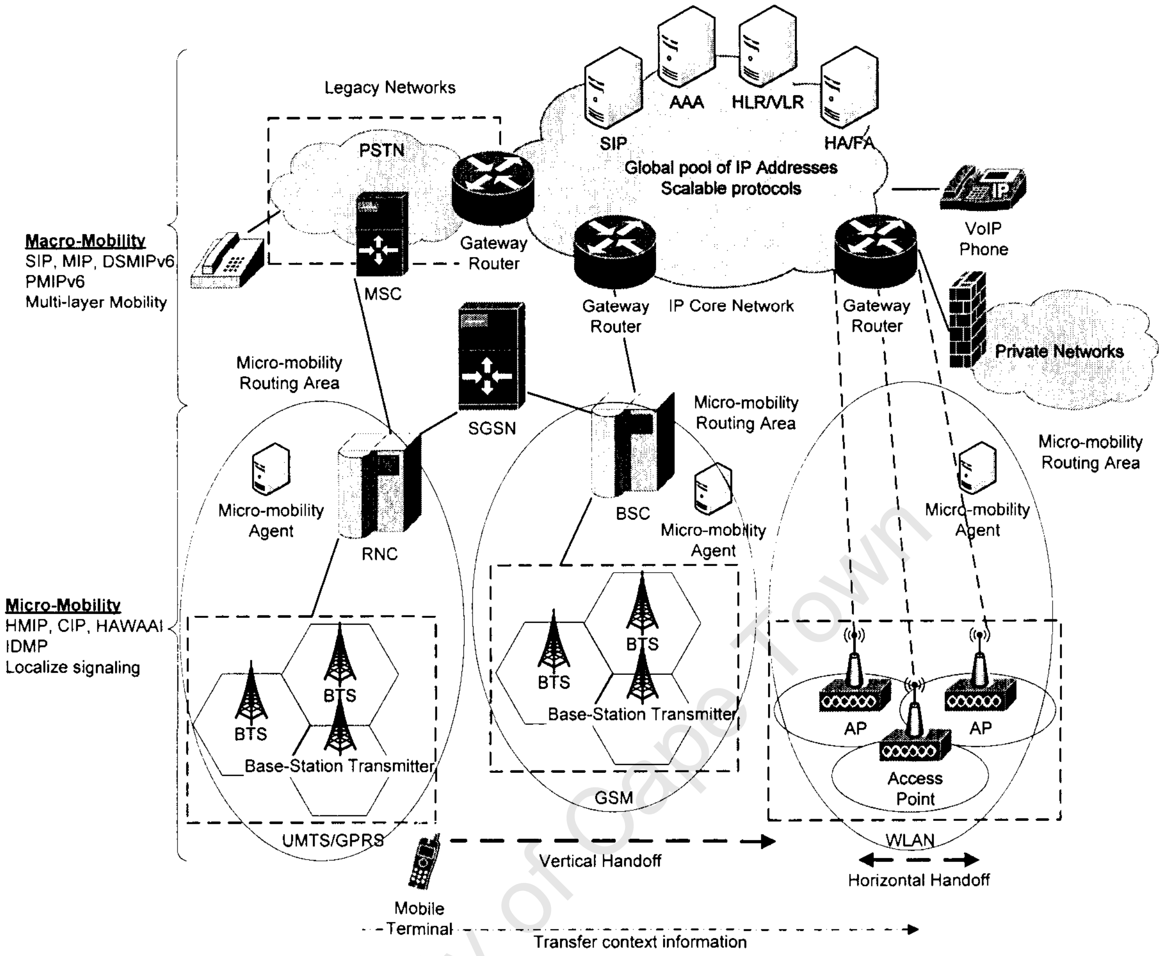

Figure 1. The architecture of B3G hybrid wireless systems as seen by Evolute.

Micromobility protocols handle mobility within an access technology whilst macromobility protocols such as Mobile IP and SIP manage interdomain vertical handoffs.

Solutions have been presented in academic literature and industry for mobility in next-generation networks. These are categorized into convergent network schemes and evolved mobility management protocols.

2.1 Convergent Mobility Schemes

Convergent or hybrid network schemes offer solutions to mobility by combining different network infrastructures for optimized inter-domain handoffs or integrating mobility protocols, in most cases MIP and SIP, for coordinated macro-mobility.

2.1.1 Mobility over Integration Architectures

In general integrated architectures implement interworking agents that allow interoperability between different Radio Access Networks (RAN). Integration architectures allow wider areas of coverage, differentiated services, and provide end-users with more access options.

Integration of different network types, especially 3G and WLAN networks, are generally categorized into tightly-coupled integration and loosely-coupled integration architectures. The tightly coupled integration scheme integrate WLAN and 3G together to a point where WLAN hotspots seem to be a connected RAN (Radio Access Network) to the 3G core backbone nodes. Even though QoS provisioning is easily manageable, tightly-coupled integration suffers from complexity in introducing costly changes in existing network infrastructures for mobile terminals and access gateways. The loosely-coupled integration architecture connects WLAN and 3G networks through a PDN (Packet Data Network) which is usually the Internet [14]. Loosely-coupled architectures reduce implementation costs in limiting the number of network changes. Terminal mobility is supported in both integration architectures where a WLAN hotspot is seen as a routing area within the UMTS domain. Generally, the type of architecture depends on how far along the backbone architecture a router goes to resolve the WLAN-connected node's location. If the location is resolved from the serving router then the architecture is tightly-coupled. Conversely, if the node's location is resolved by traversing till the gateway router then the architecture is loosely-coupled. The 802.11 network architecture will have to emulate several UMTS functionalities for compatibility. However, each technology retains their respective terminal mobility management features and inter-domain updates occur between the tier nodes to refresh the location database. For session mobility, session handoffs can be transferred to the target technology and packets can be routed to the new UMTS or WLAN hotspot assuming an all-IP scheme [14].

In [15] an integrated 4G/WLAN 3-tier architecture is proposed to satisfy seamless macro-mobility by using a generalized 3-level mobility architecture based on MIP which is similar to UMTS mobility tiers. In addition handoff and route latency are reduced by sending location updates (binding updates) to an inter-domain agent. The architecture overcomes MIP limitations by localizing mobility into domains. When a MT moves between domains, location updates are sent to a domain gateway with a General Mobility Agent handling movement between 3G and WLAN networks. This solution reduces packet loss through agent buffering and allows for AAA (Authentication, Authorization and Accountability) security measures. However this solution suffers from the need to install additional entities to current 3G and WLAN architectures which is difficult for wide-scale implementation.

Another integrated architecture is also presented in [16] where UMTS cells and WLAN hotspots connect through border routers that allow for call re-establishment for inter-domain handoff. The solution entails a multimodal terminal device with two air interfaces, an 802.11 and a UMTS network interface. The architecture allows for image resource reservation with a RSVP implementation for WLAN and UMTS using its native PDP context transfer. Signaling between WLAN's access router and 3G's SGSN and SRNC allows for inter-domain mobility.

The integration schemes rely on the native mobility management features of access technologies which are invariably in-built layer-2 features. For inter-domain handoff, network equipment is used for re-routing and maintaining location records. For a comprehensive survey on integration schemes refer to [14].

These solutions are generally not scalable to encompass all next-generation wireless systems since they only aim to integrate certain network types. Furthermore these architectures implement costly changes in the network infrastructure and do not take advantage of layer-3 macromobility for global roaming.

2.1.2 Hybrid SIP/MIP Solutions

On the other hand SIP and MIP protocols can be used for inter-domain handoff of sessions without costly changes to native network infrastructure. MIP and SIP handoff sessions at a point above the data-link layer bypassing the access technologies' native mobility management schemes.

Two approaches to mobility are introduced and evaluated in [17] with a pure SIP approach and a hybrid SIP/MIP mobility solution. MIP and SIP are used for inter-domain mobility whereas HMIP or CIP are used for micromobility within administrative domains. In the pure SIP approach Network Address Translation (NAT) functionality is necessary to interwork SIP sessions with IP since SIP's address format is different from that of IP. The SIP approach uses Dynamic Host Configuration Protocol (DHCP) to map IP addresses to handoff sessions. The hybrid SIP/MIP approach divides traffic, and assigns SIP for real-time traffic over UDP and MIP for non-real-time traffic. SIP was chosen for real-time traffic to avoid the latency incurred for IP encapsulation and tunneling brought about by MIP. For TCP connections, SIP required multi-layer coordination between TCP and IP to record on-going TCP connections and for the end-to-end encapsulation and decapsulation of the packets at the MH. The paper concluded that SIP performs well for real-time sessions but poorly for non-real-time sessions. MIP suffers from tri-angular routing (assuming route optimizations are not used) and IP tunneling.

Another hybrid solution is presented in [18] where multi-layer design is used to exploit the synergy between SIP and MIP's network registration. MIP's home agent (HA) and SIP's Registrar server are co-located so that only a single Binding Update and Re-INVITE message update is required. This, in addition to disabling Duplicate Address Detection (DAD), reduces the signaling overhead which results in low latency handoff.

These solutions assume a uniform platform for handoff between different networks or intradomain movement. Note that these studies do not take into account application requirements in order to target the best-suited access network platform for the handoff service sessions. Some studies have been proposed to take into account application requirements and even user preferences; these are treated in the following.

2.1.3 Policy Decisions for Advanced Handoff

A context-aware mobility management architecture is proposed in [19] which introduces middleware for multi-layer functionality. The scheme requires applications to be content-aware to take advantage of the benefits that different networks offer. The solution showcases Bluetooth and WLAN as the primary access mechanisms though the solution can be generalized to encompass more technologies. The solution uses NCSOCKS (Nomadic Computing Sockets) as a middleware link-abstraction mechanism as it supports object-orientated end-to-end communication regardless of underlying technology. The middleware passes mobility-related information such as link events to the upper layers. The primary defect in this solution is that application-level services are required to be mobility-aware. A solution that de-couples mobility from the applications is needed. In addition the newly-introduced IEEE802.21 standard may invalidate or replicate many of the functions that exist in the mechanism and middleware layers in the proposed solution.

In [20] fuzzy logic is the basis for running algorithms that tradeoff the demands and criteria of users, applications and networks to target an optimal network. A 3-stage decision strategy was proposed that uses metrics obtained from user demands (costs, battery life, expectations) and network QoS factors (data rate, SNR) to filter through candidate networks.

Another optimal network targeting scheme was introduced in [21] which weighs normalized decision parameters and aggregates them for a certain network. The cost function of different networks is calculated and compared to select the optimal network. The decision parameters are network metrics based on bandwidth, cost and power consumption. The cost function takes into account user preferences which are entered into a policy table. A bias is entered into the decision strategy for highly stable networks that consistently have the lowest cost function. This is to prevent destructive toggling between networks due to transient changes which may ostensibly rank a candidate network higher than the other.

IEFT Seamoby (Seamless Mobility) group introduced two protocols that can be used to transfer context information amongst neighbouring access routers. Candidate Access Router Discovery (CARD) [22] allows access routers (AR) to populate a list of neighbouring routers' MAC (or link) addresses, IP addresses and capabilities (security, resources, and available access technologies) to reduce the handoff latency in case a mobile host moves to a neighbouring network. Context Transfer Protocol (CTP) allows authentication, header compression and QoS information to be transferred between ARs before or after handoff of a host takes place. It aims to allow a handoff session to continue with minimal disruption through exchanging context messages.

For global roaming a generic mobility model needs to be defined for all network types and since these solutions apply to only specific network pairs, further research is needed the interoperation of the plethora of access networks. Furthermore, minimal changes to network infrastructure is desirable for a cost-effective implementation and for scalability.

2.2 Evolved Mobility Management Protocols

Enhancements have been added to existing protocols to allow them to cope with new challenges. These include micromobility protocols, MIPv6 extensions, and using IEEE802.21 for Fast Mobile IPv6 and SIP.

2.2.1 Micromobility Protocols

For highly-mobile hosts it is useful to reduce location updates to micro-domains to reduce the overall signaling in a network. Micromobility protocols localize host movement within a routing area domain with intradomain movement being transparent to external domains. Updates are only sent to the home agent (HA) when the host moves outside a domain to update its routing list. One of the more popular IP-based micromobility protocols is Hierarchical Mobile IP (HMIP) which introduces multiple hierarchies of micromobility domains for reduced location update delays. A similar micromobility protocol is Mobile IP Regional Registration (MIP-RR) which, with Intradomain Mobility Management Protocol (IDMP), is classified as tunnel-based schemes [23]. Routing based schemes include Cellular IP (CIP) and HAWAII and are comprehensively surveyed in [23].

Terminal Independent Mobility for IP (TIMIP) [24] decouples mobility from a host's stack by allowing any device connected to roam using layer-2 mechanisms to associate itself with an Access Point (AP). Routers along the network path to the Internet core maintain route lists with an entry for the mobile terminal for routing and location management. The protocol allows legacy terminals that do not have Mobile IP functionality to roam and operate in IP subnets that are divided in hierarchical domains for efficient micromobility.

These micromobility protocols indeed reduce handoff delay and are very useful for managing mobility within a limited administrative area. For macromobility the protocols fall onto Mobile IP to handle inter-domain movement.

2.2.2 Extensions to Mobile IP

Mobile IP (MIP) allows for macromobility in all-IP networks by re-assigning mobile nodes a globally routable address in any network connected to the IP core. MIPv6 is part of the upgraded protocol suite of IPv6 and allows for route optimization which mitigates triangular routing by ensuring the Correspondent Node (CN) gets a Binding Update (BU) copy.

A MIPv6 solution in NGWN would see multi-homed devices for multimodal capabilities, with SCTP (Stream Control Transport Protocol) a favoured multihoming control mechanism. However the reactive MIPv6 is still limited and inadequate to handle advanced mobility scenarios present in 4G networks [7] and it suffers from slow movement detection [7]. However many enhancements have been augmented to the base MIP which leads to evolved mobility management protocols. These include Dual Stack MIPv6, Proxy MIPv6, a QoS Framework for

Mobile IPv6, Network Mobility (NEMO) and a paging extension.

Dual Stack MIPv6 (DSMIPv6) in [25] allows for the concurrent deployment of Mobile IPv4 and Mobile IPv6 protocols in a single stack. Both versions are recognized by network equipment with a DSMIPv6 stack including Network Address Translation (NAT) edge routers [25].

Proxy MIPv6 (PMIPv6), presented in [26], enables MIPv6 to perform mobility in a network-based manner where the host is not involved in any direct mobility signaling. This is to ensure MIPv6 which is traditionally host-based can be deployed in a network that supports network-controlled mobility. However unlike TIMIP/sMIP, the host has to have an IPv6 stack for address configuration. A proxy foreign agent is introduced to manage roaming hosts in foreign networks to make movement transparent. PMIPv6 allows for session continuity by establishing a tunnel between the HA and the proxy agent for packet forwarding.

Network Mobility (NEMO) allows mobile routers, which are part of a mobile network, to change their PoA to the core Internet dynamically without disrupting the communication channel of the nodes attached to the mobile network [27].

In [11] QoS support is allowed in mobile IPv6 nodes through extending the protocol signaling by adding a QoS option that allows routers along the MH-CN data path to map the host requirements onto their domain. This guarantees mobile hosts a best-effort connection to the CN when changing its PoA. The QoS requirements are injected in a QoS object which is carried as either a Hop-By-Hop or Destination option along the path that Binding Updates (BU) or Binding Acknowledgements (BUack) take. It is assumed that the BU or BUack path is the same path that the end-to-end data takes between the MH and the CN.0

In [28] a paging extension for IP is introduced for the location management of roaming hosts. The extension allows hosts to enter an idle mode where the host is exempted from sending periodic location updates that are power exhaustive as long as it is within a paging area, where the host is known to be roaming around. Only when the host leaves a paging area, an area which delimits idle modes, does the host send a location update indicating movement into a new paging area.

Surrogate MIP (sMIP) is used for macromobility by implementing surrogate Home Agents (sHA) and surrogate Foreign Agents (sFA) that act as proxies for Mobile IP signaling for terminals that lack Mobile IP stacks [29].

Handoff Protocol for Overlay Networks (HOPOVER) shown in [22] is based on Mobile IP where a roaming host (MH) registers its information with neighbouring APs that cache routes, QoS, and transit packets relating to the MH. This is done with APs that might serve the MH in future, so that when a MH moves to a new AP, resources for that MH are already reserved. The old serving APs are updated with the MH's movement in order to release the resources held for that MH and forward any packets to the new serving AP. HOPOVER however wastes resources and calls for wide-spread adoption among different networks.

Omnicon surveyed in [22] allows handoff between GPRS and WLAN networks by introducing a virtual interface, tcptun, which throttles link advertisements to the Mobile IP module. tcptun registers with either the WLAN as the outgoing interface or tunnels packets to the GPRS FA for 3G connectivity. The choice of networks depends on the signal strength of both technologies with thresholds set for handoff [22]. However this is insufficient for advanced mobility scenario roaming prevalent in future systems.

2.2.3 High Layer Protocols

Mobile Stream Control Transport Protocol (mSCTP) is a transport layer protocol that allows seamless handoff for a roaming mobile host by selecting from a pool of IP addresses. With SCTP, a device can be multihomed, so a device can maintain several valid IP addresses each with its associated transport layer connection. With mSCTP, the ADDIP extension is used to add or delete an IP address or change the primary IP address. For seamless handoff a multihomed host can configure an IP address, send it to the CN and then change it to become primary before the previous primary address becomes invalidated [30].

Session Initiation Protocol (SIP) as introduced in the previous chapter is used for session or terminal mobility. Because SIP is an application layer protocol it may incur handoff delays due to the intensive text processing required at nodes during handoff [22]. However SIP's downfalls can be mitigated by using IEEE802.21 MIHS for handoff prediction to achieve seamless handoff as shown in the next section.

2.2.4 IEEE802.21 Centric Approaches

It has been shown in [31] that IEEE802.21 MIHS can provide seamless handoff in conjunction with SIP for multimedia content. Close to zero packet loss was achieved by pre-empting a connection break. This is achieved using MIES for movement detection, and then subsequent authentication-association messages are exchanged whilst maintaining the current connection.

In [32] Fast Mobile IPv6 (FMIPv6) [32] is used in combination with IEEE802.21 to optimize handoffs by reducing signaling overhead by eliminating FMIPv6's RtSolPr/PrRtAdv, radio access discovery, and candidate AR discovery. A new Information Element is defined in a MIH_Get_Information to transfer L2 and L3 information for pre-network registration.

Both IEEE802.21-assisted solutions assume a mobility management entity (MME) for network selection through running decision algorithms. However they do not define the MME [32].

2.3 Summary and Conclusions

Currently there is no holistic solution to manage mobility for roaming nodes in hybrid access networks. All the solutions reviewed lack a critical aspect, rendering it inadequate to handle vertical macromobility for real-time multimedia sessions as seen in Table 1.

A solution needs to be found that can be widely adopted with minimal changes introduced to existing network infrastructure to allow global cost-effective adoption. The mobility management solution has to be scalable across different radio access technologies with different network operators and administrators. The handoff decision metrics that provide criteria for targeting networks are more complex in next-generation networks than in traditional systems. Thus simple handoff parameters such as signal strength are not sufficient for triggering handoff. This requires advanced decision algorithms that factor in application constraints, user preferences, QoS parameters, security configurations as well as traditional triggers such as SNR, RSS and BER. The solution should support generalized mobility, with uniform authentication, configuration and billing for global roaming. Lossless handoff is a key service requirement for mobility management protocols (MMP) as user's Quality of Experience (QoS) and the bandwidth-rich applications will not tolerate packet loss and handoff latency.

The following requirements, presented in [22], are the key issues that need to be resolved before a holistic mobility management scheme is realized for Beyond 3G (B3G) systems:

1) Multimodality

The ability to traverse different radio access technologies without losing coverage is a challenge for next-generation terminals. A uniform and transparent mechanism is need to allow devices to switch application flows to different outgoing network links that have different specifications such as frequency range, modulation, data rates etc..

2) Seamless Handoff

Lossless handoff that provides a transparent platform for users and applications is paramount for fulfilling 4G service requirements. Packet loss and handoff latency should reduce to near-zero for smooth handoff amongst networks. Handoff should be seamless for vertical and horizontal micromobility or macromobility.

3) Advanced Decision Algorithms

Advanced network selection algorithms must exploit the differences in the varying services offered by access networks. Algorithms should target optimal networks while taking into account user preferences, application constraints, network coverage, security policies, and billing. In addition, the run-time of these algorithms should be minimal as not to incur handoff delays that disrupt session continuity. Intelligent and adaptive schemes should cope with the volatile nature of wireless systems to mitigate adverse effects such as hysteresis, line-of-sight (LOS) shadowing, and the ping-pong effect.

4) Quality of Service

Even though the access technologies have different QoS specification, the mobility management solution should provision, in a best-effort manner, a high quality connection regardless of PoA. As part of seamless mobility, the throughput, delay and reliability should not degrade beyond a point where the perceived quality after handoff is different from before handoff [22].

5) Security

The mobility management protocol should provide some level of security when roaming across networks. Two-way authentication should occur prior to handoff to prevent spoofing.

6) Scalability

The mobility solution should appeal globally and be scalable for wide scale adoption. This requires minimal changes to network infrastructure, low-cost implementation, and uniform agreements between different service providers and network operators.

Currently, no mobility management schemes adequately tackle all the challenges presented in this chapter. No approach taken holistically caters for all the requirements of next-generation MMPs. In the next chapter, we provide a mobility management framework that helps meet the challenges of next-generation mobility.

Table 1. Summary of mobility management protocols in next-generation all-IP wireless systems.

Protocol	Category	Capability for Widespread Adoption	Design	Drawbacks
3G/WLAN Integration	Integration Architecture	Limited	Loosely coupled/Tightly coupled integration points for managed mobility	Costly changes to network infrastructure, network-specific, non-scalable mobility
Hybrid SIP/MIP	Multi-layer mobility	Significant	Either use SIP for real-time and MIP for non-real-time applications or combine network registration entities	Assumes multimodality uniform handoff between networks. Basic handoff decision metrics
Integrated Handoff and Content Awareness Support	Advanced macromobility	Limited	Capture context information through middleware and report to context-aware services	Assumes context-aware applications.
HMIP/CIP/TIMIP/IDMP HAWAII	Micromobility protocols	Significant	Localize signalling in mobility domains	Do not support macromobility
HOPOVER	Advanced macromobility	Limited	Keep QoS information in neighbouring ARs for quick handoff	Wastes resources, requires widespread standardization
Omnicon	Macromobility	Limited	Introduce virtual interface to toggle between WLAN and GPRS based on RSS	Network specific, limited handoff decision metrics
mSCTP	Macromobility	Moderate	Use SCTP with ADDIP to toggle between IPs for handoff	Authentication problems, requires location management from different layer, limited user and content handoff decision metrics
SIP	Macromobility	Significant	Application layer terminal mobility by redirecting sessions	Lossy handoff, lack of user preferences metrics, mobility-aware applications
FMIP using MIHS	Macromobility	Significant	Use MIES for candidate router discovery and optimized network registration	No decision engine for handoff algorithms
SIP using MIHS	Macromobility	Significant	Use MIHS for movement detection and fast authentication-association messages	No decision engine for handoff algorithms

Chapter 3

Chapter 3 The Cross-Layer Mobility Management Framework

In this chapter, we propose a cross-layer mobility management framework that aims to mitigate some of the challenges faced by mobility management protocols in next-generation all-IP networks.

Here we first identify four advanced mobility scenarios found in next-generation hybrid wireless systems where the framework captures different handoff parameters and involves certain handoff decision metrics. After data capture, movement detection and network registration can take place using Mobile IP and SIP mechanisms. Context transfer including Quality of Service (QoS) provisioning ensures handoff transparency to the user and applications. The last part of the chapter shows how the handoff decision algorithms that account for Hysteresis, shadowing and the ping-pong effect, can stabilize handoffs.

3.1 Advanced Mobility Scenarios

Four mobility scenarios, identified in [2], for hybrid access networks, are categorized according to their handoff parameters and decision metrics. Handoff parameters trigger handoff when they cross a certain threshold indicating channel deterioration or a violation of host-to-network criteria. Each network providing coverage to the host displays characteristics quantified by decision metrics. Matching a host's user and application criteria to the decision metrics allows a host to select a network satisfying its host-to-network criteria from a pool of available candidate networks. Note that some handoff parameters can also be decision metrics.

The radio signal footprints of the access networks either overlap or overlay each other providing limited or continuous coverage. This can either induce forced handoff –where a node handoffs to prevent loss of connectivity, or unforced handoff –where a better access network is targeted.

3.1.1 Adjacent Homogeneous Networks

This is a simple mobility scenario where a terminal moves among networks of the same type. This is found in traditional networks where terminals horizontally handoff amongst cells of the same radio access technology. As a result, the handoff parameters and decision metrics are trivial since both networks would generally offer the same services.

Handoff Parameters

Typically, these link layer characteristics reflect the state of the channel. These can be received signal strength (RSS), noise level, and bit error rate (BER).

Handoff Decisions

Utilization is a factor that influences handoff decisions. A basestation controller would induce handoff to a cell that has lower utilization to balance the load across the edge network. Signal strength could also be decision metric where the better of two signal levels is the most reliable.

Typically, micromobility protocols such as HMIP, IDMP or CIP can manage this scenario.

3.1.2 Adjacent Heterogeneous Networks

This scenario causes vertical handoff among edge networks with different radio access technologies. It requires a multimodal device to switch to a different network interface for session continuity if handoff occurs. Here the roaming node forces vertical handoff due to an imminent loss of coverage with the current network attachment.

Handoff Parameters

The handoff parameters also reflect the L2 characteristics: signal strength, noise level, access media congestion/utilization, and bit error rate (BER). However, these values require normalization across the heterogeneous networks since the frequency, modulation and power consumption parameters are different. The handoff parameters also include communication deterioration through lower throughput and higher delays.

Handoff Decisions

There is no decision involved for loss-of-connection imminency that induces forced handoffs.

This scenario requires vertical macromobility with IPv6 address reconfiguration for session continuity if the communication channel passes through the IP packet data network (PDN).

3.1.3 Always Best Connected

Overlapping or overlaid radio signal footprints form the Always-Best-Connected (ABC) mobility scenario. Handoffs in ABC are unforced and only take place when the best-suited network, in terms of available resources, is targeted or when the current network cannot provide adequate support for the application or user. This scenario calls for advanced decision algorithms that take in multiple parameters.

Handoff Parameters

Signal strength, Committed Information Rate (CIR), data rate, reliability, delay, and power consumption.

Handoff Decisions

Security, user preference, available resources (CIR), cost budgets, mobile node trajectory, and network congestion.

ABC scenarios can only offer the best-effort network to the current user and application requirements. ABC aims to provide seamless handoff by making a link-level connection to the target network before committing to breaking the previous link.

3.1.4 Multimedia Aware

Handoff algorithms in Multimedia Aware scenarios only consider service types and high-level QoS parameters. This ensures the current service session runs on the best network. This is for bandwidth-rich real-time multimedia applications. This scenario supports soft vertical handoff.

Handoff Parameters

Data rate, reliability, delay, and service class.

Handoff Decisions

Available resources, QoS parameters, congestion, and service priority.

This QoS-aware scenario requires high-level awareness of the carrier content.

The categorized mobility scenarios entail different handoff parameters and decision metrics involved in the handoff algorithm. The mobility management framework copes with the advanced mobility scenarios.

3.2 Handoff Management

The cross-layer mobility management framework handles advanced mobility scenario roaming by employing an integrated decision strategy where content-awareness information matches context-awareness information through the architecture shown in Table 2.

Fig. 1 shows the architecture of the mobility management framework for a dual-mode client terminal. It is important to note that the Media Independent Handover Function (MIHF) has a remote counterpart within the network.

Table 2. Partial content and context information for advanced handoff.

Content	Protocol Stack Layer	Factors
User Preferences	L6 User plane	Cost budgets, power consumption, network preference, application priority, security level
Application Requirements	L5	Service class, service type, service priority, data rate, delay tolerance, security
Traffic Flows	L4	Achieved throughput, traffic class, connection-oriented/less, congestion
Network Topology	L3	Router discovery, Route, security, packet loss
Context		
Link Characteristics	L2	Throughput, media access utilization, noise level, signal strength, loss ratio, radio signal coverage, link connectivity, jitter
Node Trajectory	Cross-protocol	Neighbour maps, mobility pattern, moving velocity
Frequency Characteristics	L1	Antenna characteristics, power consumption

The IEEE802.21 Media Independent Handover Services (MIHS) captures context-awareness information whilst layer-specific Service Access Points (SAP), which interfaces with mobility-related protocols, capture content-awareness information. The Cross Layer Manager (XLM) houses 3 functional entities: Link Information Manager (LIM), Decision Engine (DE), and the Handoff Manager (HM).

The LIM captures context information through Media Independent Event Services

(MIES) and Media Independent Information Services (MIIS) from the MIHF and content information through layer-specific SAP, each defined by the attached mobility information source. The information is stored in a data structure called the State Table, which maintains a snapshot of the network. The LIM sends a network report to the DE that houses all the decision algorithms for target network selection. The DE may wake up the HM, which handles handoff procedures by utilizing Media Independent Command Services (MICS) that allow link configuration and switching network interfaces. The IEEE802.21 draft [4] defines the MIHF and the lower-layer SAP.

The XLM is a host-based manager that interfaces with the MIHF directly and uses the offered MIHS for movement detection and link management. Furthermore, the XLM uses MIP and SIP for network registration whilst docking their handoff management features. The XLM does not follow the 3GPP network-controlled mobility paradigm as it induces handoff locally within the host's stack.

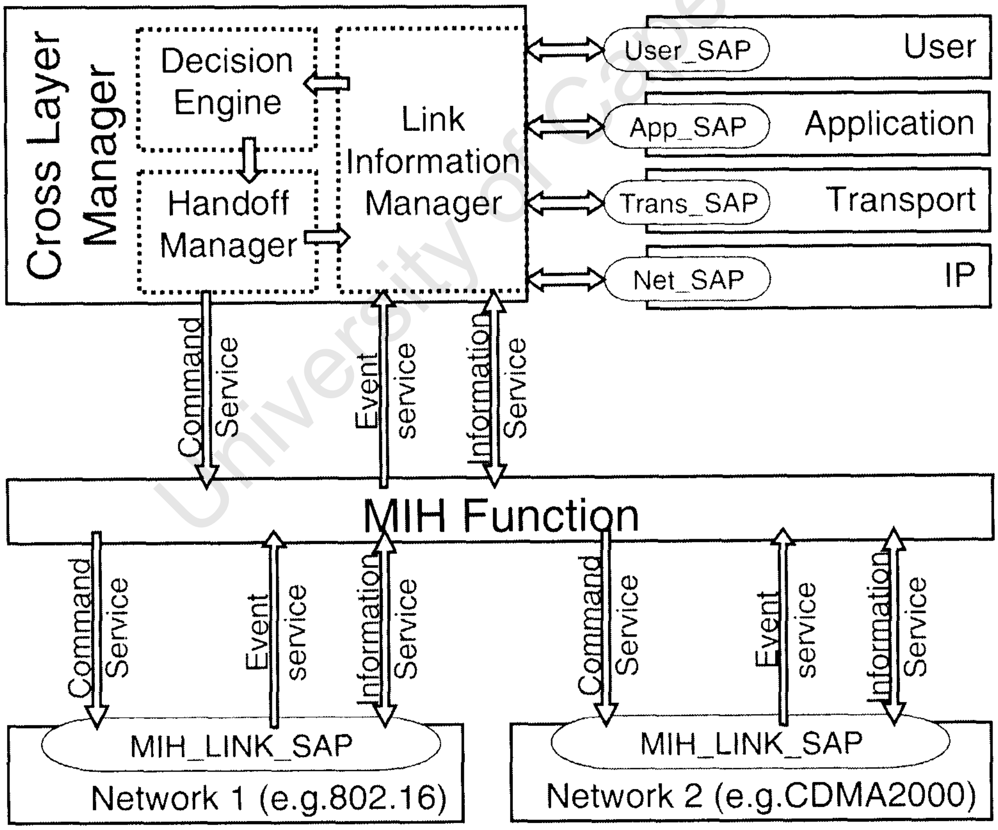

Figure 2. The Proposed Cross Layer Manager Framework based in the Media Independent Handover Function.

3.2.1 Parameter Capture

The Link Information Manager (LIM) periodically polls for network state updates when the XLM is in an idle state. The XLM is in an idle state when the mobile node is stationary for a foreseeable time in the future reflected by the movement detection parameters in the link characteristics. If the XLM is actively scanning for networks then the LIM gets state updates more often.

Thresholds programmed into the SAP serve to earmark a change in the link conditions or user/application criteria. When a SAP threshold is crossed (a change in the handoff parameters) for any layer, the LIM wakes up to update the State Table. Thresholds are set to ensure power-efficiency in the XLM. To control the number of updates, the thresholds are set to reflect a significant change in the monitored handoff parameters.

The LIM stores the updates as handoff parameters and handoff decision metrics in the State Table. The State Table maintains context-information for each connected link, whilst there is only one copy of content information. If a significant change occurs in the State Table, the LIM wakes up the DE with a copy of the State Table as a network report.

3.2.2 Handoff Initiation

Handoff procedures commence on movement detection or when a better-suited network is found for the current content and user criteria.

Movement detection is derived from the link parameter reports that are sent by the MIHF periodically. If the link characteristics continuously and deliberately deteriorate, the XLM assumes movement. This then calls for forced handoff.

When the State Table shows an idle link, with available network coverage, that is better suited to handle the current service session, then the node should handoff.

3.2.3 Target Network Selection

The Decision Engine runs its handoff algorithms for optimal target network selection when it receives an updated network report via the LIM's State Table.

The DE can easily categorize which mobility scenario the host is in by monitoring the active links and their types. If the node moves on one available active link, and the other links have no network coverage, then this shows horizontal micromobility. If another link is active, and the current link is going down, then the DE assumes vertical macromobility. If the current link is not deteriorating but one is improving then it is either ABC or multimedia aware depending on the handoff parameters. Table 3 summarizes this concept.

Depending on the mobility scenario, the DE evaluates the choice of candidate networks based on the scenario's decision metrics when a handoff parameter triggers a DE wakeup.

Table 3. Classification of mobility scenarios by the Decision Engine depending on the context information.

Link1	Link2	Mobility Scenario
Going down	No connection	Movement across homogeneous networks
Going down	Available	Movement across heterogeneous networks
Active	Improving	ABC or Multimedia Aware depending on handoff parameters

The algorithm is based on the cost functions shown in [21] where the choice of network depends on the cost function with the least value. The lowest value represents the lowest cost to the network and is thus the ideal network to handoff to. A cost function is the aggregation of normalized network parameters with corresponding weights. The costs functions are used as in [2], with the weights obtained from the content information and the cost parameters obtained from the context information.

The XLM can support fast handoffs by anticipating when a handoff is imminent. If the handoff algorithm shows that a network is fast becoming favourable then the DE can assume handoff to it is imminent. If only limited handoff parameters increase the cost function and they are changing the XLM can subscribe to the certain events that correlate to the handoff parameters, and poll them continuously. This results in fast handoff because the DE does not run the cost function again since it only depends on limited parameters changing. Other optimization techniques can be used for the handoff algorithms to ensure fast network selection.

3.2.4 Handoff Procedures

Once the target network has been selected the DE wakes up the Handoff Manager (HM) with a handoff initiate message that includes the selected network. The HM handles the authentication and association signaling and switches the current service session to the new link. If the terminal supports multi-service sessions then each service flow can be switched to the best-suited link.

The HM uses Media Independent Handover Services (MICS) for switching and configuring links. This gives the XLM a generalized means to initiate handoff from which native L2 connection establishment takes place between the host's network interface and access point. Once a L2 connection is established, the session can continue if the node is active, otherwise a session can be set up on the new link.

The IEEE802.21-assisted handoff is shown in Fig. 2.

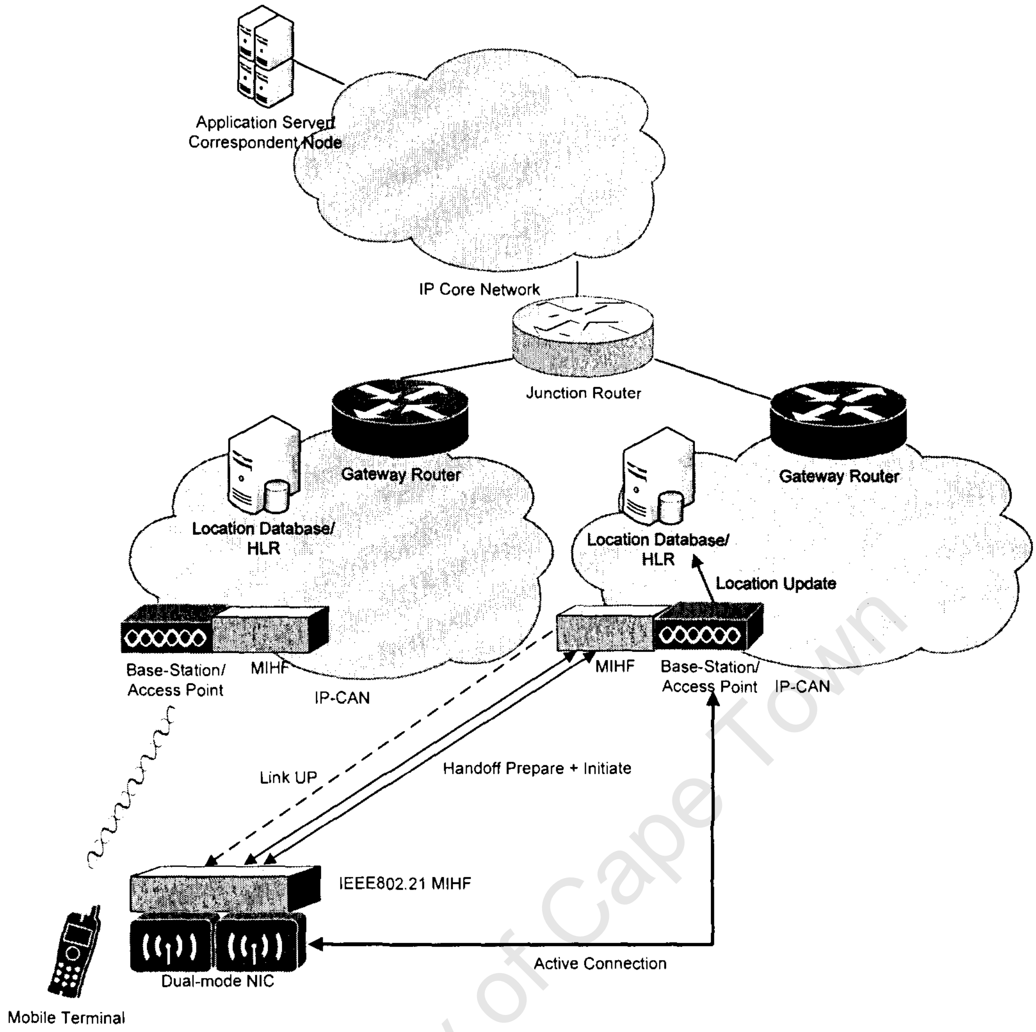

Figure 3. IEEE802.21-assisted Handoff for a Dual-Mode Mobile Host.

3.3 Network Registration

For macromobility the IPv6 address of the host needs to be configured to reflect the current topology. The XLM uses standard MIPv6 mobility mechanisms [8] for CoA configuration and sending Binding Updates (BU) including route optimization to avoid tunneling, IPsec for security, and AAA for authentication. Thus security and authentication is available for IP connections when they re-establish with a new network after handoff. The AAA and IPsec agents are available in the core network to control and manage the end-host's authentication and authorization.

For link-level network registration, native authentication and authorization procedures are used that are specified by the link-layer technology standard. These authentication and authorization procedures are different for every access technology; however the assumption here is that the node also supports the technology that it is handing off to due to its multimodal nature.

However an extension is made to Mobile IPv6 to ensure minimal packet loss when a host migrates away from its current subnet. Packets are buffered at the serving or old access router (oAR) when the link-local address of the host changes or when it does not detect the host. When a BU is sent from the host to the oAR via the new access router (nAR) the buffered packets are forwarded back to the host via the nAR. This is shown in Fig. 2.

Two methods can be used for fast address configuration:

1) FMIP [33] can be used for fast address configuration through router solicitation for proxy advertisement (RtSolPr). This obtains the target network's prefix from the old access router (oAR) which stores neighbouring routers' MAC addresses and corresponding network prefixes.

2) The neighbouring prefix can be sent through MIIS so that the hosts can autoconfigure the CoA to reduce network association time.

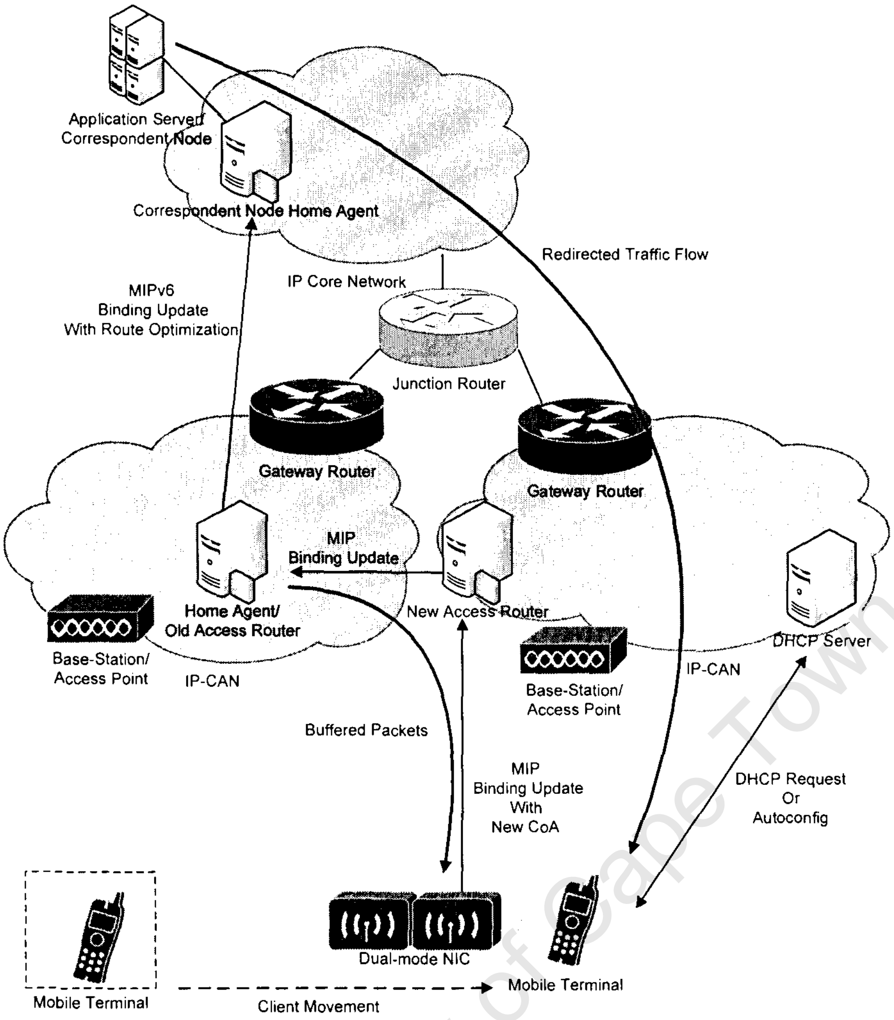

Figure 4. Mobile IPv6 Registration and Packet Buffering.

The XLM also uses standard SIP [12] for session mobility by using mid-call mobility. A hybrid SIP/MIP solution can be used by both collocating SIP's Registrar and MIP's Home Agent as shown in [18] and piggybacking the newly configured IPv6 address on SIP's Re-INVITE. The advantage of using the scheme in [18] is reducing the signaling delay when registering with a new network.

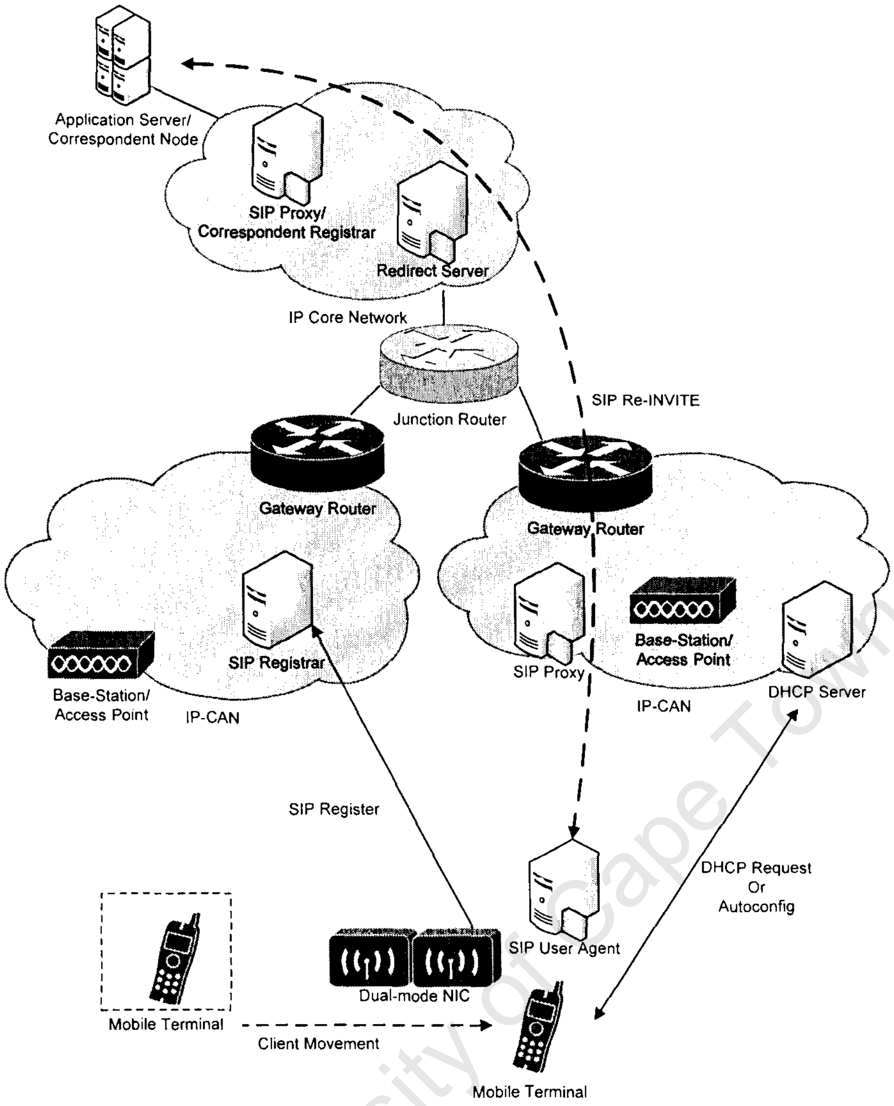

Figure 5. **Mid-call mobility SIP network registration.**

The signaling for handoff management is shown in the message sequence in Fig. 5.

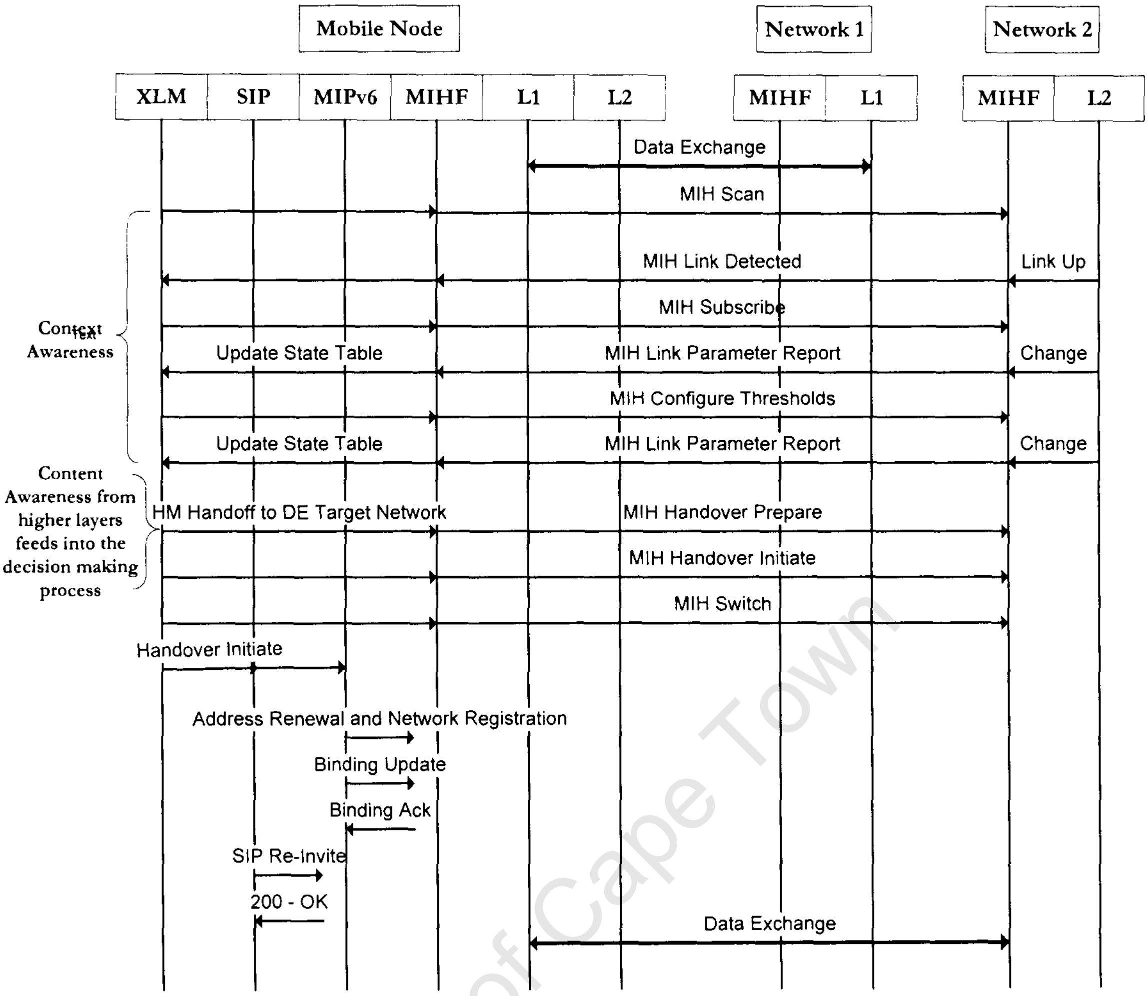

Figure 6. Message Sequence for an example of XLM Handoff Management.

If the node detects movement early enough to register for a layer-3 connection on the target network before losing the L2 connection then there will be no discernible break in IP connections. However, if the terminal can support two (or more) active links simultaneously then a totally seamless handoff is possible by shifting the session to the ideal link.

3.4 Quality of Service and Context Transfer

Mobility management protocols have to ensure sufficient QoS provisioning for a handoff session otherwise the service requirements are violated even though QoS is not a mobility management criterion. The XLM provides QoS guarantees for:

1) End-to-end QoS provisioning for the path between the host and the correspondent node (CN).

2) Mapping the QoS guarantees from the old radio access network (RAN) to the new RAN through transferring context information.

The XLM uses the QoS extension drafted in [11] to ensure an end-to-end service guarantee before network registration. The QoS service requirements from the State Table are injected into the QoS object by the XLM. The QoS object is piggybacked on a BU which is then sent on a Hop-by-Hop or Destination option basis to the CN via the Home Agent (HA). If route optimization is supported then the HA maintains the QoS option in the BU when it forwards the BU to the CN. Similarly if the node still receives CN packets from the HA subsequent to sending the first BU, then it can send a BU with a QoS option to the CN directly. If the core or access networks do not support route optimization then a triangular path is setup between the node, HA and CN. If the HA supports the QoS extension then only a QoS guaranteed path is setup between the HA and the node. Since a connection already exists between the home network and the CN, no QoS path is needed between the HA and the CN. It is assumed that the HA-CN path is safe since QoS negotiation is performed when the connection was initially setup. However IPv6 stacks are expected in NGWN with support for route optimization. The routers along the path then examine the contents of the QoS object in the BU and map the QoS requirements onto their domain. This guarantees a safe reliable connection for the new route between the host and the CN if route optimization is supported. Since this happens prior to network registration it allows for resource reservation. This can be seen in Fig. 6 where a QoS path is setup prior to registering with the new network.

The routers along the path implement QoS policies for their domain, typically a GGSN gateway in 3GPP. However, these can be any routers enforcing QoS policies along the node's data-path. If the routers along the path cannot provide adequate support to the QoS requirements according to their policy a Binding Update negative acknowledgement (BUack) is sent back to the host. If this happens in a timely manner the XLM can then select a different route for the data path.

The QoS path guarantee only acts to stabilize the network when handing off a node so that the new connection is not perceived as lossy. This serves to make the handoff process transparent to sessions and users. It is expected that other QoS network mechanisms would kick in after the node has stabilized with the newly established connection.

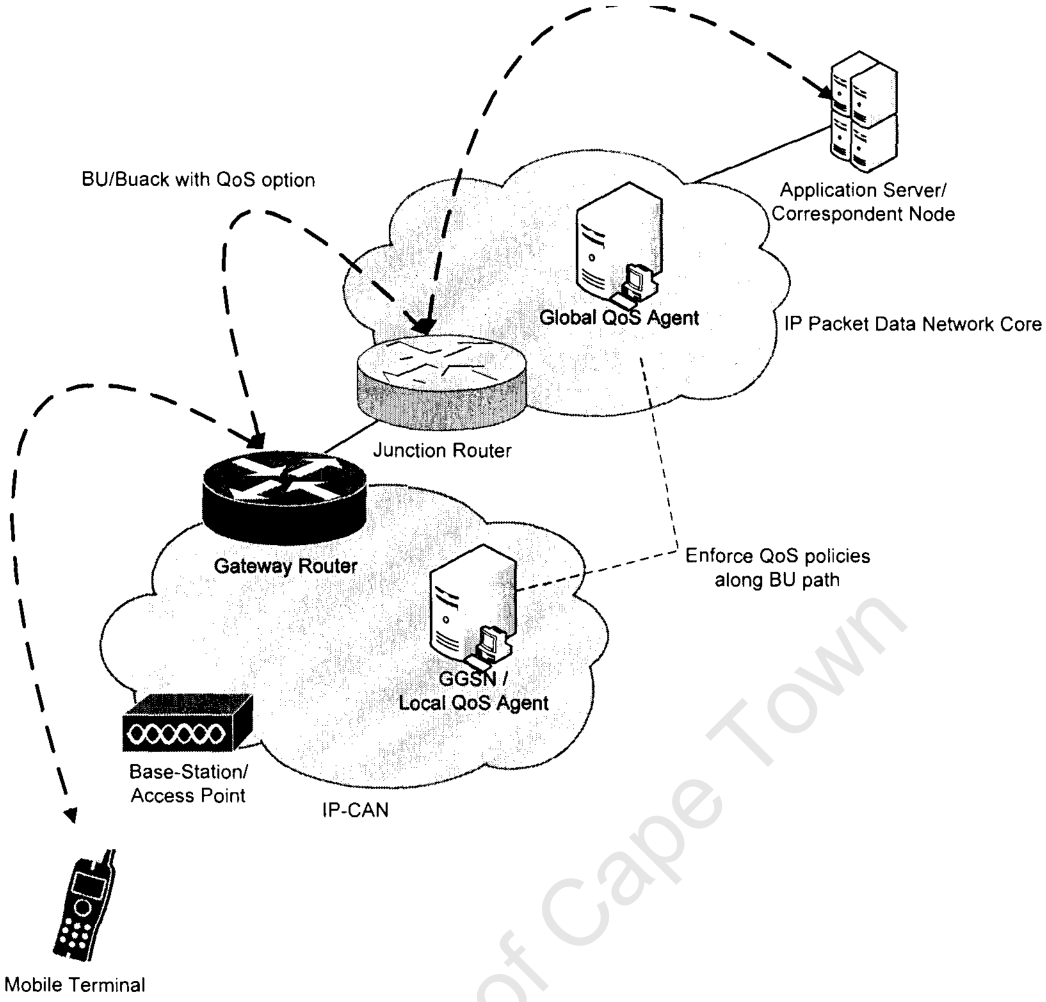

Figure 7. Resource reservation along the handoff data-path between the host and CN for QoS guarantees.

To ensure QoS reliability the XLM can transfer context information from the old RAN to the new RAN. The context information includes security (high-layer), QoS, and header compression information. The context transfer is to at least ensure the same perceived quality for both access networks for seamless mobility. Note that this is a best-effort service and the new access network should at least provide the same perceived QoS quality as the previous RAN. This ensures minimum disruption when moving across different technologies after handoff. For context transfer CARD and CTP are negated because the XLM uses IEEE802.21 MIIS for fast relay of the context information.

3.5 Handoff Stability

Due to the sporadic asynchronous nature of wireless environments intelligent handoff algorithms need to be developed that adapt to the ever-changing conditions. There are adverse conditions found in wireless systems that cause the ping-pong effect.

The ping-pong effect is when a host toggles amongst networks frequently in a short time frame due to ostensible handoff imminency. However eager handoff is unnecessary and has profound negative effects on the performance of the network especially for active communication sessions. The ping-pong effect may occur due to line-of-sight shadowing, sudden spikes and dips in the link, and network congestion due to multiple hosts realizing handoff.

3.5.1 Line-of-Sight Shadowing

When objects, such as trees and buildings, obstruct the line-of-sight (LOS) radio transmission between the host's antenna and the base-station, the terminal will experience a drop in signal strength usually resulting in bad link characteristics. Simple handoff algorithms might construe this as loss of coverage initiating handoff to a different network eagerly. However the LOS shadowing would only be a temporary dip in the link characteristics of the originally connected network which could cause the ping-pong handover back to the network. Shadowing margins are needed to introduce delays to ensure the algorithm does not immediately initiate handoff. However the algorithm needs to maintain a balance between delaying handoff unnecessarily and immediately handing off.

The XLM's Decision Engine has to continuously monitor the link characteristics and if the handoff parameters deliberately deteriorate, the XLM assumes movement. In addition because the XLM can actively scan for other links it can determine the cost of handoff. If there is LOS shadowing on an active link, with good conditions on a separate idle link, then the cost of handoff is low and the XLM can adaptively decrease the shadowing margin. However if the cost of handoff is high (peer networks have bad conditions) then the shadowing margin should increase to protect running sessions from disruption.

3.5.2 Signal Strength Amplification

In wireless environments it is not uncommon if a network suddenly shows better characteristics than the current serving network. However spikes in the link characteristics of an idle interface may be temporary and advanced handoff algorithms should not immediately handoff.

Only if a link shows continuous and consistent improvement should handoff take place. A bias towards networks with consistently strong signal strength gives better performance. The bias ensures that even though at times the network might not display the best conditions out of all available RANs, handoff does not occur because the serving network is the most reliable. The bias would latch onto a superior network that displays long-term favourable conditions. The bias can also be adjusted with the cost of handoff. If the cost of handoff is high, the bias will be correspondingly strong and vice versa. This is to ensure the XLM does not propagate the ping-pong effect through eager handoffs.

3.5.3 Mass Handoff Realization

Handoff realization has adverse effects on the performance of network systems and propagates the ping-pong effect. Handoff realization is when multiple terminals are on the same network and simultaneously realize a better network is in the vicinity and switch to it. This massive migration increases the network load on the new network, and the increase in the congestion may lead the terminals to switch back to the old network [2]. This cycle can be broken through intelligent decision algorithms.

A possible solution is by enabling nodes with XLM-stacks to 'enquire' whether they can handoff to a network by using a BU. A BUnack could be sent back to the node to deny it from connecting to the network. This would necessitate augmenting the MIPv6 protocol by including a Call Admission Controller into the handoff procedure. The Call Admission Controller (CAC), which is typically collocated with the RNC in cellular systems or access routers (AR) in nomadic networks, can measure the network load by monitoring the media access utilization. This would require the RNC or AR to act as a foreign agent so that it could intercept the BU and apply admission control algorithms for the enquiring node. If the CAC finds the load unfavourable, it sends a BUnack back to the node. The XLM would then search for a different network.

We have proposed a cross-layer mobility management framework with the main element being a Cross Layer Manager (XLM). The XLM houses advanced decision algorithms which are used for selecting optimal networks. The XLM is optimally placed to gather context and content information to cope with the stringent service requirements in B3G networks. The XLM uses IEEE802.21 for managing links and polling network conditions. Network registration is accomplished through dual MIP and SIP registration for terminal and session macromobility. The XLM guarantees a certain level of QoS when handing off to a new network by transferring context information and reserving resources along the data path to the CN. In addition suggestions are given in order to mitigate the ping-pong effect.

Chapter 4

Chapter 4 Experiments

In this chapter we evaluate the impact on performance of introducing the Cross Layer Manager (XLM) to the protocol stack. Performance is critical to ensure seamless and timely handoff. The results will allows us to see if XLM degrades the performance in a mobility scenario.

It has already been shown in [31] that IEEE802.21 Media Independent Handover Services (MIHS) allows for the seamless handoff of multimedia services over heterogeneous networks. Therefore we take a different approach and evaluate the best Media Independent Handover User (MIH User) from the two popular MMPs, SIP and MIP, and the proposed solution, XLM.

We add content-awareness to SIP and MIP by exposing them to SAP across the protocol stack to ensure fairness among the 3 protocols since content-awareness is already included in the architecture of the XLM. We then monitor the reactivity of the protocols to changes in the state of the protocol stack and the number of messages generated by the protocols as a result of these changes.

4.1 Evaluation Framework

The experiment was carried out using OMNet++ (v3.2 beta 1) simulator with the IPv6SuiteWithINET extension. The simulator was installed and run under Cygwin v.2.5 on a Windows XP SP2 Pentium 4 3.2 GHz, 512 Mb RAM machine.

OMNet++ is an open source discrete event simulator that supports C++ programming. The IPv6SuiteWithINET is a module extension of OMNet++ originally written by Monash University. The suite implements a host of IPv6 RFCs including RFC 3775 for Mobile IPv6. Cygwin is a free Linux emulator for the Windows platform that allows pre-compiled binaries to run in a Linux-like environment.

A mobility framework was created to evaluate SIP, MIP and the XLM as an MIH User. An MIH User is a mobility management protocol (MMP) that utilizes MIHS by interfacing with the MIHF. The Media Independent Handover Function (MIHF) was built with the full functionality and it exports all the Media Independent Handover Services (MIHF) seen in [4]. The MIHF maintains a first-in first-out (FIFO) service queue of all the command and event messages that need to be translated to MIHS. The MIHS were programmed from template .msg files which reflect the format and field definition seen in the standard. Since the SIP protocol is not supported, a skeleton image of the protocol was coded as an application layer protocol. The coded SIP module does not support any functionality outside what is needed to act as an MIH User.

The MIHF converts link events into Media Independent Event Services (MIES) which it then passes to the higher-layer MIH user.

In addition MICS messages are also translated into link-specific commands and are passed on to the lower layers. The protocol stack is multimodal with support for WLAN, WiMAX and CDMA2000 technologies. The MIH User receives MIES and issues MICS to the MIHF. For content-awareness all MIH Users are connected to a single higher layer and a lower layer.

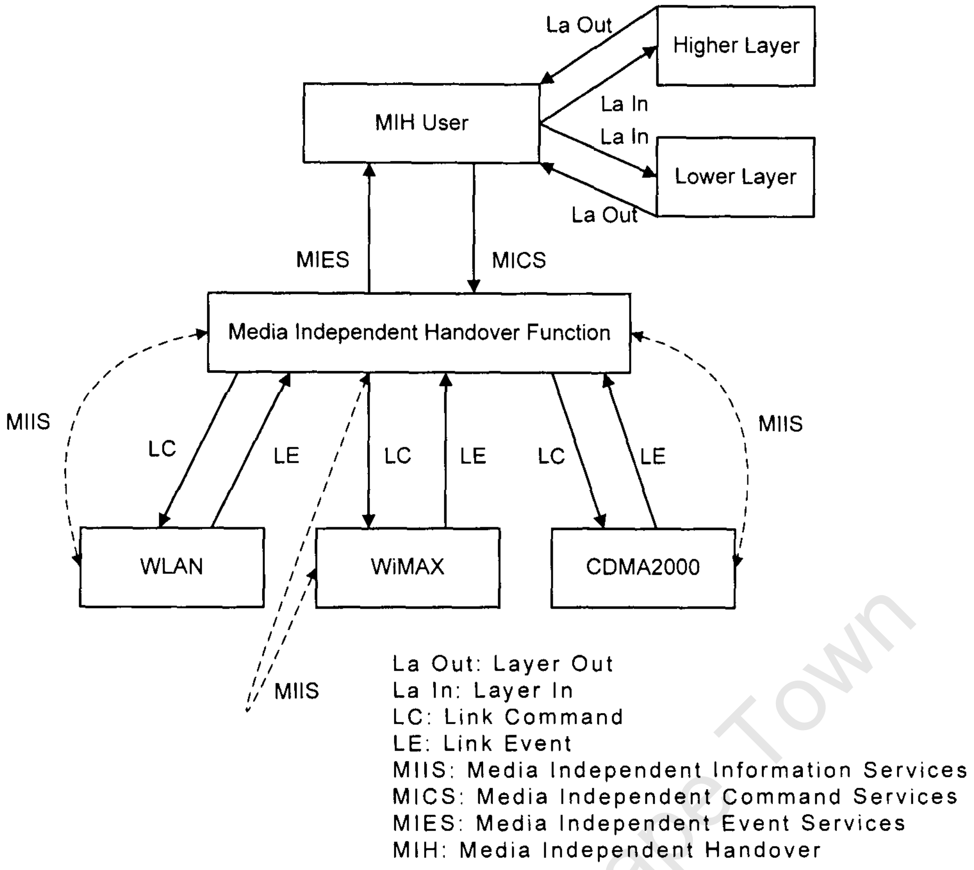

Figure 8. Simulation Framework for Evaluating Stack Reactivity.

4.1.1 Test Cases

There are 3 test cases that implement a different protocol stack. The test cases are for SIP, MIP and XLM. In each test case, a single MMP acts as an MIH User in order to find the ideal user. Each MMP is granted content-awareness capabilities, even though this is not the case in practice, to ensure fairness when deciding on the most efficient MIH user.

1. MIP as MIH User

Here MIP is the MIH User by issuing MICS and receiving MIES. MIP is coded as a L3 protocol and content-awareness information has to traverse through the protocol stack from the user to the application layer down to the network layer where MIP resides.

2. SIP as MIH User

SIP is an application layer protocol and resides in the layer that is just below the user which sends content-awareness information. In addition the lower layers in the protocol send also send information pertinent for handoff. MIES events travel from L2.5 whilst MICS commands are issued via the layers below SIP.

3. XLM as MIH User

The XLM resides as a cross-layer function with exposure to all the layers simultaneously. MIES, MICS and content-awareness information is captured immediately by the XLM neutralizing the need for PDUs to traverse the protocol stack.

4. Hybrid SIP/MIP as MIH User

In this case, a hybrid SIP/MIP solution is employed to utilize the MIHS exported by the MIHF. Here both the SIP and MIP can issue MICS and receive MIES from the MIHF. The MIES is limited to the event that SIP and MIP are subscribed to. MIP and SIP independently issue a MIH Subscribe event to the MIHF to define the set of services that each protocol is interested in. The MIHF then only forwards subscribed events to SIP and MIP.

In this case SIP and MIP have independent decision engines. This is the case in [17] and [18] where sessions are divided between MIP and SIP or only network registration is integrated. As a result, content-awareness is divided between the two MMPs as each requires a copy of the higher layer parameters as a feed into the handoff decision.

This is a useful test case since hybrid solutions could be implemented in protocol stacks if they continuously gain popularity in academia and industry.

4.1.2 Simulation Scenario

All test cases were run on the same network topology where the network interfaces exhibit link characteristics common for wireless nodes in a volatile environment. The terminal is seen to be moving with high velocity across overlapping or adjacent radio signal footprints belonging to the different technologies of WLAN, WiMAX and CDMA. Since the terminal is multimodal, the link events of each technology are sent to the MIHF which forwards it to the MMPs as MIES. The MMPs receive the MIES from the layer they reside in and optionally reply with MICS.

Please refer to Appendix A for more details on the implementation of the simulation.

4.1.3 Motivation for Experiment

The motivation for the experimental comparison between SIP, MIP, hybrid SIP/MIP and XLM is to target the ideal MIH user from the protocols. The ideal MIH user is the mobility protocol that is ideally situated in the protocol stack to capture the MIES and issue MICS in a timely manner suitable for a volatile wireless environment. In addition the ideal MIH user can also capture content-awareness information for advanced handoff decisions. MMPs in advanced mobility scenarios are required to respond very quickly to external stimuli for seamless handoff. The simulation scenario allows us to measure the protocol's reactivity to external stimuli and in-stack parameters.

4.1.4 Performance Metrics

For protocol reactivity we measure message arrival time and message response time. These are delay parameters that indicate how efficient an MMP is to external stimuli and internal stack parameters.

The Message Arrival Time (MAT) is the time it takes for an MMP to capture a link signal. The MAT is the aggregate time of the links triggering an MIES event, the time for MIHS translation by the MIHF, and the time it takes for the MMPs to interface with the SAP to recognize an MIES event. This is shown in the following equation:

$$MAT = t_{linkevent} + t_{MIHF} + t_{SAP}$$

The Message Response Time (MRT) is the time it takes for an MMP to issue an MICS command to the MIHF in addition to the time it takes for a link to receive it. The MRT aggregates the time it takes for an MMP to decide which MICS command to issue, the time it takes for the SAP to interface an MICS to the MIHF, the time it takes the MIHF to translate the MICS command to a link signal, and the time it takes for the link to recognize the command. The equation for the MRT is as follows:

$$MRT = t_{MICS} + t_{SAP} + t_{MIHF} + t_{linkcommand}$$

The Number of Unserviced Messages is the number of messages yet to be serviced by the MIHF. The MIHF maintains a queue for all received MIES and MICS messages and is incremented every time a new MIES or MICS message arrives. The queue is decremented when the corresponding event is delivered to the MMP or link command is issued to the interface. The number of messages remaining in the queue or buffer at the end of the simulation time is the number of unserviced messages. The unserviced messages relate to how much load is on the MIHF and how timely a MMP is to responding to events. There is some delay induced by translating a MIES into a link event or MICS into a link command. However this is generally negligible because the delay is typically close to zero and does not account for why the MIHF queue is not empty on average. The defining reason is due to the response by the MIH Users that grab events from the MIHF queue. The MIHF will always keep an event in the queue until it the MIH User is ready to receive it. MICS commands don't affect the unserviced messages significantly since they are not dependent on the MIH User's response.

The Number of Received Messages is the total number of MIES received by each MMP from the MIHF during a simulation session.

The Number of Sent Messages is the total number of MICS issued by an MMP to the

MIHF for a single simulation session.

4.2 Experimental Results

The results of the simulation are shown for the different test cases with the performance metrics that are relevant for that simulation run.

All line graphs displayed henceforth do not start at the origin exactly because the simulation takes time to initialize the bucket variables that capture the results.

4.2.1 Simulation Validation

In this section results are presented that help validate the simulation setup against what is expected from the experiment.

To ensure randomness in the type of link signals generated by the 3 wireless interfaces in the experimental protocol stack we measure the type of signal occurrence for a set simulation time period.

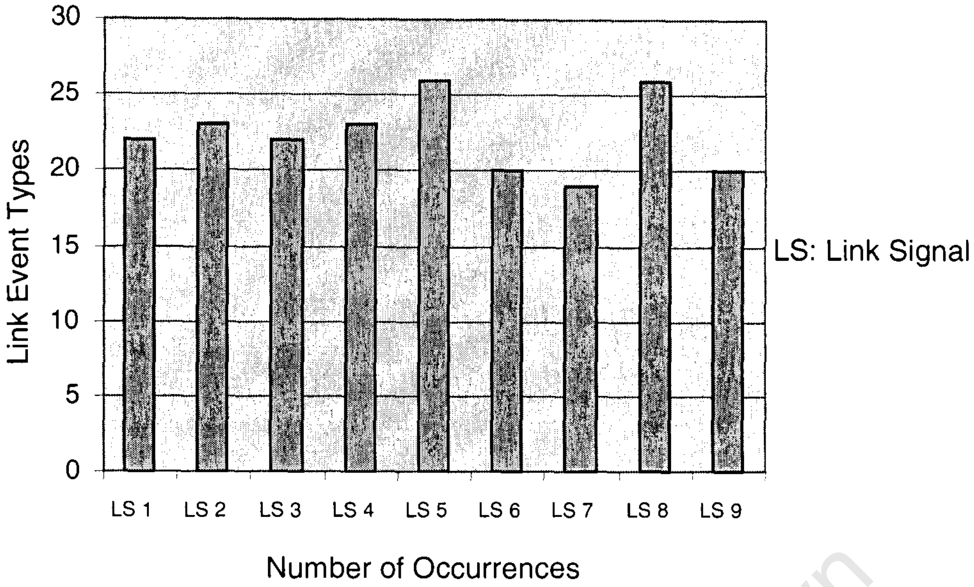

Figure 9. A Graph showing the variance of link signals occurring across the 3 different interfaces connected to the Media Independent Handover Function.

As seen from Fig. 9 there is stochasticity in the type of link signal generated by any of the 3 connected link interfaces as is expected in a mobile wireless environment. The probability of any link signal occurring from the interfaces is uniform on average.

These link signals are captured by the MIHF on occurrence. Fig. 10 shows the cumulative number of link signals in the MIHF across a simulation time.

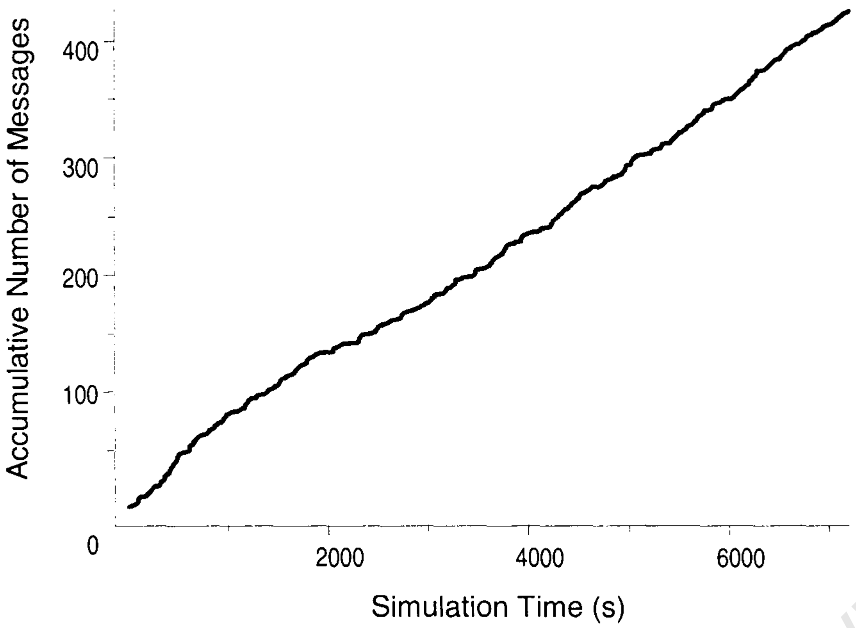

Figure 10. Graph showing the accumulative number of link signals received at the Media Independent Handover Function.

Because the probability of link signals occurring is uniform within a limited period, the accumulative number of interface signals captured by the MIHF is roughly linear.

In addition, we record the number of Media Independent Command Services (MICS) issued to the MIHF for validation. Fig. 11 shows an ascending line graph of the cumulative number of MICS captured by the MIHF from all MIH Users. MICS issue commands in response to MIH Events, which are random according to the wireless environment. Thus, the MICS have a random probability dependency on the MIH Events. However, the number of commands compared to the number of events for a single simulation run is small because not every event warrants a response through a command. In addition, the number of commands issued by MIH Users is small relative to the number of events received by the users, with polling commands being the most common. For visualization's sake we narrow down (zoom in) the simulation time-line for graphs depicting commands in order to view the number of commands clearly.

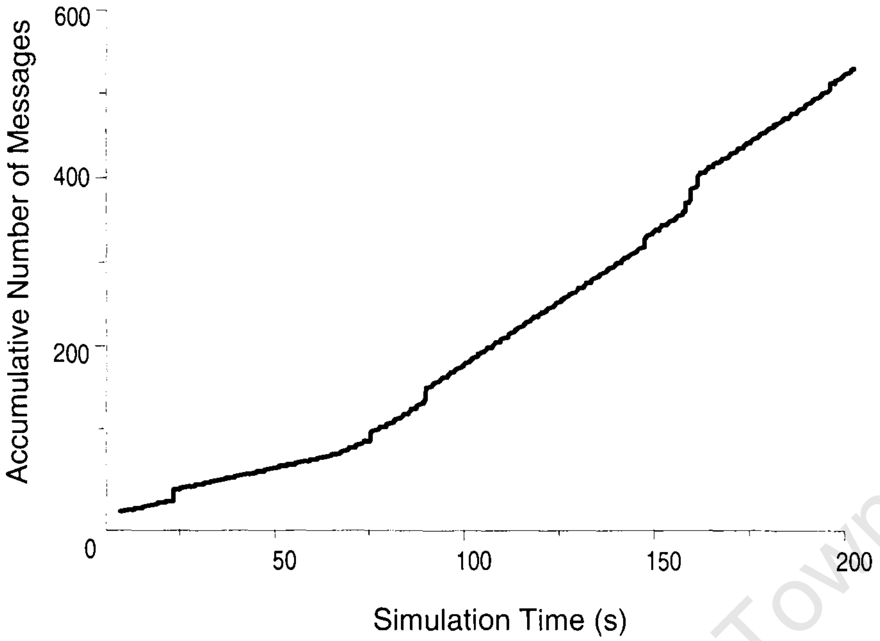

Figure 11. Graph showing the accumulative number of Media Independent Handover Commands received at the Media Independent Handover Function.

The MIHF translates MICS commands to link-specific commands for each connected interface. Fig. 12 shows the link commands generated at the MIHF.

Due to the decoding delay in the MIHF, the number of link commands issued is less than the MIH Commands received by the MIHF. The probability of a link command occurrence is dependent on the probability of a MIH Command receipt and the probability of translation by the MIHF.

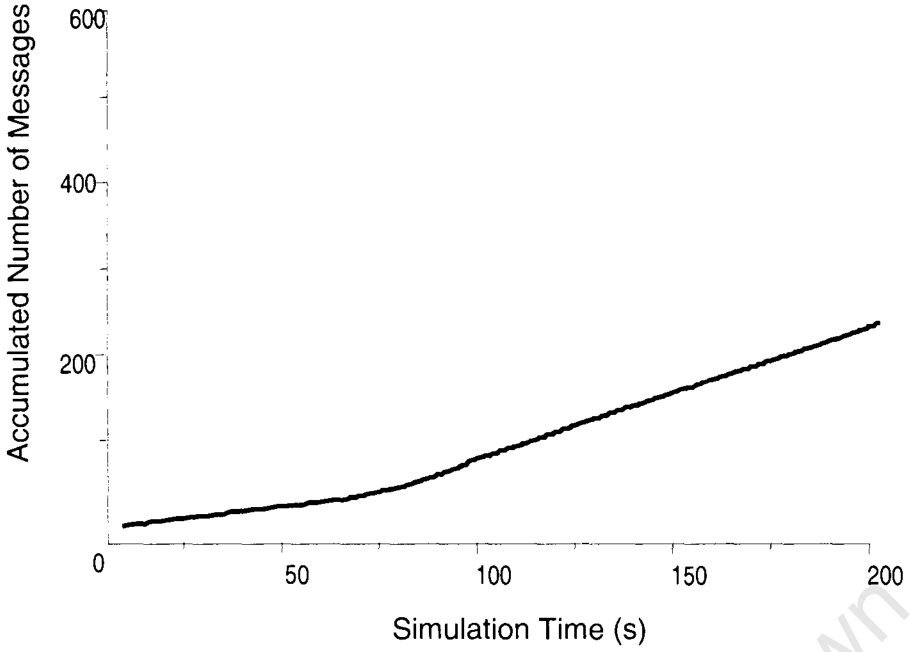

Figure 12. Graph showing the accumulative number of link commands sent from the Media Independent Handover Function.

Concurrently the link signals generated by the interfaces translate into MIH Events using Media Independent Event Services (MIES). These Events are forwarded to the MIH Users, which could issue MIH Commands. Fig. 13 shows the cumulative number of MIH Events released by the MIHF.

The probability of forwarding a MIH Event depends on the probability of link signal occurrence and the probability of the MIHF translating the signal. There is a probability that the MIHF might not translate the link signal into an MIH Event because of the decoding and encoding delay involved in creating an MIH Event. The MIHF maintains a queue of all unserviced link signals as it prioritizes event capturing over event translation.

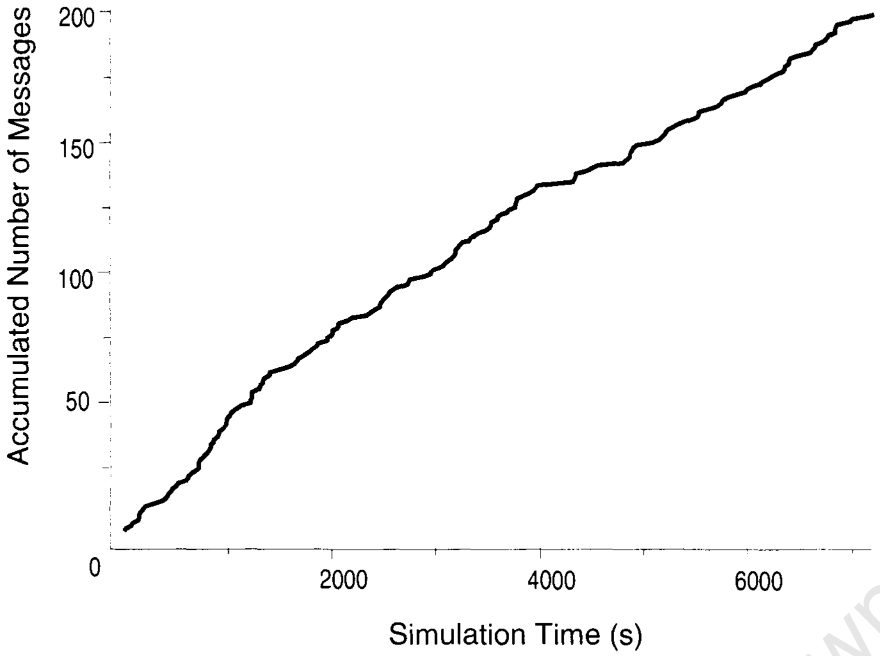

Figure 13. Graph showing the accumulative number of Media Independent Handover Events sent from the Media Independent Handover Function.

In general, the number of commands issued is less than the number of events because the MMP can choose not to send commands in response to every received event. Furthermore, the delay incurred for event notification to the MMP could result in a slow corresponding command response.

4.2.2 Performance Results

We compare the performance of SIP, MIP and XLM in terms of load and reactivity to determine which of the mobility protocols is the ideal MIH user.

4.2.2.1 Load

Here we measure the load, in terms of the number of received and sent messages, on the MIHF to determine the efficiency of the MIH Users. The mobility protocols that use MIHS are hybrid SIP/MIP and XLM. Hybrid SIP/MIP is better than sole SIP or MIP since it handoffs both a session and a terminal from one network to another. Hybrid SIP/MIP solutions could be the mobility management scheme of choice for next-generation networks. To compete with XLM's hybrid network registration, we use the hybrid model to ensure fairness.

In the first mobility run, SIP and MIP co-exist in the protocol stack since they are able to run concurrently as mobility processes. In the second mobility run, XLM is the MIH User with MIHF as its sub-layer.

Table 4. Statistics for the Simulated Mobility Runs.

SIP/MIP	Number of Generated Events = 43152 Total Message Count in the MIHF = 103461 Total Messages Sent by the MIHF = 25737 Total Messages Received by the MIHF = 77724 Number of Unserviced Messages = 34573
XLM	Number of Generated Events = 43152 Total Message Count in the MIHF = 52411 Total Messages Sent by the MIHF = 9097 Total Messages Received by the MIHF = 43314 Number of Unserviced Messages = 25005

Table 4 shows the statistics obtained for the two mobility runs where the load on the MIHF is measured. The generated events, which are the same in type and number for both mobility runs to ensure fairness, are the changes in the link conditions of the connected interfaces.

The total message count, which is the aggregate of the number of sent and received messages by the MIHF, is less for XLM's mobility run than for SIP and MIP. The sent messages are the number of MIH events sent by the MIHF due to requests by the MIH users (SIP, MIP and XLM) or due to changes in the link conditions. The received messages are the number of MIH commands issued to the MIHF by MIH users as a response to changes in the link conditions or periodic polls on the links.

The load on the MIHF is less for XLM's mobility run throughout, including the total messages received and sent by and from the MIHF.

The number of unserviced messages is the number of unserviced events and commands remaining in the MIHF queue. The MIHF services the queue by either translating generic MICS to link-specific link commands, or link-specific events into MIES. The number of unserviced messages is lower for XLM than for SIP/MIP because SIP/MIP's response to MIHS is lower and thus lags behind the events that are instantaneously occurring. SIP/MIP reacts to the latent events whilst the service list queue increases with pending events. XLM's fast response to external and internal stimuli allows more messages to be serviced resulting in a small unserviced queue.

4.2.2.2 Protocol Stack Reactivity

The reactivity of a protocol stack measures the timeliness of responding to external stimuli, such as network conditions, and internal stimuli, such as changes in user and application parameters. Here we measure the MRT and the MAT it takes for SIP, MIP and XLM to respond to Media Independent Event Services (MIES) and act accordingly with Media Independent Command Services (MICS). When measuring reactivity, the use of hybrid SIP/MIP or sole SIP and MIP does not matter since both schemes act independently when using MIHS.

The smallest mean MAT, as seen in Fig. 4, belongs to XLM whilst SIP and MIP have similar MATs, with SIP's MAT being slightly larger. The graphs in Fig. 4 smooth out after the initial period of fluctuations attributed to calculation noise. The noise is due to the small number of samples in the mean bucket, which results in erratic calculations of the mean. The graphs smooth out as the sample buckets achieve true mean.

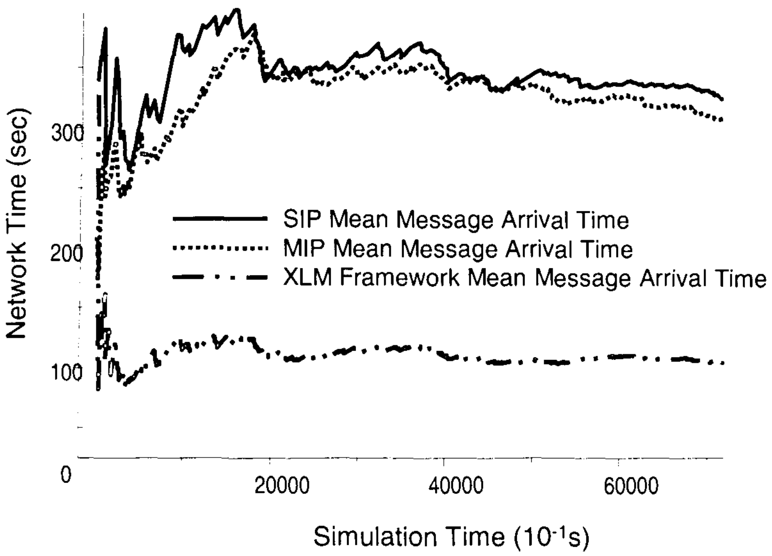

Figure 14. Mean Message Arrival Time for different mobility protocols.

As in the case for MAT, the graph in Fig.5 shows that the XLM has the lowest mean MRT. SIP has the highest mean MRT followed by MIP. As was the case previously, noise exists in the initial periods due to small sample buckets before the graphs settle down to achieve true mean.

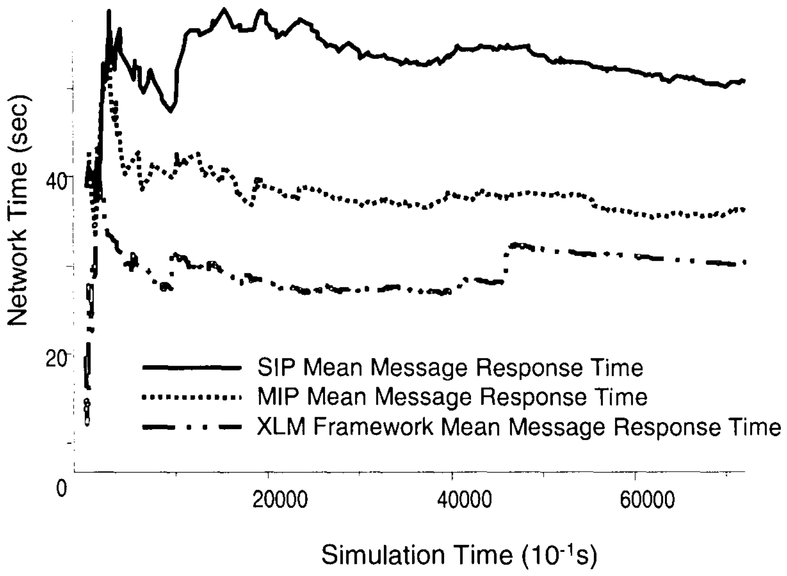

Figure 15. Message Response Time for different mobility protocols.

For both the MAT and MRT, the XLM performs better since, on average, it ensures a lower response and reaction time for both MIH events and commands.

Chapter 5

Chapter 5 Conclusions

From the analysis and experimental results in the previous chapters, we can conclude that the proposed mobility solution outperforms other mobility solutions reviewed in Chapter 2. The solution overcomes shortcomings of the other mobility schemes through the combined use of cross layer design and IEEE802.21. In addition, QoS considerations and network registration, including AAA authentication, ensures scalability and reliability for next-generation protocols.

The proposed solution introduces a framework that houses mobility algorithms and provides all the functionality required for movement in the advanced mobility scenarios presented in Chapter 3. The framework provides generic plug-in points (SAP) for capturing vital content and context information that reflect the current environment which place the algorithms in an ideal position in the mobility process. Furthermore, the IEEE802.21 Media Independent Handover Function provides link and handoff management mechanisms for multimodal devices. Thus, if the decision algorithms are efficient enough, seamless handoff is achievable through MIHS optimized handoffs.

However, the use of the proposed solution does not invalidate SIP and MIP since they are vital for network registration for session and network-layer terminal mobility. This ensures that network registration is scalable for all-IP networks and popular SIP sessions.

Specifically the proposed framework is the ideal MIH user over SIP and MIP since it reduces the detrimental message load on the MIHF. High loads on the MIHF increases the number of unserviced messages that slow down the MIH user's responsiveness to the current environment. The framework dequeues serviced messages quickly from the MIHF since the architecture of the XLM provides a direct pathway to the MIHF.

In addition, the reactivity of the framework to external and internal stimuli is higher than SIP and MIP. The XLM's Link Information Manager (LIM) directly captures external stimuli such as link events on network interfaces. The framework, due to its cross layer design that exposes the SAP to all higher layers simultaneously, reacts to internal stimuli such as changes in the content information including user preferences, application constraints and session demands, quicker than traditional single-layer protocols.

Thus, the framework provides more timely response and reaction to protocol stack events, which would make the framework better suited as an MIH user than SIP and MIP. SIP and MIP respond and react slowly since they suffer from single-layer fallibility where information has to travel layer by layer to reach its target point. [31] and [32] show that seamless handoff is achievable assisted by Media Independent Handover Services, it then follows intuitively that the framework can also achieve seamless handoff if the decision algorithms are efficient. Furthermore, SIP and MIP do not define facilities for advanced decision-making rendering them insufficient for advanced scenario roaming.

The framework features powerful mobility management tools for decision algorithms to orchestrate handoff by targeting the ideal network in next-generation networks. The algorithms can use the suggestions provided in Chapter 3 to mitigate the ping-pong effect that would otherwise lead to poor performance and QoS degradation.

Chapter 7

Chapter 6 Recommendations

The study undertaken draws some recommendations for mobility protocols that carry out next-generation roaming.

- Deployment of advanced mobility decision engines is necessary to cater for user and application-induced handoff in advanced mobility scenarios such as Always-Best-Connected and Multimedia-Aware.
- Mobility management should be seen as a cross function in next-generation protocol stacks as mobility is not a singular layer function. In addition, cross layer schemes enhance the fast capturing of higher layer preferences and parameters.
- SIP and MIP do not define decision engines that allow seamless handoff in advanced mobility scenarios, and should be used primarily as network registration protocols.

References

[1] H. Anthony Chan, "Overview of Wireless Data Network Standards and Their Implementation Issues," Proceedings of South Africa Telecommunication Networks and Applications Conference (SATNAC), Kwazulu-Natal, South Africa, 11-14 September 2005, vol. 1, pp. 117-126.

[2] G.K. Kalebaila and H. A. Chan, "Advanced Mobility Support in Next-Generation All-IP Wireless Networks: A Cross-Layer Approach - Part I", Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium.

[3] J. McNair, Fang Zhu., "Vertical handoffs in fourth-generation multinetwork environments", IEEE Wireless Communications, vol. 11, issue 3, June 2004.

[4] IEEE P802.21/D00.05, "Draft IEEE Standard for Local and Metropolitan Area Networks: Media Independent Handover Services", January 2006.

[5] G. Carneiro, J. Ruela, M. Ricardo, "Cross-layer design in 4G wireless terminals", IEEE Wireless Communications, vol. 11, issue 2, April 2004.

[6] Q. Wang and M. A. Abu-Rgheff, "A Multi-Layer Mobility Management Architecture Using Cross-Layer Signalling Interactions", Proc. IEE 5th European Personal Mobile Communications Conference (EPMCC'03), Glasgow, Scotland, UK, pp. 237-241, Apr 2003.

[7] M.A. Abdelatif, G.K. Kalebaila, and H.A. Chan, "A Comparison between MIPv6 and Cross Layer Movement Detection Mechanisms," Proceedings of Southern African Telecommunication Networks & Applications Conference (SATNAC2006), South Western Cape, South African 4-6 Sept. 2006.

[8] D. Johnson, C. Perkins, and J. Arkko, "Mobility Support in IPv6," Internet proposed standard RFC 3775, June 2004.

[9] G. Tsirtsis, H. Soliman, "Problem Statement: Dual Stack Mobility," Internet proposed standard RFC 4977, August 2007.

[10] M. Liebsch, C. Vogt, "Context Transfer for Proxy MIPv6," Internet proposed draft, July 2007.

[11] H. Chaskar, R. Koodl, "A Framework for QoS Support in Mobile IPv6," Internet proposed draft RFC 2998, Nov. 2000.

[12] J. Rosenberg et. al., "SIP: Session Initiation Protocol," Internet proposed standard RFC 3261, June 2002.

[13] D. Gatzounas, et. al., "Seamless Multimedia Services Over All-IP Based Infrastructures:

The EVOLUTE Approach", IST Mobile and Wireless Telecommunications Summit, Thessaloniki, Greece, 16-19th June 2002.

[14] Y. Xiao, K.K. Leung, Y. Pan, and X. Du, "Architecture, mobility management, and quality of service for integrated 3G and WLAN networks", ACM Wireless Communications & Mobile Computing, vol. 5, issue 7, pp. 805 - 823, November 2005.

[15] I. S. Misra, S. Dey, and D. Saha, "4G All IP Integration Architecture for Next Generation Wireless Internet", NCS 2005.

[16] M. Jaseemuddin, "An Architecture for Integrating UMTS and 802.11 WLAN Networks," Proceedings of the 8th IEEE Symposium on Computers and Communications (ISCC 2003), Turkey, July 2003.

[17] C. Politis, K.A. Chew, and R. Tafazolli, "Multilayer mobility management for all-IP networks: pure SIP vs. hybrid SIP/mobile IP", Vehicular Technology Conference, vol. 4, pp. 2500- 2504, April 2003.

[18] D. S. Nursimloo and H.A. Chan, "Mobility Management, Quality of Service, and Security in the Design of Next Generation Wireless Network, " African Journal of Information and Communication Technology (AJICT), vol. 1 no.1, September 2005, ISSN:1449-2679, pp. 16-22.

[19] P. Bellavista, M. Cinque, D. Cotroneo, and L. Foschini, "Integrated support for handoff management and context awareness in heterogeneous wireless networks", ACM International Conference Proceeding Series, vol. 115, pp. 1 - 8, 2005.

[20] A. Wilson, "Optimising Wireless Access Network Selection to Maintain QoS in Heterogeneous Wireless Environments", WPMC 2005, Aalborg, Denmark.

[21] H.J. Wang, R.H. Katz, and J. Giese, "Policy-enabled handoffs across heterogeneous wireless networks", Mobile Computing Systems and Applications, Proceedings WMCSA 1999, pp. 51-60, February 1999.

[22] F. Siddiqui and S. Zeadally, "Mobility Management across Hybrid Wireless Networks: Trends and Challenges," Computer Communications, Vol. 29, No. 9, Elsevier Science, 2006.

[23] I.F. Akyildiz, Jiang Xie, and S. Mohanty, " A survey of mobility management in next-generation all-IP-based wireless systems", IEEE Wireless Communications, vol. 11, issue 4, pp. 16- 28, August 2004.

[24] A. Grilo, P. Estrela, and M. Nunes, "Terminal independent mobility for IP (TIMIP)," IEEE Communications Magazine, Vol. 39, No. 12, ISSN 0163-6804, pp. 34-41, Dec 2001.

[25] H. Soliman, "Mobile IPv6 support for dual stack Hosts and Routers (DSMIPv6)," Internet proposed draft, July 2007.

[26] S. Gundavelli, et. al., "Proxy Mobile IPv6," Internet proposed draft, January 2007.

[27] T. Ernst, "Network Mobility Support Goals and Requirements," Internet proposed draft, October 2005.

[28] X. Zhang, J.G. Castellanos, and A.T. Campbell, "P-MIP: paging extensions for mobile IP," Mobile Networks and Applications, Vol. 7, No. 2, ISSN 1383-469X, pp. 127-141, 2002.

[29] P. Estrela, T. Vazao, and M. Nunes, "Micro-Mobility Performance Evaluation of a Terminal Independent Mobile Architecture," Second International Working Conference on Performance Evaluation of Heterogeneous Networks, EuroNGI, England, July 2004.

[30] S.J. Koh, M.J. Chang, and M. Lee, "mSCTP for soft Handover in Transport Layer," IEEE Communications Letter, Vol. 8, No. 3, ISSN 1089-7798, pp. 189-191, March 2004.

[31] A. Dutta, Y. Oshba, and H. Shulzrinne, "Seamless Handover Across Heterogeneous Networks – An IEEE802.21 Centric Approach", IEEE WCNC, 2005.

[32] Q.B. Mussabbir and W. Yao, "Optimized FMIPv6 handover using IEEE802.21 MIH services", Proceedings of first ACM/IEEE international workshop on Mobility in the evolving internet architecture, San Francisco, California, pp. 43 - 48, 2006.

[33] R. Koodli, "Fast Handovers for Mobile IPv6," Internet proposed standard RFC 4068, July 2005.

Appendix A: Simulation Files

A.1 Send Function of the MIHF

```
ev << "Message is to be serviced and sent\n";
// The self-message arrived, so we can send check the outgoing messages pipe
// send the appropriate message and remove the message request from the vector
// Check the first message on the pipe, since it has priority (first come first served)
// Check and cast the message to determine it's type
// Send the response message accordingly

// Recieved Internal/Self Message therefore record statistics
decodeDelayDistStats.collect(simTime() - previousArrivalTime);
previousArrivalTime = simTime();

if (strcmp("Link Event Message", (msg)->name()) == 0)
{
//Check and Cast
LinkEventMsg *lemsgTmp = check_and_cast<LinkEventMsg *>(msg);
bubble(lemsgTmp->getEventName());

MIHEventMsg *newMIHEventMsg;
newMIHEventMsg = generateMIHEventMessage(lemsgTmp->getEventNo());

// Get the timestamp of the message and the current simulation time
// Find the difference and write it to the message response time vector
msgResponseTimeVector.record(simTime() - lemsgTmp->getTimestamp());

// send to the XL module
MIHEventMsg *dupMsg = (MIHEventMsg *)newMIHEventMsg->dup();
send(dupMsg, "out", 0);
msgOutCount++;
msgOutCountVector.record(msgOutCount);

delete newMIHEventMsg;
newMIHEventMsg = NULL;

delete lemsgTmp;
lemsgTmp = NULL;

else if (strcmp("MIH Command Message", (msg)->name()) == 0)
{
```

```
//Check and Cast
MIHCommandMsg *MIHcmsgTmp = check_and_cast<MIHCommandMsg *>(msg);

bubble(MIHcmsgTmp->getCommand());

LinkCommandMsg *newLinkCommandMsg;
newLinkCommandMsg = generateLinkCommandMessage();

// Get the timestamp of the message and the current simulation time
// Find the difference and write it to the message response time vector
msgResponseTimeVector.record(simTime() - (MIHcmsgTmp->getTimestamp()));

// send to outgoing link layer gates
int n = 1;
for (int i=n; i<4; i++)
{
LinkCommandMsg *dupMsg = (LinkCommandMsg *)newLinkCommandMsg->dup();
send(dupMsg, "out", i);
msgOutCount++;
msgOutCountVector.record(msgOutCount);
}
delete newLinkCommandMsg;
newLinkCommandMsg = NULL;

delete MIHcmsgTmp;
MIHcmsgTmp = NULL;

}
```

A.2: Algorithm for Receiving Messages at the MIHF

```
if (msg==endServiceMsg) // Departure
{
endService( msgServiced );
ev << "Departure of " << msg->name() << "\n";

if (queue.empty()) // There is no remaining customer
{
msgServiced = NULL;
}
else
{
msgServiced = (cMessage *) queue.pop();
```

```
simtime_t serviceTime = serviceRequirement( msgServiced );
scheduleAt( simTime()+serviceTime, endServiceMsg );
}
}
else if (!msgServiced) // Arrival while server is idle
{
arrival( msg );

ev << "Arrival -while server idle- of " << msg->name() << "'\n";

// Statistics collection
//jobDist->collect (0);
jobsInSys.record(0);

msgServiced = msg;
simtime_t serviceTime = serviceRequirement( msgServiced );
scheduleAt( simTime()+serviceTime, endServiceMsg );
}
else // Arrival while server is busy
{
arrival( msg );

ev << "Arrival -while server busy- of " << msg->name() << "'\n";

// Statistics collection
// There is one customer in service, hence queue.length + 1
//jobDist->collect(queue.length()+1);

jobsInSys.record(queue.length()+1);
queue.insert( msg );
}

//For memory efficient delete the messages
//delete msg;
```

A.3: Encoding Media Independent Handover Events

```
switch (eventNo)
{
case 1:
eventType = "State Change";
eventName = "MIH Link Up";
```

```
lr[0] = lr[1] = true;
direction[0]= true;
direction[1]= true;
direction[2]= false;
break;
case 2:
eventType = "State Change";
eventName = "MIH Link Down";
lr[0] = lr[1] = true;
direction[0]= true;
direction[1]= true;
direction[2]= false;
break;
case 3:
eventType = "Predictive";
eventName = "MIH Link Going Down";
lr[0] = lr[1] = true;
direction[0]= true;
direction[1]= true;
direction[2]= false;
break;
case 4:
eventType = "State Change";
eventName = "MIH Link Detected";
lr[0] = lr[1] = true;
direction[0]= true;
direction[1]= true;
direction[2]= true;
break;
case 5:
eventType = "Link Parameters";
eventName = "MIH Link Parameters Report";
lr[0] = lr[1] = true;
direction[0]= true;
direction[1]= true;
direction[2]= false;
break;
case 6:
eventType = "Administrative";
eventName = "MIH Link Event Rollback";
lr[0] = lr[1] = true;
direction[0]= true;
direction[1]= true;
direction[2]= true;
```

```
break;
case 7:
eventType = "Link Transmission";
eventName = "MIH Link SDU Transmit Status";
lr[0] = true;
lr[1] = false;
direction[0]= false;
direction[1]= false;
direction[2]= false;
break;
case 8:
eventType = "Link Synchronous";
eventName = "MIH Link Handover Imminent";
lr[0] = lr[1] = true;
direction[0]= true;
direction[1]= true;
direction[2]= true;
break;
case 9:
eventType = "Link Synchronous";
eventName = "MIH Link Handover Complete";
lr[0] = lr[1] = true;
direction[0]= true;
direction[1]= true;
direction[2]= true;
break;
default:
eventType = "Corrupt";
eventName = "Corrupt";
lr[0] = lr[1] = false;
direction[0]= false;
direction[1]= false;
direction[2]= false;
break;
}

char msgname[20];
sprintf(msgname, "MIH Event Message");

// Create message object and set source and destination field.
MIHEventMsg *msg = new MIHEventMsg(msgname);
msg->setSource(src);
msg->setDestination(dest);
msg->setTimestamp(simTime());
```

```
msg->setEventNo(eventNo);
msg->setEventType(eventType);
msg->setEventName(eventName);
msg->setLr(0,lr[0]);
msg->setLr(1,lr[1]);
msg->setDirection(0,direction[0]);
msg->setDirection(1,direction[1]);
msg->setDirection(2,direction[2]);
return msg;
```

A.4: Link Command Format

```
message LinkCommandMsg
{
  fields:
     int source;
     int destination;
     simtime_t timestamp;
     int commandNo;
     string command;
}
```

A.5: Type-Value-Length Format

```
message TlvMsg
{
  fields:
     int source;
     int destination;
     string type;
     long length;
     string value;
```

which also contributes to service disruption during the handover. Packet loss is defined as the number of packets that were lost due to the handover of the MN. This factor highly depends on the timing of handover processes and the optimal utilisation of resources such as the packet buffer in the NAR.

Furthermore, the arrival pattern of packets is dependent on many factors, including application characteristics, network queuing behaviours, etc. Hence, packets may arrive at the NAR before the MN is able to establish its link there. These packets will be lost unless they are buffered by the NAR. Similarly, if the MN attaches to the NAR and then sends an FBU message, packets arriving at the PAR until the FBU is processed will be lost unless they are buffered.

This is evident if the handover is executed too early, because the packets start to be redirected to the NAR buffer from the PAR early. This causes buffer overloading and then the packet loss. On the other hand, if the handover is generated too late, the handover process may be in the risk of failure. As our proposed scheme defines the timing of the handover events, it provides optimal utilisation of the NAR buffer. This allows enough time for packet retransmission and reduction of packet loss. It can be shown in **Error! Reference source not found.** that our proposed scheme has been able to reduce packet loss regardless of the handover delay. In each handover case the proposed scheme has been able to reduce the packet loss compared to FMIPv6.

clarify

Figure

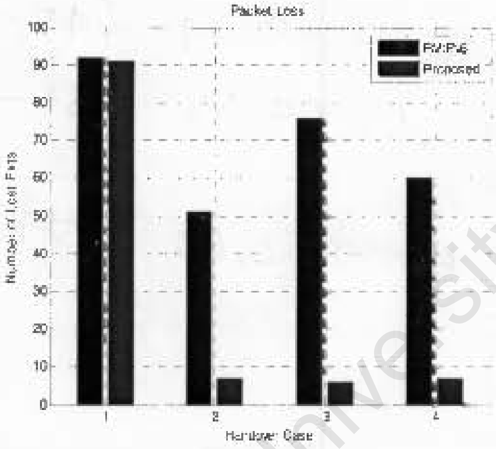

Figure 6: Packet Loss

The time it takes for a packet to move from the source to the destination also plays a vital role in providing uninterrupted service during the handover. In this case we also analysed the performance of our proposed scheme based on the time it takes for a packet to move from the CN to the MN, and how this is affected during the handover period. When calculating packet end-to-end delay each packet is traced from the CN to the MN and then the time it took the packet to reach the MN is recorded. Figure 7 and Figure 8 display the packet delays in FMIPv6 and in the proposed scheme, respectively. The packet delay is the time it takes for the packet to move from CN to MN. As each wired link has the delay time of 0.03s between the node and the distance from the CN to the PAR is equal to two hops the sum of the delay is 0.06s plus some negligible 802.11 latency. Also in the NAR the wired hops introduce the delay of 0.03s which adds up to 0.09s plus the negligible 802.16 latency. This packet delay does not remain constant it change during the handover. The handover affects the packet delay through packet buffering, re-ordering, redirection and TCP re-transmission during or after the handover.

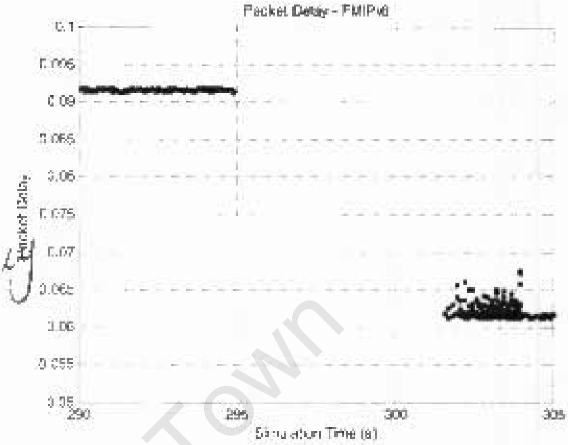

Figure 7: Packet Delay FMIP

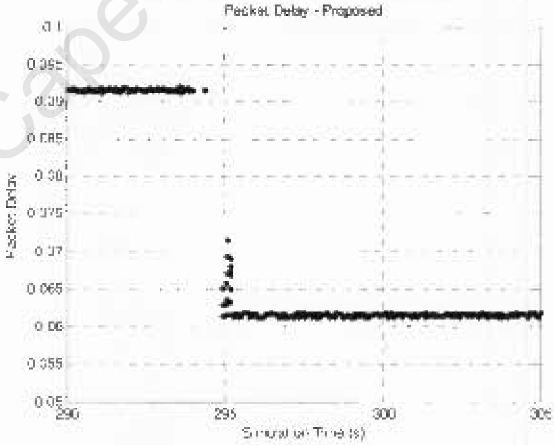

Figure 8: Packet Delay Proposed

The aim of the scheme is to overcome these factors that affect handover. In Figure 7 and Figure 8 the packet end-to-end delay is observed. The average packet delay to reach the MN while it is connected to BS1 is 0.092s. The factors contributing to this delay are defined above. As the MN moves back towards AP1 it experiences a handover at around 295s. The handover occurs earlier in the proposed scheme, but it also finishes earlier. This allows for optimal buffer as the MN receives the buffered data packets as soon as it connects to the NAR. This reduces packet congestion and buffer overload which may lead to packet loss. In FMIPv6, there is much utilisation of the buffer. This causes packet re-ordering and

[13] Inwhee Joe and MinChul Shin "A Mobility-based Prediction Algorithm with Dynamic LGD Triggering for Vertical Handover" IEEE CCNC 2010

[14] The Network Simulator - ns-2, http://www.isi.edu/nsnam/ns/

Sabelo Dlamini (S'09) completed his BScEng (Computer) degree at the University of KwaZulu Natal in 2007. He joined the Communications Research Group at the University of Cape Town as a Telkom Centre of Excellence student where he is currently working towards his MScEng (Elec) degree. His research interests are in Mobility Management in the Next Generation Wireless Networks.

Mqhele Enock-Hershal Dlodlo, (M'88) he is an Assoc. Prof., at the University of Cape Town, He completed his PhD at Delft University in 1996. He researches on wireless personal communication systems and applications. Prof. Dlodlo's international experience spans maintenance engineering, curriculum design, teaching, research, heading departments and faculties between 1980 and 2004. He is a 2003 Fulbright Visiting Scholar alumnus.

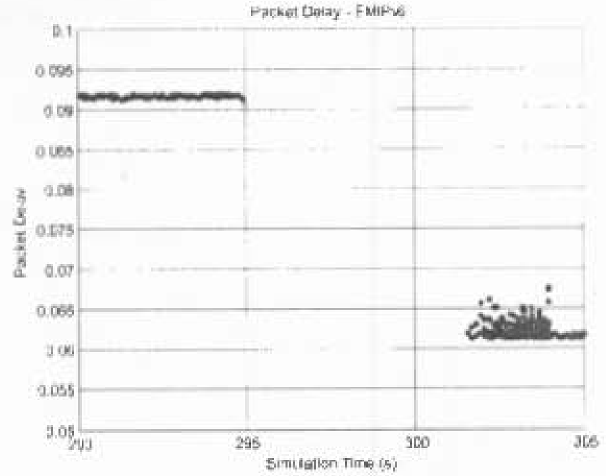

Figure 6: Packet Delay FMIP

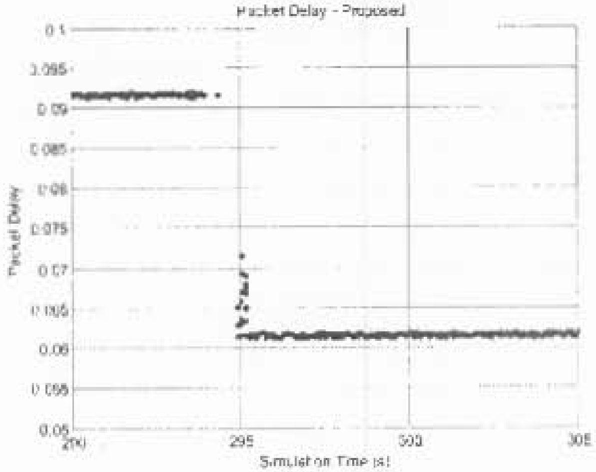

Figure 7: Packet Delay Proposed

The aim of the scheme is to overcome these factors that affect handover. In Figure 6 and Figure 7 the packet end-to-end delay is observed. The average packet delay to reach the MN while it is connected to BS1 is 0.092s. The factors contributing to this delay are defined above. As the MN moves back towards AP1 it experiences a handover at around 295s. The handover occurs earlier in the proposed scheme, but it also finishes earlier. This allows for optimal buffer as the MN receives the buffered data packets as soon as it connects to the NAR. This reduces packet congestion and buffer overload which may lead to packet loss. In FMIPv6, there is much utilisation of the buffer. This causes packet re-ordering and buffer congestion which subsequently lead to packet loss. This even causes the packets that are sent to the MN when it has already connected to the NAR to be delayed as the system will still be dealing with the buffered packets.

The proposed scheme, on the other hand, has been able to overcome the packet delay challenge. It can be noted that during the handover period there are fewer packets that took more than the average packet delay of 0.062s to reach the MN when it is connected to the NAR. This demonstrates the improved handover performance of the proposed scheme during handover. The proposed scheme has been able to reduce the effects of packet buffering, re-ordering, redirection and TCP re-transmission during or after the handover.

VI. CONCLUSION AND FUTURE WORK

The FMIPv6 reduces the long handover latency and high packet loss of MIPv6 by fast movement detection and fast binding update. But it suffers from uncertain additional anticipation time imposed by link layer trigger, especially for delay-constrained real-time traffic such as VoIP. We utilize use predictive link layer trigger with information from the MIH's information services to overcome this challenge.

In this paper we proposed scheme to optimize the performance of FMIPv6 during the handover. The proposed scheme is implemented in ns-2 and the results obtained are shown and discussed. The results have shown that the proposed scheme performs much better than the current FMIPv6 protocol. The proposed scheme achieves much shorter handover latency, lower packet loss and lower end-to-end delay. The proposed scheme has been able to reduce the effects of packet buffering, re-ordering, redirection and TCP re-transmission during or after the handover.

VII. REFERENCES

[1] D. Johnson, C. Perkins and J. Arkko "Mobility Support in IPv6", RFC 3775, June 2004
[2] R. Koodli, "Mobile IPv6 Fast Handovers", IETF RFC 5268, July 2009
[3] Yilin Song, Min Liu, Zhongcheng Li, Qi Li "Handover Latency of predictive FMIPv6 in IEEE 802.11 WLANs: A cross Layer Perspective" International Conference on Computer Communications and Networks, 2009
[4] S. Pyo and Y. Choi, "A Fast Handover Scheme Using Exponential Smoothing Method", February 2009
[5] Xiaoyu Cheng, Duyan Bi "Real-time Adaptive link layer Trigger based Cross Layer Fast Handoff Mechanism in 802.11 WLANs" International Conference on Communication Software and Networks, 2009
[6] Tara A. Yahiya, Hakima Chaouchi "An Optimized Handover Decision for Heterogeneous Wireless Networks" PM2HW2N, October, 2009
[7] Jing Nie, Liaoyuan Zeng, Jiangchuan Wen "A bandwidth Adaptive Fuzzy Logic Handoff in IEEE 802.16 and IEEE 802.11 Hybrid Networks" International Conference on Convergence Information Technology, 2007
[8] S. Haykin and B. Widrow, "Least-Mean-Square Adaptive Filter Theory," John Wiley & Sons, Inc., 2005
[9] IEEE Std 802.21™ , Amendment 1: Security; Part 21:Media Independent Handover Services. 17 Nov. 2009 URL: https://mentor.ieee.org/802.21/dcn/09/21-09-0179-00-0sec-tentative-ieee-21a-document structure-1.doc
[10] Vivek Gupta, David Johnson"A Generalized Model for Link Layer Triggers." Intel Corporation. March 2004
[11] Yong-Sung Kim, et al."Seamless handover support over heterogeneous networks using FMIPv6 with definitive link layer triggers." Springer Science+Business Media B.V. 2007
[12] S. Woon, N. Golmie, Y.A. Sekercioglu, "Effective Link Triggers to Improve Handover Performance", IEEE PIMRC06, pp.1-5, 2006

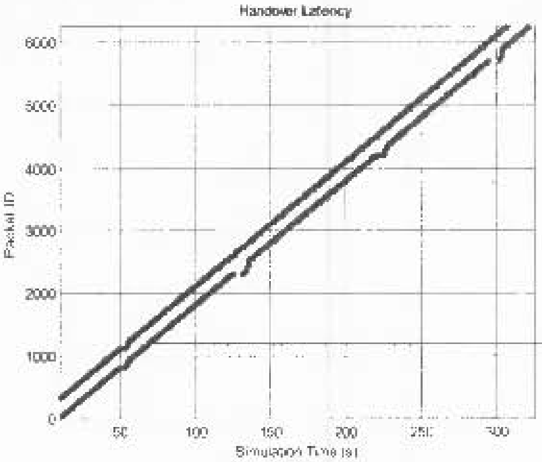

Figure 4: Handover Latency

As can be observed, the gaps in the graphs represent the handover periods and correspond to the handover delays. Thus, around 50s a vertical handover from WiFi to WiMax was experienced and the handover delay encountered in our proposed scheme was about 3.18s while that obtained for FMIPv6 was about 3.21s. In the second gap is the horizontal handover which is discussed below and the other subsequent handovers in the return trip of the MN.

The horizontal handover though was about 0.4s, shown in Figure 5, in our scheme while it was 7.7s for FMIPv6. The reason for longer handover delay in FMIPv6 can be attributed to the fact that it was not well prepared for the handover and that it had to perform the handover procedures sequentially, e.g. scanning which its value varies and contribute much of the time in the handover process. Our scheme on the other hand reduces this processing delay due to preparedness by early prediction and accurate estimation of the time to start the handover process while the MN is still connected to the old network.

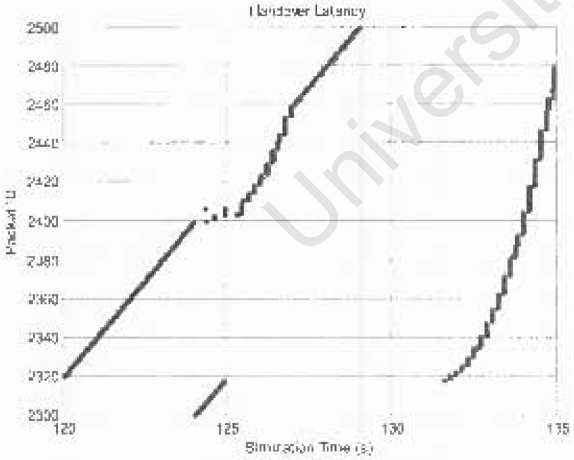

Figure 5: Proposed Scheme Handover Delay

In addition to the time that the MN is not able to send or receive traffic, the handover delay there is also packet loss which also contributes to service disruption during the handover. Packet loss is defined as the number of packets that were lost due to the handover of the MN. This factor highly depends on the timing of handover processes and the optimal utilisation of resources such as the packet buffer in the NAR.

Furthermore, the arrival pattern of packets is dependent on many factors, including application characteristics, network queuing behaviours, etc. Hence, packets may arrive at the NAR before the MN is able to establish its link there. These packets will be lost unless they are buffered by the NAR. Similarly, if the MN attaches to the NAR and then sends an FBU message, packets arriving at the PAR until the FBU is processed will be lost unless they are buffered.

This is evident if the handover is executed too early, because the packets start to be redirected to the NAR buffer from the PAR early. This causes buffer overloading and then the packet loss. On the other hand, if the handover is generated too late, the handover process may be in the risk of failure. As our proposed scheme defines the timing of the handover events, it provides optimal utilisation of the NAR buffer. This allows enough time for packet retransmission and reduction of packet loss. It can be shown in Table 1 that our proposed scheme has been able to reduce packet loss regardless of the handover delay. In each handover case the proposed scheme has been able to reduce the packet loss compared to FMIPv6.

Table 1: Packet Loss

	Packet Loss			
Handover	1	2	3	4
Proposed	91	7	6	7
FMIPv6	92	51	76	60

The time it takes for a packet to move from the source to the destination also plays a vital role in providing uninterrupted service during the handover. In this case we also analysed the performance of our proposed scheme based on the time it takes for a packet to move from the CN to the MN, and how this is affected during the handover period. When calculating packet end-to-end delay each packet is traced from the CN to the MN and then the time it took the packet to reach the MN is recorded. Figure 6 and Figure 7 display the packet delays in FMIPv6 and in the proposed scheme, respectively. The packet delay is the time it takes for the packet to move from CN to MN. As each wired link has the delay time of 0.03s between the node and the distance from the CN to the PAR is equal to two hops the sum of the delay is 0.06s plus some negligible 802.11 latency. Also in the NAR the wired hops introduce the delay of 0.03s which adds up to 0.09s plus the negligible 802.16 latency. This packet delay does not remain constant it change during the handover. The handover affects the packet delay through packet buffering, re-ordering, redirection and TCP re-transmission during or after the handover.

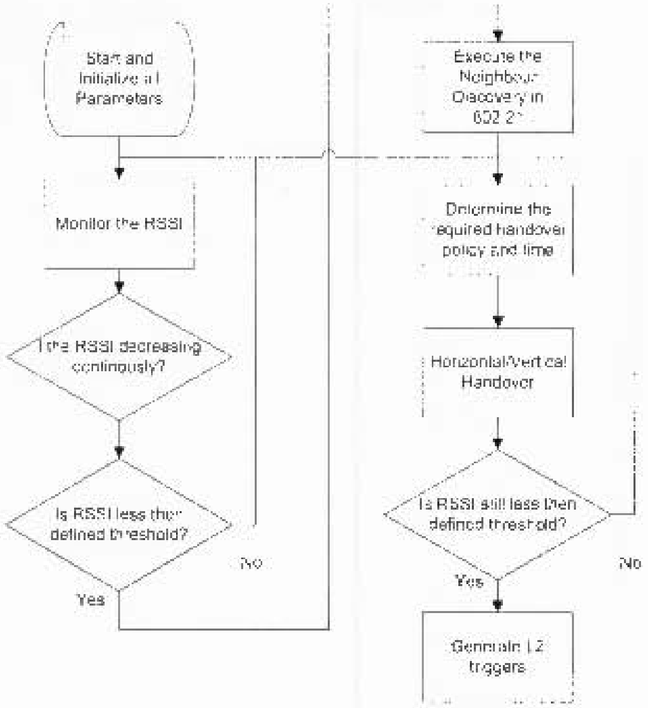

Figure 2: Proposed Scheme Flow Chart

As can be seen from Figure 2 above, after initialization, the RSS Monitoring and Prediction continually observes the RSS with respect to dynamically set threshold. The threshold is dynamically set based on the current network condition. For example, if the conditions are poor, the threshold is correspondingly and proportionally increased. Likewise, when the conditions are good, the threshold is correspondingly and proportionally reduced. In effect, the RSS Monitoring and Prediction module, predicts the next set threshold on the current network conditions or RSS.

The outcome of the RSS Monitoring and Prediction is then passed on to the Handover Trigger module. The Handover Trigger module also gets network-related information from the Neighbour Discovery module. Based on this information, the Handover Trigger module determines the most suitable network to handover to, i.e. whether is horizontal or vertical handover, as well as the expected time the handover will take. The Handover Trigger module constantly checks the inputs from the Neighbour Discovery and RSS Monitoring and Prediction modules to ensure that it is always up-to-date in the estimation it makes, otherwise a rollback to the handover events is also possible.

After the estimation of the time required for the handover as well as the type of handover required, it sends a command for handover initiation to the Handover Execution module in a timely manner to ensure that the handover starts and finishes as per estimation, hence ensuring minimal handover and packet loss.

The implementation of the cross-layer scheme to optimize the performance of the FMIPv6 has been outlined. Also the details of each component and its relation to other modules in terms of the operation to reduce the handover delay in FMIPv6 handover.

V. SIMULATION RESULTS

In this section the performance of the proposed scheme is evaluated through simulations. The handover delay, packet loss and the end-to-end packet delay is analyzed for the proposed scheme. The proposed scheme is compared with FMIPv6.

Figure 3 shows the simulation setup, which is set in area of 3000 by 3000 meters in the ns-2 simulator platform with the NIST mobility module. In the simulation the TCP traffic is used to evaluate the handover delay and packet loss, and the UDP traffic is used to evaluate the end-to-end packet delay. The packet intervals are fixed to 0.05s for all simulations. The wired link between the router and the gateway are 100Mb with link delays of 30ms. The MN moves linearly from AP1 towards BS2 past BS1 at a constant speed of 1m/s, effectively experiencing two handover; AP1-to-BS1 vertical and BS1-to-BS2 horizontal handover, and on its return trip it then makes four handovers. These handovers are shown in Figure 4.

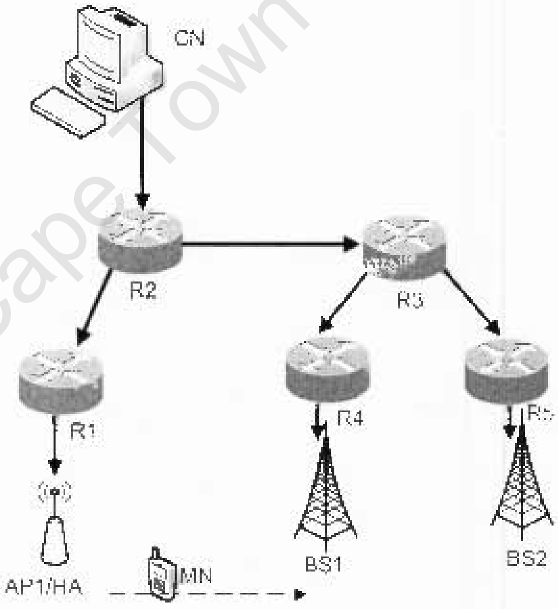

Figure 3: Simulation Setup

Figure 4 depicts the handover delay performance obtained when our proposed scheme is incorporated in FMIPv6 and when it is not. For clarity in the diagram, we shifted the graph obtained by our proposed scheme upwards in the Packet ID scale, to avoid the graphs from appearing on top of each other.

efficient manner. The MIH was designed to be such standard that can be used across different heterogeneous networks to provide seamless handovers. In this paper MIH is going to be adopted for heterogeneous network discovery and selection.

The link layer triggers are used to communicate link layer events to the network layer and the layers above. Link layer events include the anticipation and execution of a host association and disassociation with the current link. The link layer triggers have been well defined in [9] and [10]. There are numerous schemes that have been proposed to provide more definitive L2 triggers to reduce handover delays in FMIPv6. References [11] and [12] use the MIH services to provide timely L2 triggers.

The work in reference [13], also MIH Information Service to obtain the information that helps it predict the target cell for the MN. They use this information to advance the Link Going Down trigger in order to prepare for the handover in advance. The issue with this work is that it uses GPS based information which can be resourceful to an MN compared to our proposed scheme. In addition to this after generating the Link Going Down trigger in advance they do not take into consideration the timing impact that may cause handover to be generated to early.

III. MOTIVATION

The goal of this work is to propose a predictive handover scheme. This scheme will use the IEEE 802.21 for network discovery and the most optimal prediction algorithm to timely generate the L2 triggers. This will be used to ensure that all the L2 triggers are executed at the right time and increase the probability of FMIPv6 to be executed in a proactive mode. This will be done by estimating the required handover time for given neighbour network conditions. Then using the predictive L2 triggers the handover will be started at a time that is dynamically calculated to minimize the handover latency.

IV. PROPOSED SCHEME

This section outlines the operations of the proposed scheme. The first part defines the implementation of the proposed scheme in the MN. The second part previews the operation of the scheme in facilitating the handover process.

Our proposed scheme employs the cross-layer approach to enhance FMIPv6 handover performance. Figure 1 below illustrates the architectural framework of the proposed scheme. The basic idea of the scheme involves utilization of both link layer and network layer information to facilitate the handover process.

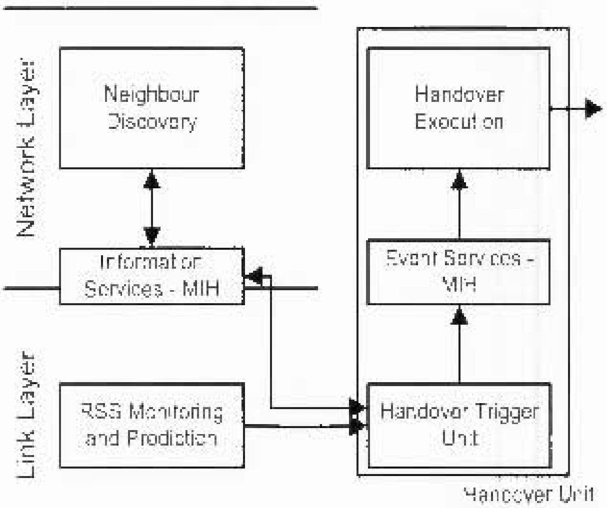

Figure 1: System Architecture

As can be observed in Figure 1, the proposed scheme consists of 4 cooperating modules to improve handover performance. The modules' functions are briefly discussed next.

The Neighbour Discovery module collects and stores both dynamic and static network layer information. The information include e.g. neighbouring PoAs addresses, authentication information etc. This module utilizes the MIIS functionality of the MIH services to achieve its aim, which involves providing information that will facilitate network selection.

The Handover Trigger module gets the information from the Neighbour Discovery module and processes it based on defined prediction algorithms and thresholds, which ensure timely triggering of the handover process. Since network conditions are very dynamic, the operations of the Handover Trigger module also rely on the latest link layer information it gets from the RSS Monitoring and Prediction module to make timely and well-informed handover triggering decisions. Based on the information the Handover Trigger has obtained from the Neighbour Discovery and RSS Monitoring and Prediction module, it estimates the delay associated with either horizontal or vertical handover. In effect, the Handover Trigger module is able to estimate the appropriate time to start the handover process. We implement an LMS prediction algorithm to ensure that the Handover Trigger efficiently performs its function.

The RSS Monitoring and Prediction, on the other hand, concentrates on obtaining dynamic link layer information that has an impact on handover decisions, in particular the dynamic RSS. It observes the behavior of the RSS with respect to same dynamically set threshold and sends the output to the Handover Trigger module as required.

Finally, the Handover Execution module, executes the command it receives from the Handover Trigger module. In fact, the Handover Execution module implements FMIPv6, which is triggered to timely start the handover procedures at the handover initiation time estimated or predicted by the handover triggers module. Figure 2 below shows a flow chart that illustrates the basic operation principle of our proposed scheme as explained above.

Improvement of FMIPv6 handover performance using MIH and timely link layer triggers

Sabelo Dlamini, *MIEEE*; Mqhele Enock-Hershal Dlodlo, *MIEEE*
Department of Electrical Engineering
University of Cape Town, Private Bag, Rondebosch, 7701
Tel: +27 650-2813, Fax: +27 650-3465
email: dlaminis@crg.ee.uct.ac.za; mqhele.dlodlo@uct.ac.za

Abstract- The Mobile IPv6 for Fast Handovers (FMIPv6) solution aims at reducing the handover latency and packet loss experienced in MIPv6. It achieves this reduction by applying fast movement detection and fast binding update procedures. However, handover latency in this FMIPv6 solution is still not sufficient for active ongoing real-time and time-sensitive applications. In fact, this latency results in packet loss hence cause service disruption during the handover period. We propose to address this drawback by using predictive link layer triggers in conjunction with IEEE 802.21 Media Independent Handover (MIH) services to facilitate an enhanced FMIPv6 handover. The obtained simulation results show that the proposed scheme enhances handover performance in terms handover latency and packet end-to-end delay.

Index Terms— Handover latency, link-layer triggers; Least Mean Square; MIH

I. BACKGROUND

Among many proposed mobility management solutions, Mobile IPv6 (MIPv6) [1] has been proposed as the standard to solve the problem of mobility. It does this by redirecting packets for the mobile node (MN) to its current location. In MIPv6 the period during which the MN looses connectivity with its current link until the time it receives the first packet after connecting to the new link is known as handover latency. The overall handover latency in MIPv6 consists of Layer 2 (L2) handover latency and Layer 3 (L3) handover latency. L2 handover latency is the period when the MN is disconnected from the air-link to the current Access Router (AR) until the time it connects to the air-link of the new AR [1]. In L3 handover, there are latencies incurred due to the processes of movement detection, Care-of-Address (CoA) configuration and Binding Updates (BU). The handover latency incurred by MIPv6 is intolerable for time sensitive and real-time traffic [2], since the MN is not able to send or receive traffic during this interval.

Various protocols have been proposed to optimize handover latency in MIPv6 e.g. Fast Handovers in MIPv6 (FMIPv6) [2] being one of them. FMIPv6 protocol has been designed to reduce handover delays incurred due to movement detection, Care-of-Address (CoA) acquisition and binding update (BU) events. This is done with the aid of anticipation based Layer 2 (L2) trigger information as well as by obtaining the subnet prefix information from the New Access Router (nAR) while the MN is still connected to its current/old Access Router (oAR). In order to form a new CoA, FMIPv6 relies on the oAR to resolve the network prefix of the nAR based on the L2 identifier reported from the MN.

The anticipation mechanism specified by FMIPv6 suffers from the problem of timing hence it may cause the handover process to start earlier or later than the actual handover. This reduces the certainty about the MN's movement. Also sudden degradation of the wireless link during the handover initiation phase may cause the MN to lose connectivity with the oAR. In this case, if the handover anticipation time is large, then the MN may not have sufficient time for new CoA (NCoA) configuration while being attached to the oAR's link. Consequently, there would be long handover latencies.

II. RELATED WORK

Timely execution of handover decision plays a vital role in handovers, particularly in heterogeneous networking environments. Various handover solutions have been proposed to reduce handover delays, but there are still issues that need to be addressed in this area. Song et al. in [3] identifies the issues that are affecting the handover latency in the predictive mode of FMIPv6. It shows that the ambiguous link layer triggering timing, the lack of assistance from the network layer entities, and the inefficient interaction between the link layer and the network layer are the primary causes of large handover latency.

The work by Pyo et al. in [4] proposes a prediction algorithm to counteract the problem of timely link layer triggering. Pyo et al. uses Fast Exponential Smoothing, while Xiaoyu in [5] use FFT the prediction algorithm to predict the decay of the received signal strength. Different prediction schemes have been used as prediction algorithms such as Neyman-Pearson in [6] and artificial intelligence in [7]. In this paper we are going to use Least Mean Square which is defined in [8]. However, any other prediction algorithm can be used to provide the RSS prediction capabilities in the scheme.

As it is always desirable to have a standardized method to handle mobility across heterogeneous networks in a

UNIVERSITY OF CAPE TOWN
FACULTY OF ENGINEERING & THE BUILT ENVIRONMENT

Submission of a Paper in partial fulfillment of the requirements for a Masters Degree

Section A - To be completed by the Student

I attach a copy of the Paper I am submitting in terms of the Masters Degree Rules.

Name of Student:	SABELO DLAMINI		
Student No:	DLMSAB003	**Dept in which registered**	ELECTRICAL
Title of Paper:	IMPROVEMENT OF FMIPV6 HANDOVER PERFOMANCE		
	USING MIH AND TIMELY LINK LAYER TRIGGERS		

For the Faculty's information, please indicate if the Paper has already been (i) submitted to a Journal or a Conference or (ii) published or accepted for publication. Please give details attaching evidence (reprint / photocopy of paper / letter of acceptance) in the latter case. (Please note, however, that (i) and (ii) are not requirements.)

South African Telecommunication Networks and Applications Conference
Spier Estate - 2010 September

Title of Dissertation (***if different from above***)**:**	OPTIMISATION OF FMIPV6 USING
TIMELY PREDICTIVE LINK LAYER TRIGGERS AND IEEE 802-21 MIH	
SERVICE	

Noted by Supervisor:		**Date:**	
Supervisor's Comments:			

Section B - To be completed by the Panel of Assessors (FOR OFFICE USE)

☐ The Paper **meets** the standard of being potentially publishable.

☐ The Paper **does not meet** the standard of being potentially publishable.

Comments:			
Convenor's Signature		**Date:**	