Bansal, S, 2006.
Framework to Facilitate Smooth
Handovers Between Mobile IPv6 Networks.
Available from Middlesex University's Research Repository.

Framework to facilitate smooth handovers between Mobile IPv6 based networks

Thesis submitted to Middlesex University in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Saurabh Bansal, BEng

Middlesex University
School of Computing Science
London, UK

2006

To my family

Abstract

Fourth generation (4G) mobile communication networks are characterised by heterogeneous access networks and IP based transport technologies. Different access technologies give users choices to select services such as levels of Quality of Service (QoS) support, business models and service providers. Flexibility of heterogeneous access is compounded by the overhead of scanning to discover accessible services, which added to the handoff latency. This thesis has developed mechanisms for service discovery and service selection, along with a novel proposal for mobility management architectures that reduced handoff latency.

The service discovery framework included a service advertisement data repository and a single frequency band access mechanism, which enabled users to explore services offered by various operators with a reduced scanning overhead. The novel hierarchical layout of the repository enabled it to categorise information into various layers and facilitate location based information retrieval. The information made available by the repository included cost, bandwidth, Packet Loss (PL), latency, jitter, Bit Error Rate (BER), location and service connectivity information. The single frequency band access mechanism further enabled users to explore service advertisements in the absence of their main service providers. The single frequency access mechanism broadcasted service advertisements information piggybacked onto a router advertisement packet on a reserved frequency band for advertisements. Results indicated that scanning 13 channels on 802.11b interface takes 189ms whereas executing a query with maximum permissible search parameters on the service advertisement data repository takes 67ms.

A service selection algorithm was developed to make handoff decisions utilising the service advertisements acquired from the service discovery framework; based on a user's preference. The selection algorithm reduced the calculation overhead by eliminating unsuitable networks; based on interface compatibility, service provider location, unacceptable QoS (Quality of service) and unacceptable cost; from the selection process. The selection algorithm utilised cost, bandwidth, PL, latency, jitter, BER and terminal

power for computing the most suitable network. Results indicated that the elimination based approach has improved the performance of the algorithm by 35% over non-elimination oriented selection procedures, even after utilising more selection parameters.

The service discovery framework and the service selection algorithm are flexible enough to be employed in most mobility management architectures. The thesis recommends Seamless Mobile Internet Protocol (SMIP) as a mobility management scheme based on the simulation results. The SMIP protocol, a combination of Hierarchical Mobile Internet Protocol (HMIP) and Fast Mobile Internet Protocol (FMIP), suffered handoff latency increases when undergoing a global handoff due to HMIP. The proposed modification to the HMIP included the introduction of a coverage area overlap, to reduce the global handoff latency. The introduction of a Home Address (HA) in Wireless Local Area Networks (WLAN) binding table enabled seamless handoffs from WLANs by having a redirection mechanism for the user's packets after handoff.

The thesis delivered a new mobility management architecture with mechanisms for service discovery and service selection. The proposed framework enabled user oriented, application centric and terminal based approach for selecting IPv6 networks.

Acknowledgments

First and foremost, I would like to express my gratitude to my supervisor, Dr A. Lasebae, for his support and guidance throughout the course of my association with Middlesex University. Above all, I would like to thank him for his help and guidance throughout the course of this project.

I would also like to thank my second supervisor, Prof. R. Comley, for his ongoing support and guidance.

In the journey to achieve the project objectives there were many critical issues which needed expert guidance and in-depth knowledge. Mr. Kenneth Chin, Mr. Christopher Kindberg, Dr Orhan Gemikonakli, Dr Robert Colson, Mr. Richard Gledhill, Dr Robert Ettinger, Mr. Nick Gerogiannis and many other professors at Middlesex University extended their help and support when it mattered most. I am truly thankful to them.

I would also like to thank the research administration staff at Middlesex University. Special thanks to Prof. Colin Tully, Emma Warne, Kerry Lane and Dr Robert Pleas.

In addition, I would like to take this opportunity to say thank you to all the wonderful people in the 'Engineering Systems Laboratory' who have been very helpful and patient with me. Special thanks to Mr. Michael Robinson, Mr. Goodwyn Griffiths, Mr. Gerald Osei-kofi and Mr. Neil Matticks for helping me on numerous occasions.

I would also like to take this opportunity to thank my family for their continued support throughout the project and my life. My parents, my sister and brother-in-law stood by me when it mattered most. Also, a special thanks to my niece for giving me a reason to smile when everything else looked miserable.

My uncle Sunil, aunt Sunita and cousin Savil, my local friends, philosophers and guides need a special mention. They have always seen to it that I do not feel homesick.

Last but not least all my friends have been wonderful throughout my time as a research student. Special thanks to Birinder Sandhawalia, Gurteg Sandhu, Abhishek Agrawal, Nilton Aranda, Jean Marie Clive Mamodesen, Saman Rasheed, Amit Bhatnagar, Kostas Vlachakis, Kostas Dovas, Kamran Younis, Barjinder Singh, Rakesh Tiwari, Nalini Selvaraj, Rachid, Andrew Georgiades, Saif, Fatima Sheikh, Agozie and John Salisbury.

Very special thanks to all my fellow PhD students for helping me whenever I needed it.

I also thank anyone who has helped me directly or indirectly to complete this project.

Finally, I wish to thank Middlesex University for the award of a research studentship, without which it would not have been possible for me to undertake this period of postgraduate study.

Glossary of terms

4G	Fourth Generation
BAck	Binding Acknowledgement
BAN	Basic Access Network
BER	Bit Error Rate
BS	Base Station
BU	Binding Update
CCN	Common Core Network
CN	Correspondent Node
CoA	Care of Address
F-BAck	Fast Binding acknowledgement
F-BU	Fast Binding update
FMIP	Fast Mobile Internet Protocol
FN	Foreign Network
F-NA	Fast Neighbour advertisement
GPS	Global Positioning Satellite
HA	Home Agent
HAck	Handover Acknowledgement
HI	Handover Initiation
HMIP	Hierarchical Mobile Internet Protocol
IP	Internet Protocol
L2	Layer 2
LCoA	Local Care of Address
MAP	Mobility Anchor Point
MIIS	Media Independent Information Server
MIP	Mobile Internet Protocol
MIPv6	Mobile Internet Protocol version 6
MIRAI	Multimedia Integrated network by Radio Access Innovation

MN	Mobile Node
MOSSA	Multiple Operator Service Accessibility Architecture
PL	Packet Loss
PoA	Point of Attachment
PrRtAdv	Proxy Router Advertisement
QoS	Quality of Service
RAN	Radio Access Network
RCoA	Regional Care of Address
RMIP	Regional Mobile Internet Protocol
RtSolPr	Router Solicitation for Proxy
SC	Switching Centre
SDA	Service Discovery Agent
SMIP	Seamless Mobile Internet Protocol
UI	Unit Interval
UIS	Universal Information Service
UIS	Universal Information Service
WLAN	Wireless Local Area Network

Table of Contents

Index of Figures

Index of Tables

Chapter 1 Introduction

"There is nothing like looking, if you want to find something. You certainly usually find something, if you look, but it is not always quite the something you were after."

J.R.R. Tolkien

In recent years, the world has witnessed an explosive growth in wireless and mobile communications. This growing industry coupled with the ever changing requirements has redefined the whole perception of wireless communication. The increasing number of users and the demand for different types of services have added constraints on network designs. Two decades ago, good quality real-time voice telephony was deemed more than sufficient for mobile communication use, but with the introduction of data services the user's outlook towards wireless communication has changed. Increasing user requirements for email on the move, multimedia, high speed data and good quality voice communications has led to the introduction of different types of wireless networks ranging from high speed data to blackberry devices for email on the move. The introduction of 3G networks was viewed by many as the answer to user requirements that led companies to invest billions of pounds in licensing and infrastructure costs [Losi, 2001]. 3G networks have not even been rolled out completely and already it is believed that they will not be able to cope with user's expectations and requirements. Researchers have already started planning for next generation (Fourth Generation) mobile and wireless networks that are deemed to be necessary in order to cope with users requirements [Wisely, Matjina, 2003].

1.1 Fourth Generation (4G)

"Fourth Generation is not yet a global standard but more of a concept that will enable devices to switch between different types of networks"

[Mckay, 2002]

There have been several conflicting theories about what actually the next generation for cellular and wireless networks is, but in principle every theory has IP in common [Wisely, 2001]. All the different wireless standards are expected to inter-work to provide users with a wide range of services at a higher grade in the fourth generation systems [Ojanperä, 2000]. Real-time services with quality comparable to traditional cellular networks should be generally accessible regardless of the technology, the access network and uninterrupted handover. Wireless Local Area Networks (WLAN) could be a cost effective complement to 3G systems in urban blanket deployments for multimedia applications [Birchler, 2003] and provide users high speed access to the Internet in hotspot environments. Hence, 4G networks would enable heterogeneous access over an IP based transport technology.

Design goals for 4G were –

- Unifying all IP architectures evolving from previous generations – IP has been widely accepted as the integration platform for the various wireless and wired technologies for 4G. 4G networks will have to evolve from previous generation of mobility to reduce the cost of the infrastructure.
- Support for multiple-operators and multiple interfaces – 4G's heterogeneous access enabled users to switch networks to look for better services but such heterogeneity poses challenges for the mobility management architecture [Montavont, et al., 2004]. The heterogeneous access does not limit the user's capability to use services over various types of interfaces; hence the mobility management architecture should be flexible enough to deal with multiple operators offering services over different types of interfaces [Montavont, et al., 2004].

- Mobility enabled end-to-end QoS architecture – The heterogeneous access across networks should not be at the expense of the user's service quality. The mobility management architecture should be competent enough to maintain service quality for entire service usage duration.
- Competent auditing, accounting, authorisation and charging (AAAC) mechanisms – Since users would be able to move across networks, AAAC mechanisms should be able to cope with it.

"4G wireless networks support global roaming across multiple wireless and mobile networks" [Varshney, Jain, 2001]

4G networks would support global roaming across multiple wireless and cellular networks, e.g. users free to switch from a cellular network to a satellite based network or to a high bandwidth wireless LAN. Multi-operator support would enable users to access different services, increased coverage and convenience of a single device, one bill with reduced total cost, and more reliable wireless access even with the failure or loss of one or more networks.

1.1.1 Visions for networks beyond 3G

Vision for the next generation of mobile telephony -

- "Design for Choice" – Users should be able to make choices (accessing several services and selecting between them) and express preferences.
- "Design for Control" – Users should be able to control (and/or negotiate) Service Level Agreements (SLA), hence being able to control QoS/cost.
- "Design for Expandability"- Architecture should enable ease of service setup and delivery, by network operators and third party service providers.
- "Design for Freedom"- The network should be able to deliver a diverse range of applications to enable "Office in the hand".
- "Design for Security"- Users should feel secure using the services.
- "Design for seamless operation" – Users should be able to switch between access networks without compromising the quality of service during the switch. Users

should be able to switch between company intranets to the wider domain with ease.

- "Design for Speed" – Users should be able to enjoy high bandwidth services with good QoS.

1.1.2 Multiple-Operators

'Multiple operators' is not an easy term to define. It is possible that multiple operators refer to the ability of users to switch networks to enjoy better services but in theory users could simultaneously enjoy services from more than one network. For example, users could enjoy high bandwidth data service in a wireless hotspot whilst enjoying real time voice communication from a cellular provider. 3G networks and networks beyond 3G are designed to cope with this new definition of multiple operators if spectrum trading becomes a reality. In this thesis, the term multiple operators refers to –

- Users ability to switch networks to seek better services
- Users ability to enjoy services from multiple operators simultaneously

The attempt to redefine multiple operators would lead to constraints in network design which is beyond the scope of this research thesis.

1.1.3 Multiple Interfaces

Future mobile networks support multiple operators offering service on different wireless networks. The advancing wireless terminal technology has led to the development of devices with more than one wireless interface [Montavont, et al., 2004]. The future wireless interfaces would be able to support multiple frequencies from a single interface [Craninckx, Donnay, 2003]. The terminal's capability to support multiple interfaces and the interface's capability to support multiple frequencies expand the user's choice of service providers but puts extra strain on the mobility management architecture.

1.2 Issues with multiple operator multiple interface scenarios

1.2.1 Scanning overhead

In 2G and 3G networks, handover measurements are controlled by the network. In the UMTS system, prior to handover decisions, neighbouring cell parameters, such as frequencies and scrambling codes are sent to the mobile node by the UTRAN/RNC. In WLAN 802.11b, however, wireless nodes have to scan all the 13 channels to find the corresponding access point, which can take up to 400ms [Zhang, et al., 2003]. Large scale scanning adds to the Layer 2 handovers. Since, future wireless terminals would be able to support more than one interface; the device will have to scan for all the usable frequencies available to that device. Large scale scanning would add to delays in call setup and handovers. Even though the scanning adds to the delays, the user has to go through the process to make a handover decision and as a consequence, drains the terminal's resources.

Research question 1 – How to deliver service advertisements to users with or without reduced scanning overhead?

Research Question 2 - Could the delivered advertisements have operator transparency?

Research Question 3 – Could the advertisements be delivered to the user, even in the absence of its main service provider?

1.2.2 Cross-Network Mobility

Mobile terminals in 2G and 3G networks send a measurement report to the network to enable the network to make handover decisions for the mobile node [Zhang, et al., 2003]. Mobile assisted handovers work well in a homogeneous network within an administrative domain, whereas in heterogeneous networks they have some disadvantages. Inter-domain handovers – Handovers within a domain do not require any security considerations whilst if the handover is cross-domain the node would have to be authenticated in the new domain. The service profile of mobile nodes will vary from one domain to the other whereas it stays the same within a single domain. Accounting and authorization information would have to be refreshed every time the node is handed over to a new administrative domain, which adds to the handover latency. Network controlled handovers would prevent users from exercising their right of choice by finding an alternative operator other than the one they are subscribed to. To enable heterogeneous access across networks, only mobile-controlled handovers should be considered [Zhang, et al., 2003].

Research Question 4 – Is it possible to reduce the cross network handover latency by pre-registering the mobile in neighbouring networks?

1.2.3 Selecting services based on user preferences

Handovers in 4G would not only be carried out to maintain connection but also to provide better services and to meet individual requirements [Varshney, 2001]. The mobile terminal should be able to make the handover decision based on pre-selected user preferences with the option of allowing users to manually change the settings for a specific channel. Handovers can be categorised as imperative and alternative [Damberg, et al., 2002]. Imperative handovers are a priority because their selection and execution have to be performed as quickly as possible in order to maintain connection of on-going channels. Alternative handovers take place to provide users with better performance or to meet a particular preference. Alternative handovers can therefore tolerate longer handover latency as compared with imperative handovers.

Users gathering service advertisements would have to decide on the operator for the desired service, which would be an extremely difficult task for the average user. Moreover, since handover decisions have to be made quickly to maintain call quality, the selection procedure has to be quick and effective.

Considerations for the selection procedure –

- The selection algorithm should be easy to use for an average user.
- The selection algorithm should be quick and effective to maintain call quality during handovers.

- The selection algorithm should not add to call setup delay.
- The selection procedure should take imperative handover situations into consideration.
- Users should be able to specify the characteristics for the desired service.
- The selection algorithm should take into account the services already in use by the consumer whilst selecting the operator for the desired service based upon user preferences.
- The selection procedure should account for interface incompatibility on the device. Even though a device may have more than one connection interface, they all might not be compatible with each other all the time.
- The selection procedure should also take into consideration the Service Level Agreements (SLA) between users and operators.

Research Question 5 – Is it possible to design an algorithm which takes into consideration most of the user desirable characteristics, when making a handover decision?

Research Question 6 – What are the service characteristics that users might consider during service selection?

1.3 Previous Work

Fourth generation is not yet a global standard but more of concepts that would enable devices to switch between different types of networks [Mckay, 2002]. Though, there are many conflicting views about 4G, e.g. the Japanese view of the network is based on orthogonal frequency division multiplexing (OFDM) offering data rates up to 100Mbits/sec [Wisely, et al., 2001], it is widely believed that IP will be the means to integrate access networks for all technologies [Clapton, 2001]. The switch from traditional circuit switched technology to packet switched provided users greater flexibility but complicated network and mobile terminal designs [Zhang, et al., 2003].

Mobile IP was designed to provide mobile devices the freedom to move from one network to another while maintaining their permanent IP address. Mobile IP is a modification of the IP that allows nodes to continue to receive datagrams (i.e. UDP packets) as long as they remain connected. It involves some additional control messages that allow the IP nodes to manage their IP routing tables reliably [Perkins, 1998]. Mobile IP deploys a Home Agent (HA) which intercepts packets bound for the Mobile Node (MN) and tunnels them to its current point of attachment to the network via a foreign agent. Tunnelling (or triangle routing) ensure that MNs continues to receive their packets once the HA is aware of their current point of attachment. MNs notify their HA when changing in between networks but this indirection requires a registration process and address resolution procedure. The indirection procedure has been shown to result in long handover latencies [Hseih, et al., 2002], which often leads to packet loss and performance degradation.

Many variants of Mobile IP have been developed to counter the problems faced by Mobile IP and to integrate access networks of different technologies. Of the many variants Hierarchical Mobile IP (HMIP) and Fast Mobile IP (FMIP) have attracted a lot of attention from the research community [Hseih, et al., 2001]. HMIP [Castelluccia, 1998] provides a scheme for performing registrations locally in the foreign network, thereby reducing the number of signalling messages forwarded to the home network as well as lowering the signalling latency that occurs when a mobile node moves from one Base Station (BS) to another. HMIP introduces a conceptual entity, Mobility Anchor Point (MAP), to separate local mobility from global mobility. MAP acts as a Mobile Node's (MN) Home Agent (HA) in the foreign network. MAP intercepts all packets destined for MN and tunnels them to its on-link CoA (LCoA). In HMIP, MN attaches itself to two Care of Addresses (CoA), a regional (or site) address and an on-link (or point of attachment) address. Figure 1-4 shows a typical HMIP network architecture.

Fast Mobile IP (FMIP) [Dommety, et al., 2002] on the other hand reduced the address resolution latency by address pre-configuration. FMIP did not define the architecture of the network but introduced additional signalling messages. Seamless Mobile IP (S-MIP) builds on fast and hierarchical mobile-IP schemes. S-MIP exploits the fact that Hierarchical Mobile IP has reduced registration latency and Fast Mobile IP reduced address resolution time where the combination has proved to reduce latency [Hseih, et al., 2002]. S-MIP handoff strategy is terminal initiated but network determined [Hseih, et al., 2003].

There are many groups working on the network design for future cellular and wireless networks. The Broadband Radio Access over IP Networks (BRAIN) project aims to design an 'ALL-IP' network, evolving from the 3G access network to support new air interfaces. It is a micro mobility based approach with QoS handling to provide seamless service provision and QoS adaptation. The Access network would be limited to transport of IP datagrams with assured QoS, security and seamless handover [Wisely, 2001].

The Drive project introduces an overlay design for the network architecture, consisting of two parts: a backbone and individual radio access system. The approach leaves the radio access system unaltered and provides a framework for access between access systems. The backbone is a pure IPv6 network employing HMIP with extensions for more elaborate traffic distribution. The backbone consists of an accounting, authorisation and authentication (AAA) server, a multi access support node (MSN) and a home registration support node (HRSN). The DRiVE interface unit (DIU) links up various types of access systems to the backbone. Access systems contain a Traffic Management Node (TMN) and Spectrum Allocation Node (SAN) [Paila, et al., 2001].

The Moby-dick project aims at developing an 'ALL-IPv6' architecture supporting seamless access across heterogeneous networks with AAA (authentication authorization accounting) and end-to-end QoS architecture. Mobility management is based on FMIP and paging whereas differentiated QoS would be supported along with a QoS broker. A WCDMA base station would be attached directly to the IP network [Einsiedler, et al. 2001].

The Wineglass project is developing a deeper integration between IP and mobile systems. The approach adopted is to re-use the state UTRAN and to attach it to the enhanced Mobile IPv6 backbone through the standard IU interface. GPRS nodes in the network have been removed altogether and specialised border routers are required. Session, mobility and authentication aspects are not duplicated at both IP and UMTS level, which is a major benefit [Annoni, et al., 2001].

None of the 4G working groups have defined a service advertisement procedure. Zhang proposed to deliver neighbouring cell specification as a UMTS value added service by the user's current point of attachment [Zhang, 2003]. After authentication, a network could provide the node with the list of candidate networks available in the surrounding area.

4G working groups have still to outline a procedure for service selection. Even though most of the architectures support multiple-operators, they have not defined mechanisms for service advertisement and selection. Chan proposed to use fuzzy logic based algorithms for channel selection [Chan, et al., 2001].

1.4 Research Aims and Objectives

This research aimed to study, investigate and design mechanisms for delivering service advertisements to the terminals with a reduced scanning overhead. Efficient service selection algorithms utilising advertisement information collected by the terminals was also examined.

The **objectives** of this study were-

1. **To research 3G and 4G capabilities with regard to multiple operator support**

2. **To review past and present work in the area Service Advertisements Data Repository**

3. **To review past and present work in the area Service Advertisements Access Mechanism**

4. **To design a Service Advertisement Data Repository layout**

5. **To design a Service Advertisement Access Mechanism**

6. **To devise and design mechanisms for Service Advertisement Delivery to the user with a reduced scanning overhead**

7. **To devise mechanisms for Service Advertisement Delivery, that can be employed in the absence of the main service provider.**

8. **To Critically study and analyse past and present mobility management architectures**

9. **To devise modifications to reduce handover latencies in heterogeneous network handoffs.**

10. **To examine and critically analyse past and present Service Selection Algorithms**

11. **To design and simulate a user friendly Service Selection Algorithm which considers user preferences as a priority when selecting a service.**

12. **To characterise the service characteristics, that can be used for Service Selection from a user perspective.**

1.5 Outline of the Thesis

Chapter 2 Evolution of mobile and wireless technologies – The chapter discusses the evolution of wireless technologies and the Fourth Generation of mobile technology. The chapter presents a critical analysis of various mobility management architectures. The chapter then introduces services discovery mechanisms including service advertisements and service selections. The chapter then critically evaluates the architectures of various working groups for their capability of supporting heterogeneous access.

Chapter 3 Service Advertisements – Design Requirements & Architecture – The chapter presents an overview of service advertisements leading to a critically analysis of the requirements. The chapter introduces service advertisement data repository and access mechanism. The data repository defines an architectural and functional layout of the service repository and a single frequency band access mechanism is introduced. The evaluation of the data repository and access mechanism is also included in the chapter.

Chapter 4 Service Selections expressing user's preference– This chapter discusses the background of selections and introduces the proposed solution. The chapter discusses interface compatibility and location based elimination approaches. It further discusses network selection parameters along with an evaluation of the algorithm. .

Chapter 5 Mobility management in future mobility networks – This chapter discusses mobility management architecture for the proposed framework. This chapter contains simulation results for HMIP, FMIP and SMIP mobility management architectures. it also includes a detailed discussion of the modifications to the HMIP protocol to reduce global handoff latency. The chapter further discusses modifications to the WLAN binding tables to enable smoother handovers between WLAN and HMIP networks. It also explained how the service advertisements and service selections fit into the mobility management architecture.

Chapter 6 Conclusion – The chapter discusses the major findings of the thesis where major results and contributions are presented.

Chapter 2 Evolution of Mobile and Wireless Technologies

2.1 Mobility Management

Freedom of movement is what differentiates wired networks from wireless networks. Although the Internet offers access to information sources worldwide, one would not expect to benefit from that access until the arrival at some familiar point whether it is at home, office, or school. However, the increasing variety of wireless devices offering IP connectivity, such as PDAs, handheld devices, and digital cellular phones, are beginning to change the perceptions of the Internet.

The use of conventional network configurations requires that a computer modify its network settings when moving or connecting to a different network. The modified network settings did not permit the packets addressed to the computer's old network configuration to be rerouted to the computer's new network settings. This caused much inconvenience for the end users. The situation became even worse when the computer kept roaming and switching to different networks, e.g., a person browsing the internet whilst travelling on a train.

To solve these problems and to provide real time mobility to mobile users, Mobile IP was introduced. Mobile IP came as a result of a modification of IP that allowed nodes to continue to receive datagram's (i.e. UDP packets) no matter where they happen to be attached to the Internet. It involved some additional control messages that allowed the IP nodes involved to manage their IP routing tables reliably [Perkins, 1998]. The aim of the design for mobility support was to keep computers and laptops connected with no additional modifications or changes while roaming or changing their point of attachments.

Mobile IP assigns an IP address for identifying the terminal's home network (home address) and its current point of association (foreign network assigned care of address). The network layer was chosen to make alternations whereas protocols in the other layers remain unchanged as much as possible. The advantage of designing Mobile IP based on modifying the network layer protocol was to make it independent of the physical layer, which meant that any communication media, including wired and wireless networks, would support Mobile IP. In IP networking, one IP address indicates the point of the attachment for each node. When a laptop connects to a different network, it needs a new IP address to indicate its current location and keep its communication over the Internet; otherwise, the packet addressed to the node connected to a different network becomes un-routable. Consequently, it is evident that the key point of Mobile IP design is making IP addresses transparent.

Mobile IP was intended to enable nodes to move from one subnet to the other. It is just as suitable for mobility across heterogeneous media as it is for mobility across homogeneous media. That is, Mobile IP facilitates node movement from one Ethernet segment to another as well as accommodates node movement from and Ethernet segment to a wireless LAN, as long as the mobile node's IP address remains the same after such a movement.

2.1.1 Requirements for mobility management

When a Mobile Node using Mobile-IP undergoes handover from one access router to another, the path traversed by the Mobile Node's packet stream in the network may change. Such a change may be limited to a small segment of the end-to-end path near the extremity, or it could also have an end-to-end impact. Further, the packets belonging to a Mobile Node's ongoing session may start using new care-of-address (CoA) after handover. Hence, they may not be recognised by some forwarding functions in the nodes even along that segment of the end-to-end path that remains unaltered after handover. Finally, handover may occur between the subnets that are under different administrative control. Requirements for Mobility Management [Bansal, 2002], [Wisely, 2001], [Zhigang, et al.], [Karagiannis, Heijenk, 2000], [Varshney, 2001]

- **Minimising handover latency** - When a mobile node changes its point of attachment, the latency should be minimised and minimum QoS channel specification must not be compromised.
- **Packet loss at handover** - Packet loss due to the handover should be minimised. Increased packet loss would result in retransmission of lost packets and would add to handoff latency.
- **Interoperability within mobility protocols** - The QoS mechanism for Mobile-IP should take advantage of the complementary mobility protocols for optimised performance. However, the QoS scheme must have provisions to accomplish its tasks even if one or more of these mobility protocols is not supported by an operator.
- **Interoperability with other access technologies** – Handovers between different access technologies should be supported. Next generation of mobile networks would enable heterogeneous access across operators providing services on various interfaces; hence interoperability between access technologies should be supported.
- **Maintain QoS status post handover** – When handing over across administrative domains a terminals QoS state should be maintained.

Various mobility management schemes have been designed to satisfy some of these requirements.

2.1.2 Mobile IPv6 (MIPv6)

2.1.2.1 Registration

When a mobile node is in a foreign network, a care-of address is necessary to present the mobile node's current location. Due of the absence of foreign agents in MIPv6, the temporary addresses for mobile nodes are co-located care-of addresses. These care-of addresses can be generated by the use of stateless address auto-configuration or dynamic host configuration protocol (DHCPv6) [Perkins, 1998].

One mobile node can have several care-of addresses at the same time. Once it receives a Router Advertisement with a different network prefix from a particular router, the mobile node asks for a new care-of address to be generated. [Perkins, 1998]

On getting the care-of addresses, the mobile node then registers one of them to its home agent by sending a Binding Update to the home network. A Binding Acknowledgement would be returned to notify the mobile node of the successful registration. The care-of address registered to the home agent is referred to as a ***primary care-of address***, and the router that has the same network prefix as the primary care-of address is denoted as the ***default router*** [Damberg, et al., 2002].

2.1.2.2 Packet Delivery

For delivering a packet addressed to a mobile node, if a correspondent node has no binding entry for the mobile node, the packet would be sent to the mobile node's home network. Consequently, the home agent intercepts it by making use of a proxy *Neighbour Advertisement.* Thereafter, the packet is tunnelled to the mobile node by the use of IPv6 encapsulation. On receiving the encapsulated packet, the mobile node realises that the packet travelled via the home agent, and informs the correspondent node to set up a binding entry for the mobile node. However, it is not compulsory for the mobile node to notify the correspondent node of its care-of address. If the correspondent node contains a binding entry of the mobile node, the packet is directly dispatched to the mobile node. This is achieved by placing the mobile node's care-of address as the last intermediate address in the Routing Header, and setting the mobile node's home address as the destination address.

When the mobile node wants to deliver a packet to a correspondent node, the packet only needs to be sent in the conventional way. In addition to setting the source address to the mobile node's care-of address, the packet must include a Home Address option, which indicates the mobile node's home address.

2.1.2.3 Handover latency

Mobility management in Mobile IP is dependent on movement detection by the mobile node and the registration process with the Home Agent. A quick look at the handover process –

- MN forms a new stateless address based on the BS prefix and sends a Binding Update (BU) to the BS.
- BS updates its binding cache and sends a Binding Acknowledgement (BAck) to the MN.
- MN sends a BU to its HA (and CN) notifying it of its site movement.
- HA updates its binding cache and sends a BAck to the MN.

This process introduces two types of latencies –

- **Move detection latency [Hsieh, et al., 2002]** – it is defined as the time required by a mobile node to detect the change in its point of attachment (PoA). Since, the movement detection is only done on the basis of expiry or address prefix comparison of two different agent advertisements, the latency introduced is quite noticeable.
- **Registration Latency [Hsieh, et al., 2002]**– Since the HA could be located anywhere on the Internet, the latency could vary considerably and could potentially take a very long time.

MN notifies HA when it moves between networks but this indirection requires a registration process and address resolution procedure. The indirection procedure results in long handover latencies, which often leads to packet loss and performance degradation [Hsieh, et al., 2002]. Since, the HA would have to be notified of every move, the amount of control packets travelling on the network could be high if the MN is moving quickly across networks and adds to the handover latency [Bansal, et al., 2003]. Along with movement detection and registration latency, control packets also affect the handover latency [Bansal, et al., 2003].

2.2 Mobility Management in Future Heterogeneous Environments

Mobility in future mobility networks would not be limited to the same network and the same access technology. The table 2-1 classifies heterogeneous and homogeneous handovers.

Table 2-1 Handover classification

Category	Same wireless access technology	Different wireless access technology
Same administrative domain	Homogeneous	Heterogeneous
Different administrative domain	Heterogeneous	Heterogeneous

Mobile assisted handovers work well in homogeneous networks within the administrative domain, whereas in heterogeneous network they have some disadvantages. However, inter-domain handovers don't require any security considerations whereas if it is a cross-domain handover, the node would have to go through the authentication process before accessing the new domain. The service profile of mobile nodes would vary from domain to domain whereas it stays the same within a single domain. Accounting and authorisation information would have to be refreshed every time the node is handed over to a new administrative domain, which adds to the handover latency. Network controlled handovers would prevent a user from exercising their right of choice by looking elsewhere other than the subscribed to network. To enable heterogeneous access across networks, only mobile-controlled handovers will be considered [Zhang, et al., 2003].

2.3 Mobility Management Architectures

2.3.1 Predictive Mobility management schemes

Predictive mobility management schemes predict which Access Router a user would handover to before the event. The prediction is based on the user's movement profile

[Fang, Reeves, 2004]. The predictive mobility management schemes rely heavily on the capability of the prediction estimation. Reliability of handover predictions had been around 50% [Schimdt, et al., 2005]. The predictive mobility management schemes have not been considered because of their low prediction success rate. An incorrect prediction leads a user to undergo another handover, which adds to the handover latency.

2.3.2 Reactive Mobility Management Schemes

Reactive mobility management schemes undergo a handover by reacting to an event, e.g. change in received signal strength, etc. The following reactive mobility management schemes had been considered.

2.3.2.1 Regional Mobile-IP (RMIP)

RMIP [Suh, 2002] is intended to reduce the amount of signalling to a Mobile Node's Home Agent and its Correspondent Nodes. The protocol does not assume a certain network topology of Access Routers, for example hierarchical structure. As a result, the protocol is flexible enough to adapt to any network topology assumed by IPv6.

The RMM (Regional Mobility Management) mechanism is intended to reuse the Care of Address that was obtained by MN in the previous AR (Access Router) and make the AR (visited by MN) as a RAP (Regional Anchor Point) for the MN. The forwarding path from a RAP to a MN can be a controllable or configurable, because the RAPs are dynamically determined by MNs (mobile determined region) or RAPs themselves (network determined region).

RMIP is based on the evaluation of the cost function (distance limitation) leading to four scenarios:

- The old AR (Access Router) supports RAP
    - The old AR (Access Router) of the mobile Node supports the RAP functionality and the cost evaluation is satisfied, MN can use the pCoA (previous Care of Address) as its HMIP MAP address.

    - The old AR of the MN supports RAP functionality but the cost evaluation is not satisfied, then MN reverts back to the normal MIPv6 operation. This would lead to longer delays.
- The current AR does not support RAP functionality and it is located within distance limitation, the mobile node uses the CoA which is obtained from the previous AR (AR1) as an RCoA.
- The current AR does not support RAP functionality and it is located out of distance limitation, MN stops using RMM/RMIPv6 and reverts back to the basic Mobile IPv6.

RMIP is designed for reducing cost of operation for the networks but is only useful if the old new access routers are close to each other. Since if there is no link between them, the packet would have to travel from the old AR to the new AR over the Internet. This would add to the handover latency. Such a scheme is not the ideal one for heterogeneous access as the packets would have to travel across domains and it would be hard to justify the cost function in such a scenario.

2.3.2.2 Hierarchical Mobile-IP (HMIP)

Macro mobility schemes overcome the problem of registration latency by differentiating between inter and intra domain movement [Castelluccia, 1998]. Hierarchical Mobile IP (HMIP) provides a scheme for performing registrations locally in the foreign network, thereby reducing the number of signalling messages forwarded to the home network as well as lowering the signalling latency that occurs when a mobile node moves from one Base Station (BS) to another. HMIP introduces a conceptual entity, Mobility Anchor Point (MAP), to separate local mobility from global mobility. MAP acts as a Mobile Node's (MN) Home Agent (HA) in the foreign network. MAP intercepts all packets destined for MN and tunnels them to its on-link CoA (LCoA).

In HMIP, MN attaches itself to two Care of Addresses (CoA), a regional (or site) address and an on-link (or point of attachment) address. Figure 2-1 shows a typical HMIP network architecture [Castelluccia, 1998]. The regional and local CoA's of the MN are formed in a stateless manner, i.e. MN forms these addresses based on the site (and local) prefix in router advertisements. To avoid duplicate addresses all the nodes run a duplicate address detection algorithm [Karagiannis, 1999].

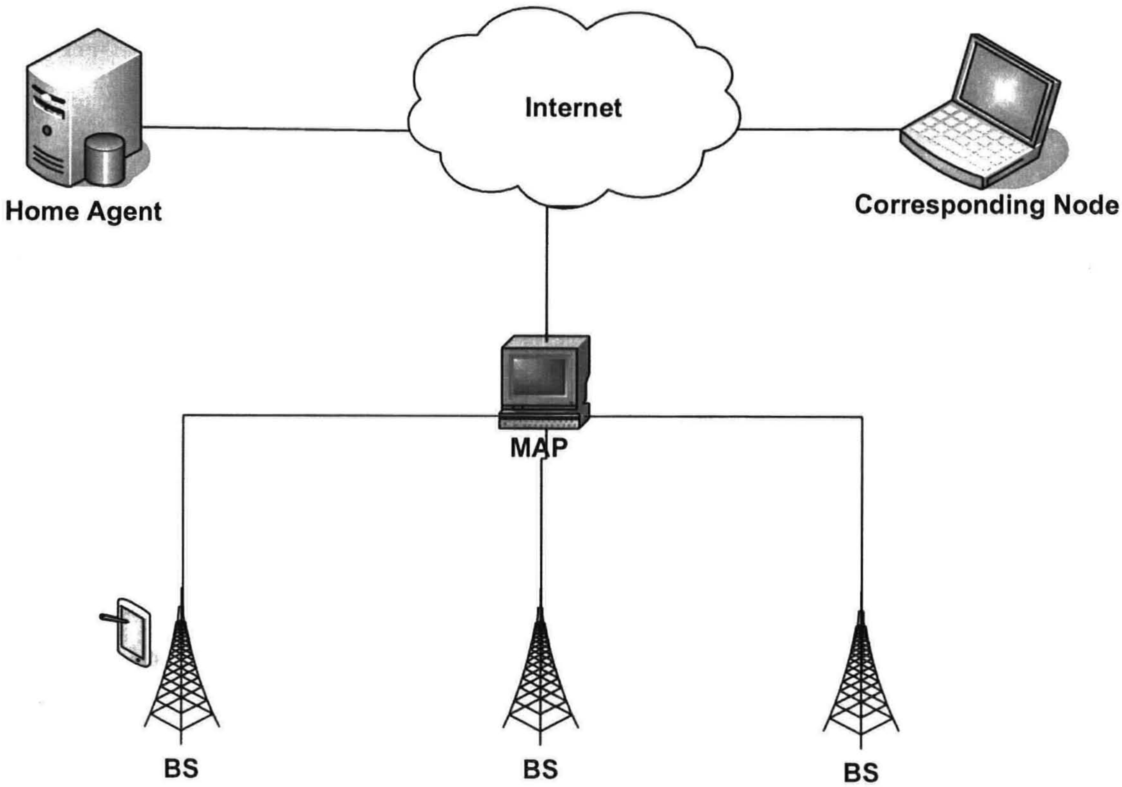


Figure 2-1 Typical HMIP Network Architecture

2.3.2.2.1 Global Handover HMIP

When MN enters a new site, it goes through a global handover. A global handover consists of these stages [Bansal, et al., 2003]–

- MN forms a new stateless address based on the BS prefix and sends a Binding Update (BU) to the BS.
- BS updates its binding cache and sends a Binding Acknowledgement (BAck) to the MN.
- MN forms a new stateless site address based on the MAP address in the router advertisements and sends a BU to MAP.
- MAP updates its binding cache and sends a BAck to the MN.
- MN sends a BU to its HA (and CN's) notifying it of its site movement.

* HA updates its binding cache and sends a BAck to the MN.

Figure 2-2 shows the sequence of message transfers between MN, BS, MAP and HA during the handover [Castelluccia, 1998]. In the process, we had assumed that there were no duplicate addresses detected.

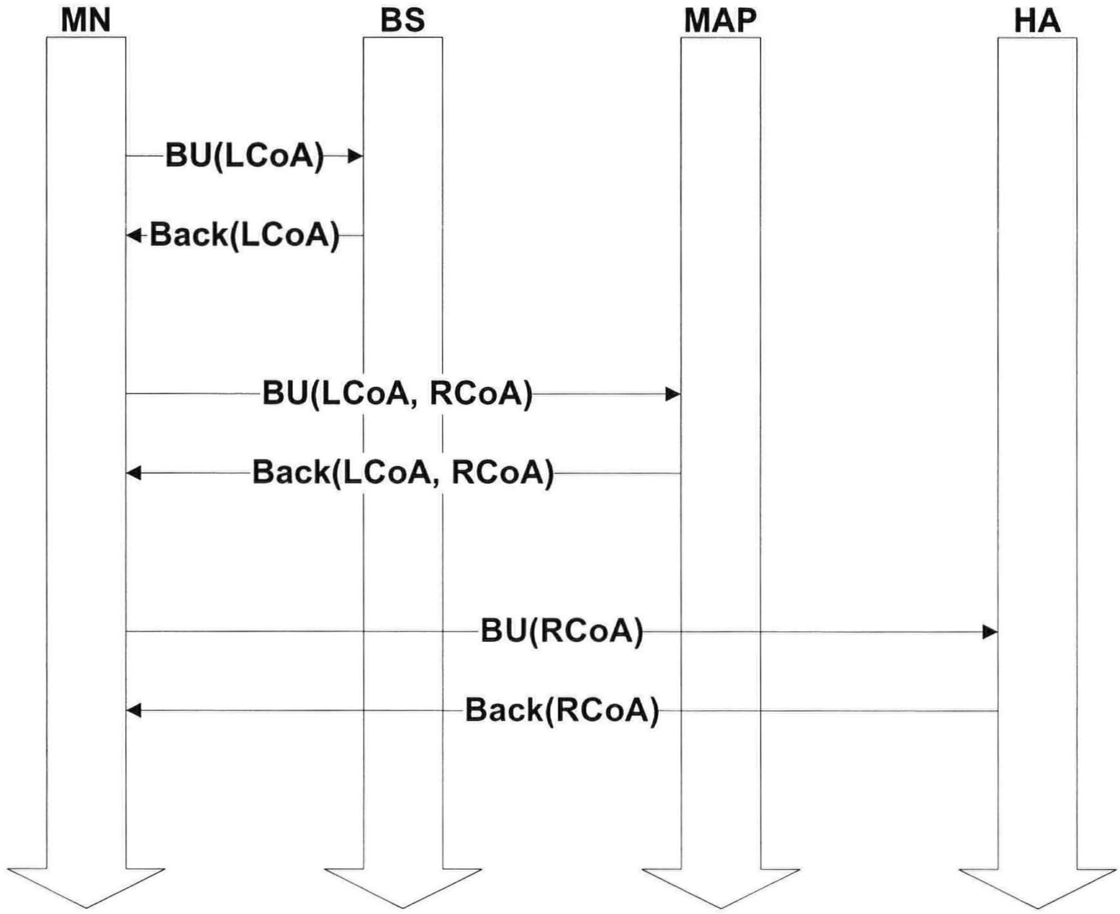


Figure 2-2 Interaction between MN and BS, MAP, HA during a global handover [Castelluccia, 1998]

2.3.2.2.2 Local Handover HMIP

Local handover (handover within a MAP domain) consists of the following stages [Bansal, et al., 2003], as shown in Figure 2-3–

* MN forms a new stateless local address based on the router advertisements and sends a BU to BS.
* BS updates its binding cache and sends a BAck to the MN.

- MN sends a BU to the MAP notifying it of its site movement.
- MAP updates its binding cache and sends a BAck to the MN.

The main differences between local and global handover is that MN does not need to notify HA of its movements within a site thus reducing handover latencies.

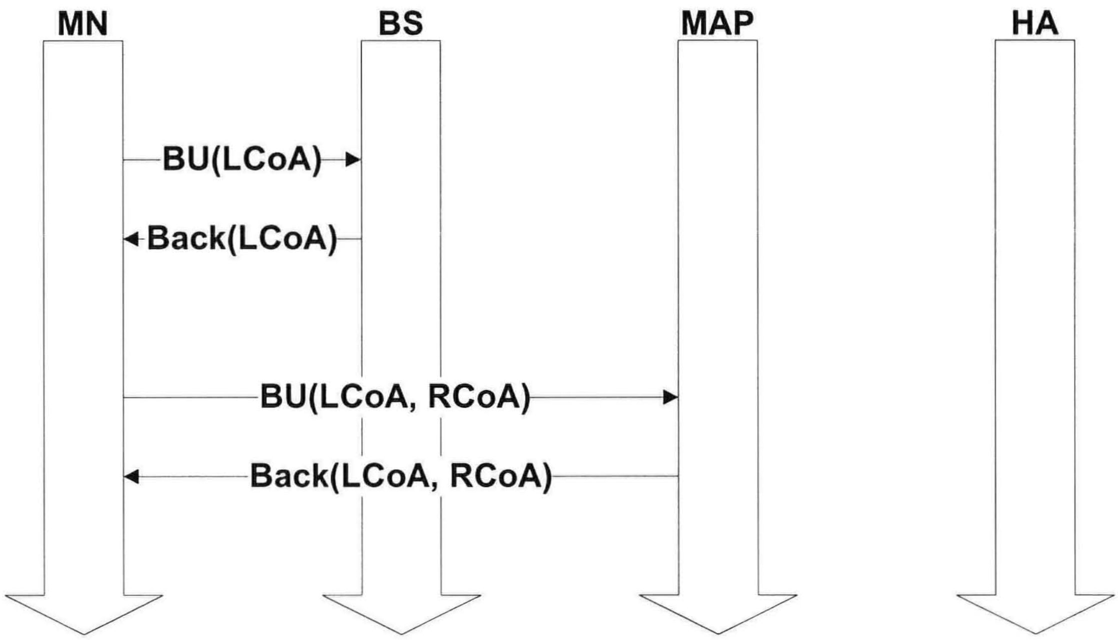


Figure 2-3 Interaction between MN, BS, MAP and HA during a local handover [Castelluccia, 1998]

2.3.2.2.3 Drawbacks of HMIP

- Since, all the communication within a MAP domain pass through the MAP, it can reduce quality of service [Hsieh, et al., 2003].
- The failure of the MAP would lead to the breakdown of the whole network [Hsieh, et al., 2003].
- The inter-MAP handover would lead to longer delays which become even greater if the MAPs are topologically far apart [Reinbold, 2002].
- Signalling load within the network increases but the global signalling load decreases [Costa, et al., 2002].

2.3.2.3 Fast Mobile-IP (FMIP)

Fast handoff schemes reduce the move detection latency through address pre-configuration [Koodli, 2002]. Fast-handoff schemes in MIPv6 introduce seven additional message types for use between access routers and the mobile node.

These seven messages are [Koodli, 2002]:

- Router Solicitation for Proxy (RtSolPr)
- Proxy Router Advertisement (PrRtAdv)
- Handover Initiation (HI)
- Handover Acknowledgement (HAck)
- Fast Binding acknowledgement (F-BAck)
- Fast Binding update (F-BU)
- Fast Neighbour advertisement (F-NA)

Fast Mobile IP operation -

- A fast-handoff is initiated on an indication from a wireless link-layer 2 (L2) trigger. The L2 trigger indicates that the MN will soon be handed off. Upon receiving an indication, the fast handoff scheme anticipates the MN's movement and performs packet forwarding to the nAR(s) accordingly. This is achieved by the MN sending a RtSolPr message to the oAR indicating that it wishes to perform a fast-handoff to a new attachment point. The RtSolPr contains the link-layer address of the new attachment point, which is determined from the nAR's beacon messages.
- In response, oAR will send the MN a PrRtAdv message indicating whether the new point of attachment is unknown, known or known but connected through the same access router. Further, it may specify the network prefix that the MN should use to form the new CoA.
- Based on the response, the MN forms a new address described using the stateless address configuration [Davidson, 2003].

- Subsequently, the MN sends an F-BU to the oAR as the last message before the handover is executed.
- The MN receives an F-BAck either via the oAR or the nAR indicating a successful binding. As the exact handoff instance is unpredictable, the oAR sends a duplicated F-BAck to the nAR to ensure the receiving of F-BAck by the MN.
- Finally, when the MN moves into the nAR's domain, it sends the Fast Neighbour Advertisement (F-NA) to initiate the flow of packets at the nAR.
- In addition to the message exchange with the MN, the oAR exchanges information with the nAR to facilitate the forwarding of packets between them and to reduce the latency perceived by the MN during the handoff. This is realized by the oAR sending a HI message to the nAR. The HI message contains MN's requesting CoA and the MN's current CoA used at the oAR. In response, the oAR receives a Hack (Hi Acknowledgement) message from the nAR either accepting or rejecting the requested new CoA. If the new CoA is accepted by the nAR, the oAR sets up a temporary tunnel to the new CoA. Otherwise, the oAR tunnels packets destined for the MN to the nAR, which will take care of forwarding packets to the MN temporarily.

Fast Mobile IP reduces the address resolution time but the registration latency is still a problem but FMIP is still quicker than HMIP [Costa, et al.,2002].

2.3.2.4 Seamless Mobile-IP (S-MIP)

S-MIP builds on fast and hierarchical mobile-IP schemes. S-MIP exploits the fact that hierarchical Mobile IP reduces registration latency and fast Mobile IP reduces address resolution time and the combination has proved to reduce latency [Hsieh, et al., 2002]. S-MIP handoff strategy is terminal initiated but network determined [Hsieh, et al., 2003]. Since the handoff decisions are made by the network, access across competitor network would be an issue.

S-MIP extends the HMIP architecture by introducing a Decision Engine (DE). The DE is similar to a MAP in its scope, and makes handoff decisions for its network domain. DE maintains a global view of the network through periodic feedback information from AR's and terminals through their periodic movement status update. DE is capable of offering load balancing services.

S-MIP introduces these additional messages –

- **Current Tracking Status (CTS)** message from the MN to DE. It contains location tracking information.
- **Carrying Load Status (CLS)** message from the ARs to DE. The CLS message contains the information regarding how many mobile devices an AR is currently managing.
- **Handoff Decision (HD)** message from the DE to ARs. The HD message contains the outcome of the handoff decision at the DE, namely which AR a MN should handoff to.
- **Handoff Notification (HN)** message from the oAR to MN. The HN contains the indication from the oAR to the MN, directing exactly which nAR the MN should handover to (this is sent in combination with the PrRtAdv). The oAR derives the content of the HN message from the received HD message.
- **Simulcast (Scast)** message from oAR to MAP. The Scast message triggers the start of the SPS process.
- **Simulcast Off (Soff)** message from nAR to MAP. This message terminates the SPS process.

SMIP would reduce handoff latencies but adds to the weakness of HMIP by adding another crucial entity to the architecture, which if unavailable could lead to potential chaos and the network having to revert back to normal HMIP operation. Also, the scheme hands most of the control of the handoff decision to the network which would harm the goal of heterogeneous access across networks and it also increases the signalling load.

2.3.2.5 PROTON

PROTON is policy based mobility management architecture [Vidales, et al., 2004]. The main components included context management, policy management and policy enforcement. The model focussed on defining an efficient policy structure on top of MIPv6 mobility management architecture. Proton's context gathering retrieves information required to process user preferences individually from every network. Negotiating with individual networks would add handover latencies [Dutta, et al., 2005]. Proton did not define any means for service discovery, without which the context management would not be as effective.

2.4 Service discovery

Increased mobility and a heterogeneous network environment result in mobile devices frequently connecting to networks about which they lack information. The terminal would scan the available frequencies to locate accessible networks, which results in resource consumption and increased handover latency. Service discovery mechanisms enable users to gather information about candidate networks with or without the overhead of scanning.

2.4.1 Operator Dependent

2.4.1.1 CARD Protocol based approach

Melazzi et al. have outlined a procedure for resource discovery on a mobile node's current network to outline various access mechanisms offered by the operator [Blefari-Melazzi, et al., 2004]. The Candidate Access Router Discovery (CARD) protocol enables identification of the IP addresses of Candidate Access Routers (CAR's) and finding their capabilities [Liebsch, et al. 2005].

CARD approaches for service discovery-

1) MN's can obtain Layer 2 Access Router mapping from the beacons of surrounding Access Points and pass these to the current point of attachment for resolving the addressing, under the CARD approach. MN can send the address resolution to a centralised CARD server, where all the services are registered, if the L2 mapping is not found in the local cache.

2) Another approach is based on distributing and exchanging information between neighbouring Access Routers. Access Routers which have handled handover events are defined as neighbouring.

Melazzi's approach is based on the first CARD service discovery mechanism and the following is a brief explanation of this approach.

The Melazzi approach –

This approach is summarised in the following points

1) Terminal listens to the beacons on the same radio access technology
2) Terminal notifies its current Access Router of the new access points
3) Terminal's Access Router would issue a multicast request, requesting information from the Access Router managing the Access point, if it does not hold information about the Access Point.
4) The Access Router would then send the IP address and service capabilities along with an invitation to join its multicast group of the new Access Point to the MN's current Access Point.
5) MN's current Access Router would respond with the L2 id of its own, service capabilities and an invitation to join its multicast group.
6) Both the access routers can then join each others multicast groups.
7) MN's current Access Router would then notify it of the service capabilities of the new Access Point, which could then be utilized by MN's service selection algorithm.

This approach discourages MNs from scanning for beacons on more than one interface at a time. The approach fails to satisfy any of the service discovery requirements and is only suitable for discovering new Access Routers in the area. The discovery for new access routers is done via beacons and the MN is unable to deploy more than one interface under the approach.

2.4.1.2 Neighbourhood resource discovery

Neighbouring cell specification [Zhang, et al., 2003] could be delivered by the MN's current point of attachment upon authentication. MN's current AP could send it a list of neighbouring cells. After authentication, the network could provide the node with the list of candidate networks available in the surrounding area. MN's current network might not be willing to provide a list of candidate networks as this could be used to the advantage of its competitors. Also, if the MN's subscribed network does not offer services in a particular area, then MN would be unable to find out about the candidate networks. UMTS networks could provide this information as a value-added location based service. Neighbouring cell specifications fails to satisfy most of the service advertisement requirements.

2.4.2 Operator independent

2.4.2.1 IEEE 802.21 based Information Services

2.4.2.1.1 Universal Information Service (UIS)

UIS comprises [Dannewitz, et al., 2006] –

- ⇒ Information Service – The component is used to store and deliver service data. The component delivers configuration data to the terminal as well.
- ⇒ Event Service – Informs mobile entity about changes to its environment.
- ⇒ Command Service – Reacts to the change in environment and interacts with it.

The UIS reduces the scanning overhead by providing an information service but fails to standardise details, which would be used to compare services. The architecture did not define a layout of the information service repository. The architecture describes a context-aware service discovery mechanism but fails to standardise service information.

2.4.2.1.2 Media Independent Information Server (MIIS)

MIIS [Yaqub, Madhani, 2007] created a database of existing networks mapped with their Global Positioning Satellite (GPS) co-ordinates. The information received from the MIIS was stored locally at the terminal for future use. The terminal would lookup its own repository for service information and if the information does not exist, it would consult the MIIS. The architecture identifies QoS, cost, connection and security as parameters for network selection but did not define how these parameters would be acquired by the server. The architecture assumes that location of the MIIS would be known to the terminal and does not define how to acquire these details if the location of the MIIS is not known to the device. Mapping the GPS co-ordinates of every network would affect the handoff latency; refer to Figure 3-6. The mechanism did not specify a layout for load balancing.

2.4.2.2 Model Based Service Discovery

The architecture incorporates a service registry for holding service details [Kun, et al., 2006]. The architecture further includes a Service Discovery Agent (SDA), which discovers service details from the local registry. The service discovery mechanism did not standardise the service details held by the service registry. The distinction between types of services was also not made. The layout and architecture of the registry has also not been defined.

2.4.2.3 Multimedia Integrated network by Radio Access Innovation (MIRAI)

Gang Wu proposed an IPv6 Common Core Network (CCN) acting as a central hub for all mobile terminals [Wu, et al., 2001]. All access points were principally connected to this network. The network would ensure QoS guaranteed routing and seamless handovers among radio access networks.

Gang Wu's MIRAI architecture supports service advertisements delivery over the Common Core Network (CCN) [Wu, et al., 2001]. The service discovery mechanism does not specify how the information would be transferred to the terminal. The architecture though capable of delivering service advertisements does not specify how the information would be stored and made accessible to the terminal.

2.4.3 Service Selection

The service discovery mechanism enabled the terminal to acquire service information, service selection mechanism enables it to select a service based on the desired characteristics.

2.4.3.1 Policy enabled handoffs across heterogeneous wireless networks

Wang, Katz and Giese drafted a policy based handoff decision system where the policies were user defined [Wang, et al., 1999]. The following policy parameters were considered–

1) Cost of using the service
2) Network status which included available bandwidth, latency and reliability. Network status was considered to be a dynamic parameter and the authors proposed to use this parameter to make cost performance tradeoffs. A performance agent developed by the authors collected performance related data at the base stations and broadcast that information to the terminals. To avoid the scenario of several terminals discovering and switching to a better network simultaneously, the authors introduced a random handoff stability period which in turn produced random waiting periods.
3) Wireless interfaces have different power consumptions based on the network with which communication has been established. An access point closer to the terminal would require lesser signal strength as compared to an access point further away. The user could wish to use networks with lower power consumptions if the terminal is low on battery.
4) Connection setup time in connection oriented networks could also be employed as a policy parameter.
5) Velocity could also be used as policy parameter. Users could disregard certain types of networks when travelling at considerable speeds. For example, users would disregard wireless LAN networks when travelling at 50Mph, since the terminal would be out of reach of the access point in a short while. Whilst on the

move, a user should not consider networks with a short coverage area, as the terminal could be out of reach of the network fairly quickly.

2.4.3.1.1 Cost Function

Cost of using a network 'n' at a certain time is a factor of:

1) Bandwidth offered by network (B_n)
2) Power consumption of using the network access device (P_n)
3) Cost of using the network (C_n)

The cost function of the network therefore can be represented as

$$Cost_n = f(B_n, P_n, C_n)$$

Equation 2-1 Wang Cost Function

The user can specify the maximum cost they are willing to bear for using the service. The terminal's battery life and bandwidth available on the network are the other fixed variables in the equation. The user can also specify the importance (or weighting) for each of these parameters.

Total Weighting (W_i) = Bandwidth Weighting (W_B) +

Power Consumption Weighting (W_P) +

Cost Weighting (W_C) = **1**

Equation 2-2 Wang cost function weight factor

The normalised cost function proposed by the authors is

$$f_n = W_B \cdot \ln\frac{1}{B_n} + W_P.\ln P_n + W_C \ln C_n (\sum W_i = 1)$$

Equation 2-3 - Wang cost function for network selection

2.4.3.1.2 Analysis

The amount of control message signalling was one of the drawbacks of Mobile IP based networks. The scheme proposed to use a performance agent, which on a regular basis broadcasts performance information to the terminals. This further added to the control messages travelling on the network as well as the workload of base stations. The performance agent relied on the co-operation of the base stations which might not be appreciated by the operators. The authors have not indicated that how they intended to use the information relayed by the performance agents. The cost function relied only on the parameter 'bandwidth offered' as the only performance related selection criteria.

2.4.3.2 Zhu and McNair's vertical handoff decision algorithm

Zhu and McNair proposed a policy based handoffs and developed a cost function with elimination factor [Zhu, McNair, 2004]. The policy parameters considered are -

1) Network performance related parameters like reliability, latency and data rate.
2) Cost of using the service
3) Network condition – traffic, available bandwidth, network latency and congestion.
4) System performance parameters – channel propagation characteristics, path loss, inter channel interference, signal-to-noise ratio, bit error rate and battery power.
5) Mobile terminal conditions – velocity, moving pattern, moving histories and location information
6) User preferences

2.4.3.2.1 Cost function

For each network, n, the cost of using the service, s, is determined by-

$$C_s^n = E_s^n Q_s^n$$

Equation 2-4 - Cost of using a service

where E is the elimination factor for a requested service and Q is the QoS factor for the network.

The QoS factor in the cost function is determined by –

$$Q_s^n = \sum_j W_{s,j}^n Q_{s,j}^n$$

Equation 2-5 - QoS factor in network selection

where Q_s^n is the normalized QoS provided by network n for parameter j for service s, and $W_{s,j}^n$ is the weight indicating the impact of the QoS parameter on the user or the network.

The network elimination factor in the cost function is determined by-

$$E_s^n = \prod_i E_{s,j}^n$$

Equation 2-6 - Network elimination factor

where $E_{s,j}^n$ represents whether the minimum constraint i for service s can be met by the network n. The elimination factor would result in a large value when the constraint cannot be met.

2.4.3.2.2 Analysis

The approach suggested optimisations to the vertical handoff decision algorithms, hence did not specify policy criterion when determining the cost of using the service. The approach relied on other algorithms or the user (/application) to define the policy criterion. The approach was not a handoff decision algorithm in itself but rather a cost function optimisation process. The elimination factor introduced would have to be dynamically calculated, since it was the capability of a network to deliver a service, which would have added to the handoff latency. The elimination factor calculation had been performed using QoS parameters in the example which implied the elimination would have been performed on the basis QoS requirements being satisfied.

2.5 Working Groups

This is an overview of the visions and approaches of the main working groups for 'ALL-IP' architectures.

2.5.1 MIRAI

The MIRAI architecture consisted of two main parts, a Common Core Network (CCN) and a Basic Access Network (BAN) [Wu, et al., 2001]. BANs provide a common signalling and control channel for access of all MNs. A BAN would assist in wireless system discovery and would have a larger coverage area than the Access Networks under its control. A BAN would provide MNs with currently available wireless networks.

MIRAI is capable of delivering service advertisements via the CCN, however did not define a framework for delivering them. MIRAI architecture did not specify whether the CCN and BAN would be independent of larger operators or perhaps would be managed and controlled by them. This raised a serious competition issue where if they were not managed by the main operators, then the issue of controlling and maintaining these networks would be paramount.

2.5.2 BRAIN and MIND

As discussed in section 1.3, the project employs an ALL-IP network coupled with QoS micro-mobility to provide seamless service provision and QoS adaptation. . The Access network would be limited to transport of IP datagram's with assured QoS, security and seamless handover [Wisely, 2001].

2.5.3 Drive

As discussed in section 1.3, Drive network consists of IPv6 backbone and individual radio access system. Drive supports various types of access systems by employing the DRiVE interface unit (DIU). Access systems contain a Traffic Management Node (TMN) and Spectrum Allocation Node (SAN) [Paila, et al., 2001].

2.5.4 Moby-dick

The project aimed at developing 'ALL-IPv6' architecture supporting seamless access across heterogeneous networks with AAA (authentication authorization accounting) and end-to-end QoS architecture. Mobility management is based on FMIP and paging whereas differentiated QoS would be supported along with a QoS broker. A WCDMA base station would be attached directly to the IP network [Einsiedler, et al., 2001].

2.5.5 Wineglass

4G networks evolving from GPRS and UMTS consider the issue of IP as two separate questions:

- Internet connectivity to the mobile station
- IP technology in the interior of the mobile network

Wineglass is developing a deeper integration between IP and mobile systems. The approach adopted is to re-use the state UTRAN and to attach it to the enhanced Mobile IPv6 backbone through the standard IU interface. GPRS nodes in the network have been removed altogether and specialised border routers are required. Session, mobility and authentication aspects are not duplicated at both IP and UMTS level, which is a major benefit [Annoni, et al., 2001].

2.5.6 Comparison of mobility management architectures of major working groups

Comparison of the main 'ALL-IP' network architectures with regards to mobility management and heterogeneous access:

Table 2-2 Comparison working groups

Architectures → / Parameter↓	BRAIN/ MIND	DRiVE	WINE GLASS	Moby-dick	MIRAI	UMTS
Network	IPv6	Overlay with IPv6 backbone	MIPv6 enhanced	MIPv6 (extended FMIP)	IPv6	Enhanced GSM +GPRS + Multimedi a add-ons
Access Network	IP based network supporti ng QoS, micro	Any system that enables IP	UTRAN and WLAN (IEEE 802.11)	Any system that enables IPv6	IPv6 common core network	UTRAN air interfaces

Architectures → **Parameter↓**	**BRAIN/ MIND**	**DRiVE**	**WINE GLASS**	**Moby-dick**	**MIRAI**	**UMTS**
	mobility and several access technologies					
Micro Mobility	Operator choice	Backbone assumes micro mobility on access system level	UTRAN (radio protocol based) and WLAN (MIP based)	Operator choice	Operator choice	Based on radio protocols (RNC/NBAP/RNSAP)
Macro mobility	MIP or SIP	HMIPv6	M-IP	FMIP	MIP	SRNS relocation + MAP
Mobile Initiated Handovers	Yes	Yes	Yes	For WLAN and vertical handover	Yes	NO
Network Initiated Handovers	Yes	Within some access systems	For UTRAN	No	No	Yes
Mobile controlled Handovers	Yes	Yes	For WLAN horizontal handover	No	Yes	No
Network Controlled Handovers	No	Within some access systems	UTRAN and vertical handover	Yes	No	Yes
Mobile assisted Handovers	No	Within some access systems	UTRAN	Yes	No	Yes
IP based	Yes	Yes	Yes	Yes		Optional

Architectures → Parameter↓	BRAIN/ MIND	DRiVE	WINE GLASS	Moby-dick	MIRAI	UMTS
terminal						
Application support	Transparent IP service. No special application support	Multi-access support, location aware and QoS aware services	Location and QoS services	Transparent IP service with support for QoS aware applications	Location and QoS services possible via the common core network	
Multi-operator scenario	Possible	Yes	Possible	Yes	Yes	Not yet
Service Advertisements	Possible	Possible	Possible	Possible	Possible	Possible as a values added service
Service selection	Possible	Possible	Possible	Possible	Possible	No
Interface selection	Possible	Yes	Possible	Possible	Possible	No

2.6 Summary

Mobile and wireless technologies have come a long way since the first generation radio telephony service. IP based networks would push it one step ahead but are hampered by their share of problems ranging from handover latency to service selections. The chapter discussed MIPv6, HMIP, FMIP, RMIP and SMIP mobility management architectures. SMIP would be the preferred mobility management architecture, since it reduces both the move detection and registration handoff latency. The chapter further discussed service discovery and service selection mechanisms. The mobility management architectures of different working groups were evaluated for their heterogeneous movement ability and service discovery provisions. The comparison table 2-2 points out that service advertisements and selections are possible in all the post UMTS architectures but yet to be clearly defined. Subsequent chapters will discuss service advertisements, selections and handover latency in more detail.

Chapter 3 Service Advertisements – Design Requirements & Architecture

Heterogeneous access across networks provides users with great flexibility in selecting services based on their requirements, but such heterogeneous access networks pose great challenges. Heterogeneous access would enable mobile users to control the handover process and decide to which network the mobile node should be handed over. Users would get greater flexibility in exercising choices when making handover decisions. Future mobile networks would cater for various access mechanisms and operators would vary in terms of coverage and access technology. Support of multiple-operators would require efficient service advertisement mechanisms [Bansal, et al., 2004]. Since the future mobile terminals will support more than one access mechanism; operators may vary in size; for instance, there could be a coffee shop offering high speed data access in one particular hotspot to a country-wide operator offering voice, video and data services. Users in a particular area should be able to explore services offered by different networks by enabling them to explore a list of services available with the use of efficient services advertisement mechanisms.

Mobile IP protocols enable nodes to execute IP layer handovers between Access Routers where they assume the details of new access routers are known. The service advertisement can be categorised into:

1) **Service discovery access mechanism** – The access mechanism defines how users can deploy the service discovery mechanism on their terminals.
2) **Service discovery data repository** – The data repository outlines the layout of the service discovery data.

3.1 Requirements for service advertisement mechanisms

Formal requirements for a service discovery mechanism have yet to be defined. To understand the requirements of a modern day mobile user, a scenario based study was conducted.

3.1.1 Scenario

A scenario of a typical young wireless user a few years from now might be -

Linda is a young girl some years ahead from now. She is used to small handheld devices and also to new technologies from a user point of view. Linda does not care about the technical components inside her smart and tiny terminal or what kind of infrastructure was deployed in order to provide her the services she uses. Linda's handheld device provides her with a platform for all her communication needs. Linda's device is equipped with service selection software, which is capable of selecting networks based on her preferences.

The terminal in this scenario is a mobile handheld terminal e.g. a PDA (Personal Digital Assistant) with a variety of wireless interfaces capable of connecting to various networks. In addition to all her local needs - diary, notebook, games, etc - Linda is using it for all her communication requirements and for accessing information. For example, Linda wishes to go to a movie and desires to find out which movies are being shown in town. She initiates a web query to gather a list of movies being shown in different cinemas. The list of movies would be displayed on her PDA with hyperlinks of movie trailers. Linda makes a videoconference call to her friends joining her for the movie. Since, Linda's device is capable of selecting a network based on her requirements, she instructs it to pick the cheapest network for her social calls as call quality is not her primary concern for social calls. Linda and her friends watch the movie trailers to agree upon a movie. After she has chosen her movie, the way to the next theatre is shown on a map and she can order the tickets by looking at the available seats of the respective theatre.

Linda moves quite frequently and uses her PDA for all her communication requirements. Linda utilizes her PDA not only for communication but also entertainment by listening to music either stored locally or downloaded from the web.

Linda's parents in search of sun and sand decided to spend their retired life in Spain. Linda, being much attached to her parents, calls them biweekly. As her conversations with her parents are quite lengthy and often made away from home, she expects her PDA to use the cheapest network available for the calls.

In the office Linda uses her PDA for all her business communication. Whilst using the device for business, Linda's primary concern is the quality of the call rather than cost. She expects her PDA to select the most optimum network for her communications based on her requirements.

3.1.2 Requirements

3.1.2.1 Support for Multiple Operators

Since, Linda expects her wireless terminal to pick the most suitable network based on her requirements; she requires access to more than one service provider. Multiple operators offer great flexibility to the users wishing to seek alternate networks for better services. Since, one of the design goals for future mobile networks is to provide heterogeneous access across networks, it is very important for the service advertisement mechanisms to support multiple operators.

3.1.2.2 Operator transparency and dependence

The operator dependent service advertisement mechanisms would raise competition issues and the terminal would be limited in its ability to explore services. Operator independent service discovery mechanisms allow greater flexibility to the users in exercising their right of choice.

3.1.2.3 Reduced scanning overhead

Linda communicates a lot on the move which would lead her terminal to constantly scan the frequency spectrum to find the available networks for her communication requirements. This constant scanning would lead to delays in selecting the network and would drain the resources of her terminal.

In 2G and 3G networks, handover measurements are controlled by the network. In the UMTS system, prior to handover decisions, neighbouring cell parameters, such as frequencies and scrambling codes are sent to the mobile node by the UMTS terrestrial radio access network/ radio network controller (UTRAN/RNC). In WLAN 802.11b wireless nodes have to scan all the 13 channels to find the corresponding access point (AP), which can take up to 400ms [Zhang, et al., 2003]. Large scale scanning adds to the Layer 2 handovers. Since, future wireless terminals would be able to support more than one interface; the device would have to scan for all the usable frequencies available to the device. Large scale scanning adds to the delays in call setup and handovers. Large scale scanning also drains resources of the terminals. Service advertisement mechanisms should not increase the scanning overhead of terminals.

3.1.2.4 Inter-network recourse discovery

Since, the future mobile networks may offer various access mechanisms; the users should be able to explore services within the network [Melazzi, et al., 2004].

3.1.2.5 Service discovery in absence of main operator

In areas, where Linda's service provider is not providing services, Linda's terminal would have to revert to scanning to locate alternate service providers. Service discovery mechanisms, which enable the users to explore services in an area where its main service provider is absent, would enable Linda to communicate no matter where she is.

3.1.2.6 Service discovery by coverage area

The access technologies of a mobile node would be limited in their reach, so the service advertisement mechanisms should enable the user based on a coverage area. Linda's terminal is equipped with various wireless technologies with varying coverage. Some of the services available to her via her device are very local, e.g. data access via a Wi-Fi hotspot which has typical coverage area of 50m [Schiller, 2003].

3.1.2.7 Support for multiple interfaces

Linda's wireless device is equipped with a variety of interfaces, so the terminal should be able to use any of them to explore service advertisements. The access mechanisms should not limit the interfaces that can be employed by the terminals.

3.1.2.8 Provisions for specialised searches

The usage of mobile users can be broadly classified into –

2. Standard services – Standard services do not vary in terms of price, e.g. browsing the Internet, etc.
3. Specialized services – Specialized services vary in terms of price, e.g. a local call would not be priced the same as an international call even though they both belong to the real-time voice telephony services. Specialised services include any non-standard services, which the operator may offer, e.g. live broadcast of a concert, Linda's call to her parents in Spain, etc.

Storing the pricing for every service by every operator would be a waste of the terminal's resources. Therefore, the advertisement mechanism should enable the mobile node to establish the cost of using the services as and when required.

3.2 Proposed approach – Multiple Operator Service Accessibility Architecture (MOSAA)

3.2.1 Data Repository

The data repository should be independent of the access mechanisms employed. The repository aims to –

1) Reduce the terminal's scanning overhead
2) Allow service discovery by coverage area
3) Allow inter-network resource discovery

4) Support specialised services

3.2.1.1 Distributed architecture with zonal divisions

Nationwide wireless coverage area would be split up in zones based on the volume of wireless traffic. Zonal division of London may look like the London borough map in the Figure 3-1. Zonal division would enable us to focus on more localised services and differentiate between large and small operators. Traditionally any operator providing services citywide or nationwide was termed as a big operator even though that operator may have a very limited presence in a particular area. This thesis refers to operators providing services throughout the zone as large operators whereas operators having a small presence in the borough as small operators.

Figure 3-1 – London Borough Map (Source http://www.ldan.org.uk/cms/view/findservice/ViewLondon.asp)

It is proposed to have a distributed hierarchical database instead of having a central database as shown in the Figure 3-2. The database would have several hierarchies based on country, city and zone divisions. The hierarchical distribution would enable terminals to locate repositories and search their local repository. It will also enable zonal repositories to look after a localised user.

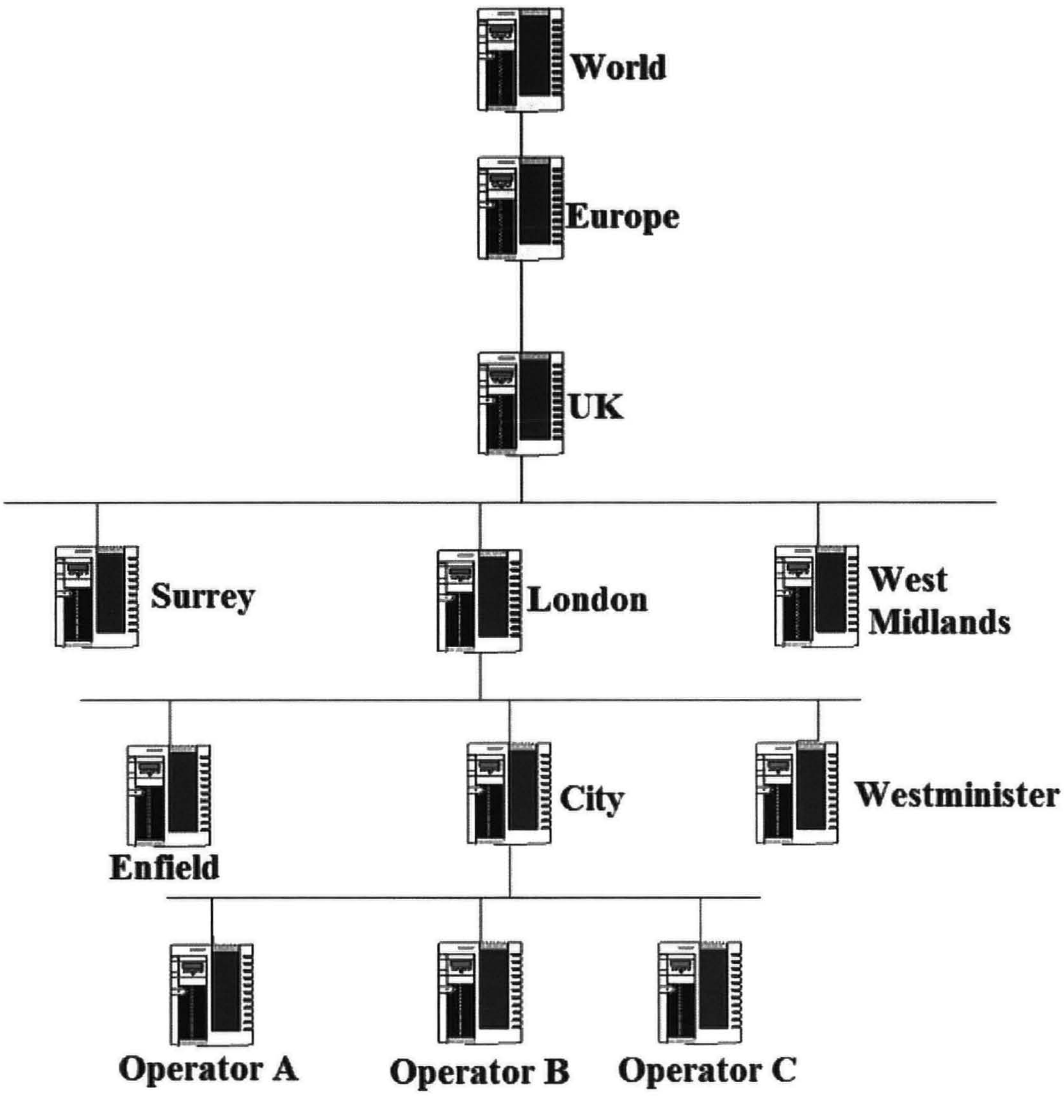


Figure 3-2 - Zonal database distribution

The zonal databases would cater for standardized services would work in conjunction with the operator run databases providing for their own standard and specialized services, as shown in the Figure above. Operator databases would enable the user to explore other services offered by its operator. Distributed repositories would enable much less information being stored on the central database.

Different levels of hierarchies would hold a varied base of information. The table 3-1 outlines data to be held on the databases –

Table 3-1 Database hierarchy

Database	Contents
World	Links to all the continent databases
Europe	Links to all the European country databases
UK	Operational details of all the standardised services of UK wide operators
	Links to all the County databases
London	Operational details of all the standardized services of London wide operators
	Links to all the Zonal databases
City	Operational details of all the standardized services of Zone 'City' wide operators
	Standard and Specialized details of small operators
	Links to the databases of large operators
Operator B	Standard and specialized services of operator B

The varied levels of hierarchies would enable the user in locating service providers in a foreign territory. The top level hierarchies and user' subscribed to providers can be pre-configured on the subscriber identity module (SIM) of the terminal.

3.2.1.2 Location based service update

The current 3G specification [Naghian, 2001] enables the retrieval of location co-ordinates of the terminal by the network or the terminal. Terminals can extract the Global Positioning System (GPS) information from the Radio Network Controller (RNC) or the Servicing Mobile Location Controller (SMLC). The terminal can gather the information periodically or on demand. MN would be able to retrieve the zonal information and its co-ordinates in the zone.

Location based information retrieval would enable the MN to greatly reduce the information overload by only retrieving what it requires. MN would specify its location and coverage area it seeks when retrieving information.

Although location information is very valuable it puts extra strain on the database server by having to calculate a terminal's position with reference to the operator's access point. To reduce such calculations one would only allow location tracking for smaller operators rather than all the operators. Any operator providing services throughout the zone would be termed as large operators whereas operators catering for smaller areas would be termed as smaller operators. Larger operators would have their location based services listed under smaller operators and would be available for a location based search. For example, Operator A provides 3G voice, video and data services but also provides a high speed Wi-Fi Internet service in Heathrow Airport Terminal 4. In this scenario operator A would have two entries in the central database but only the Wi-Fi service would be available for a location based search.

A terminal's location from the Access Point can be worked out using the triangulation algorithm [Jonge, 2005], Eq 3-1.

Distance between terminal and access point =

$$\sqrt{(X'-X)^2 + (Y'-Y)^2 + (Z'-Z)^2}$$

Equation 3-1 - Triangulation to compute distance

where X, Y, Z are the co-ordinates of the terminal and X', Y', Z' are the co-ordinates of the terminal.

The distance between the access point and terminal enables the repository to analyse whether the access point is within the coverage area requested by the MN. To reduce the burden on the database distance, calculations would only be performed when the operator fulfils all the other requirements of the user, which will be discussed in the subsequent sections.

3.2.1.3 Quality of Service (QoS)

QoS is defined as a "collective effect of service" performance, which determines the degree of satisfaction of a user of a service [Wisely, 2001]. QoS in itself is a collection of traffic engineering, policy management, QoS middleware, session management, network QoS mechanisms and signalling mechanisms [Wisely, 2001].

QoS is classified into –

1) Best Effort - This is the most basic level service available and does not include any kind of policy based handling of packets. Best effort [Cisco, 2001] does not employ any service level agreements and these packets would be given the lowest priority in the network. Best effort relies on the capability of the network to route the packet to its destination. Best Effort is associated with an applications requiring no QoS like file transfers or e-mail. No special provisions are taken for moving hosts.
2) Integrated Services – Integrated services maintain an end-to-end QoS state by negotiating the QoS on the end-to-end path. Integrated services are a per flow behaviour QoS type.
3) Differentiated Services – Differentiated services do not define a standard per flow QoS state but rather deals with the packet on a hop by hop basis. Based on the service level agreements, the routers would deal with the packets appropriately.

Quality of service is a measure of various QoS attributes discussed below-

1) **Latency** - The time between initiating a request for data and the beginning of the actual data transfer. Network latency is the delay introduced when a packet is momentarily stored, analyzed and then forwarded.
2) **Jitter** - is when a bit arrives either ahead or behind a standard clock cycle, generally, the variable arrival of packets

3) **Bandwidth** – Bandwidth is the amount of data that can be carried from one point to another in a given time period.
4) **Packet Loss** - Packet loss is where network traffic fails to reach its destination in a timely manner. Most commonly, packets get dropped when the time to live expires before the destination can be reached, usually due to poor (high) network latency. Also traffic gets rejected when the packets fail integrity checks, often due to signal degradation over a poor connection.
5) **Bit Error Rate (BER)** – BER is an error ratio of the number of bits incorrectly received to the total number of bits sent during a specified time interval.

Different application types have varying QoS requirements where such requirements for data, voice and video applications are discussed below-

1) **Data** – Delivering data does not pose as big a challenge as voice and video. Since the Internet Protocol (IP) is an accepted worldwide standard for delivering data, interoperability and transport are easy to support. Data traffic has a high tolerance for delays and latencies in the network. Data packets have a low tolerance for bit error rate (BER), even though higher layer protocols address erroneous bits. High BER affects the throughput of the channel, thus affecting the performance of the channel. Therefore, a data traffic stream is capable of handling varying levels of capacity and latency but requires a low error rate transmission.
2) **Voice** – Voice traffic is handled primarily in a circuit switched manner due to its unique requirements. Voice traffic has a low bandwidth requirement, 64 Kbps in PSTN and 15Kbps in cellular telephony, but is very intolerant of latency and jitter. Industry standard mandates a latency of up to 250ms (1/4th of a second) one way latency as acceptable voice quality. Voice data are not very affected by bit error rate, as dropped or corrupted bits results in noise but does prevent the call from taking place. Latency, jitter and packet loss are voice's main concerns.
3) **Video** - Video is one of the most difficult types of traffic to support. It requires the low latency and jitter performance of a voice call, but places high demands on the data capacity at the same time. In addition, it is also intolerant of errors. It can be argued that video takes the most stringent requirements of both voice and data

traffic. Thus, the successful video transport system will combine low latency with high capacity and low bit error rates.

Voice, video and data applications have different QoS requirements. The table 3-2 outlines the QoS attribute of different types of applications.

Typical application QoS requirements are-

Table 3-2 QoS sensitivity for various application types

QoS Metrics ⇓	**Application Type ⇒**	**Voice**	**Data**	**Video**
Bandwidth		Moderate	Moderate	High
Random Drop Sensitivity		High	High	High
Delay Sensitivity		High	Low	High
Jitter Sensitivity		High	Low	High
Bit Error Rate		High	High	Moderate

Looking at the table 3-2, it is evident that QoS attribute sensitivity varies significantly from application to application. The attributes would vary whether the application is a real-time application or not. Wide ranging requirements for different applications makes QoS very complicated.

Difficulties in achieving good grades of QoS -

- Cost/Complexity/Strength of QoS Compromise [Wisely, 2001]– better grades of QoS can be achieved by providing users more capacity than they would ever require but that would be unrealistic due to the cost factor. Standard telephone networks achieve high levels of QoS by placing restrictions on the type of applications that can be supported.

$$QoS \propto Complexity \propto Cost$$

Equation 3-2 QoS relationship with complexity and cost

- QoS Co-operation [Wisely, 2001]– The fact that in IP networks packets could travel through a range of networks makes co-operation increasingly important to achieve high levels of QoS. A weak link in the chain could degrade the experience of the user perceived QoS. Also, excellent network performance might not be helped by the user running many applications simultaneously and degrading the level of experienced QoS.

The users (or applications) perceive QoS based on the requirements rather than the various QoS classes or sub classifications. Users would have little knowledge about various QoS technical jargon, therefore it is proposed to document QoS in the service advertisement data repository as a combination of bandwidth, packet drop, latency, jitter and bit error rate.

To aid the selection process the followings are recommended:

1) Minimum bandwidth/User offered by the network
2) Average dropped packets/Million
3) Average latency – The average delay experienced by the users in the network.
4) Average jitter – The average jitter experienced by the users in the network.
5) Average bit error rate – The average bit error rate experienced by the users in the network.

This approach relies on the honesty of the operators to advertise true information to the users. Another approach is to have the user calculate all of the QoS performance metrics mentioned above, but the approach is marred with several drawbacks –

1) Mobile terminals have limited resources and would end up spending considerable time carrying out these calculations as well as the fact all the other values are dynamic and therefore would have to be refreshed by the user at regular intervals.

2) Mobile IP's main drawback is the amount of control data on the networks. If all the users started maintaining their own performance metrics, the networks would have even more control information flowing through the network.
3) The terminal would have to maintain QoS performance metrics for every operator available, which could be several in certain areas.

To deal with the issue of honesty the user can maintain a feedback (or the past experience) of the operators it has used services from. The user can then discard operators with an adverse rating. Maintaining a feedback mechanism would be much more convenient and manageable than maintaining local QoS performances. Feedback based selections will be discussed in more detail in Chapter 4.

3.2.1.4 Cost

The cost of using a service is often ambiguous because of the varying units. Some operators charge per min whereas some charge per byte downloaded for browsing services. To enable users to compare service provider's charges, we propose to break this information as-

1) Charge Model – Charge model would hold information regarding the cost policy of the Provider. For example, per min, per sec, per byte downloaded, etc.
2) Cost – Cost would hold cost value per unit of the charge model.
3) Usability period prices - Sliding scale prices vary depending on the time the services are used for [Virgin Mobile, 2005]. The user might get an attractive price to start with but if the price increases with usage the network would lose its cost benefit and may not satisfy user's requirements and vice versa, the cost of using the service could drop. To safeguard the interests of the users, cost would be documented on the database as usability chunks of the charge model. The usability chunks proposed are 1, 5, 10, 50 and greater than 50. The chunks represent the usage against the charge model, e.g. 5 mins call at 10 pence per min would be documented in the database 50 pence.

3.2.1.5 Conceptual Design of the Database

The MOSSA service advertisement database would be categorised by the main service type's data, voice and video. The data would also be subdivided based on operator type: large or small. The main difference between large and small operators would be location tracking. Only zonal databases would allow location tracking and accommodate smaller operators. Larger operators would list out details relevant to the zone, e.g. operator's frequency band is 1.8 GHz to 2.2GHz but only uses frequencies 1.90 to 1.92 GHz in a particular zone [Ofcom]. It would reduce MN's scanning overhead if unused frequencies in the zone are removed from the information passed to it. Zonal databases will only allow location tracking for smaller operators but operators can choose to provide location tracking on their databases. Location based tracking can also be a complement to predictive handover schemes [Naghian, 2003], if any such handover management policy is implemented by the operator.

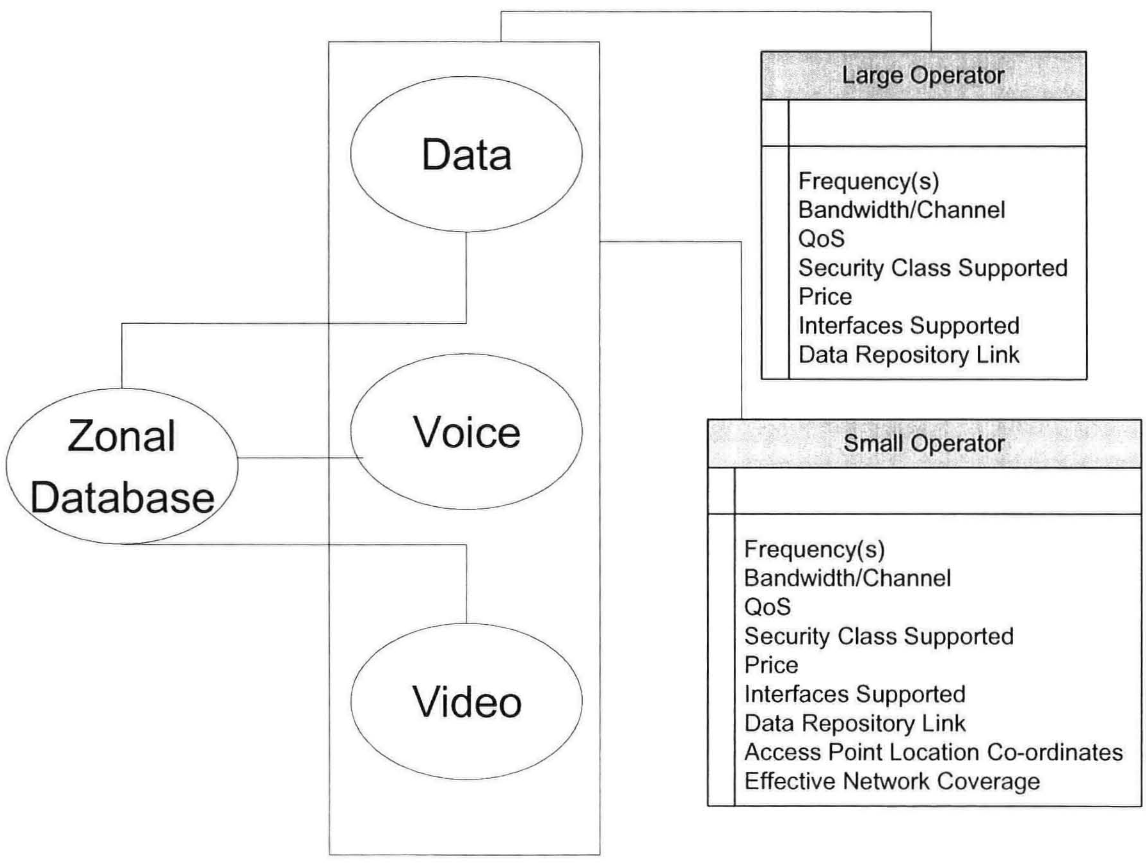


Figure 3-3 Conceptual framework of the database

Figure 3.3 provides a schematic diagram of the conceptual framework of the MOSSA service advertisement database. The first level of classification is the service type and the second is based on the coverage of the operator. Tables 3.3 and 3.4 provide a dummy record for the large and small operators. Operators would have to have a separate entry for every different combination. For example, A different QoS, Security and Price combination would have a bearing on a user's service selection algorithm. Operator's could also provide the IP address of its mobility anchor point (or zonal head) responsible for registering and authenticating a terminal. The MAP details would enable MN to register with the network before it hands over its services to the network.

Table 3-3 - Dummy record Large Operator (*All the values are for demonstration purposes only and may not be true)

Field	Value	Description
Provider	Vodafone	Operator's name
Provider ID	12775	Operator's ID number
Frequency	3GHz	Usable frequency
Bandwidth	100 MBps	Available bandwidth
Service Description	Browsing	Description of the service
QoS	1 (Standard)	QoS measures
Security	1 (Standard)	Security measures
Price	15 Pence/Minute	Cost of using the service
Interfaces Supported	UMTS	Access mechanism supported
Data Repository	Repository.enfield.vodafone.co.uk	Location of the operators repository
MAP IP	10.25.146.71	Location Mobility Anchor Point / Zonal head

The table 3.4 provides a dummy record for the small network operator. The table contains more detail as compared to the large operator table. The extra details mainly relate to the location tracking and coverage area.

Table 3-4 - Dummy record Small Operator (*All the values are for demonstration purposes only and may not be true)

Field	Value	Description
Provider	Starbucks	Operator's name
Provider ID	10880	Operator's ID number
Frequency	11MHz	Usable frequency
Bandwidth	11 Mbps	Available bandwidth
QoS	0 (Best Effort)	QoS measures
Security	0 (Best Effort)	Security mechanism
Interfaces Supported	802.11b	Supported access mechanism
Service Description	High Speed Internet	Description of the service
Price	5 Pence/Minute	Cost of using the service
Location co-ordinates	101,501,1700	Location of the access point
Data Repository	Repository.enfield.starbucks.co.uk	Location of the operators repository
Coverage Area of Access Point	50m	
Coverage area of network	75m	The area of the biggest circle possible with the access point as the centre
MAP IP	10.25.146.71	Location Mobility Anchor Point / Zonal head

Along with location co-ordinates of the network, Table 3.4 lists the coverage radius of the access point and the coverage area of the network. The coverage area of the network is not the absolute coverage area of the network because the radius of the coverage area of the network might provide misleading information to the terminal. For instance in the Figure 3.4, R is the radius of the network coverage and r is the radius of access point coverage area. Providing R would not provide any useful information to the MN because it would not be able to relate it to the centre of the network whereas comparing r and R' would give MN a rough idea of guaranteed coverage in any direction if it moves out of the access point coverage area by the same network. R' is the biggest circle in the network that can be drawn with AP's centre as its centre.

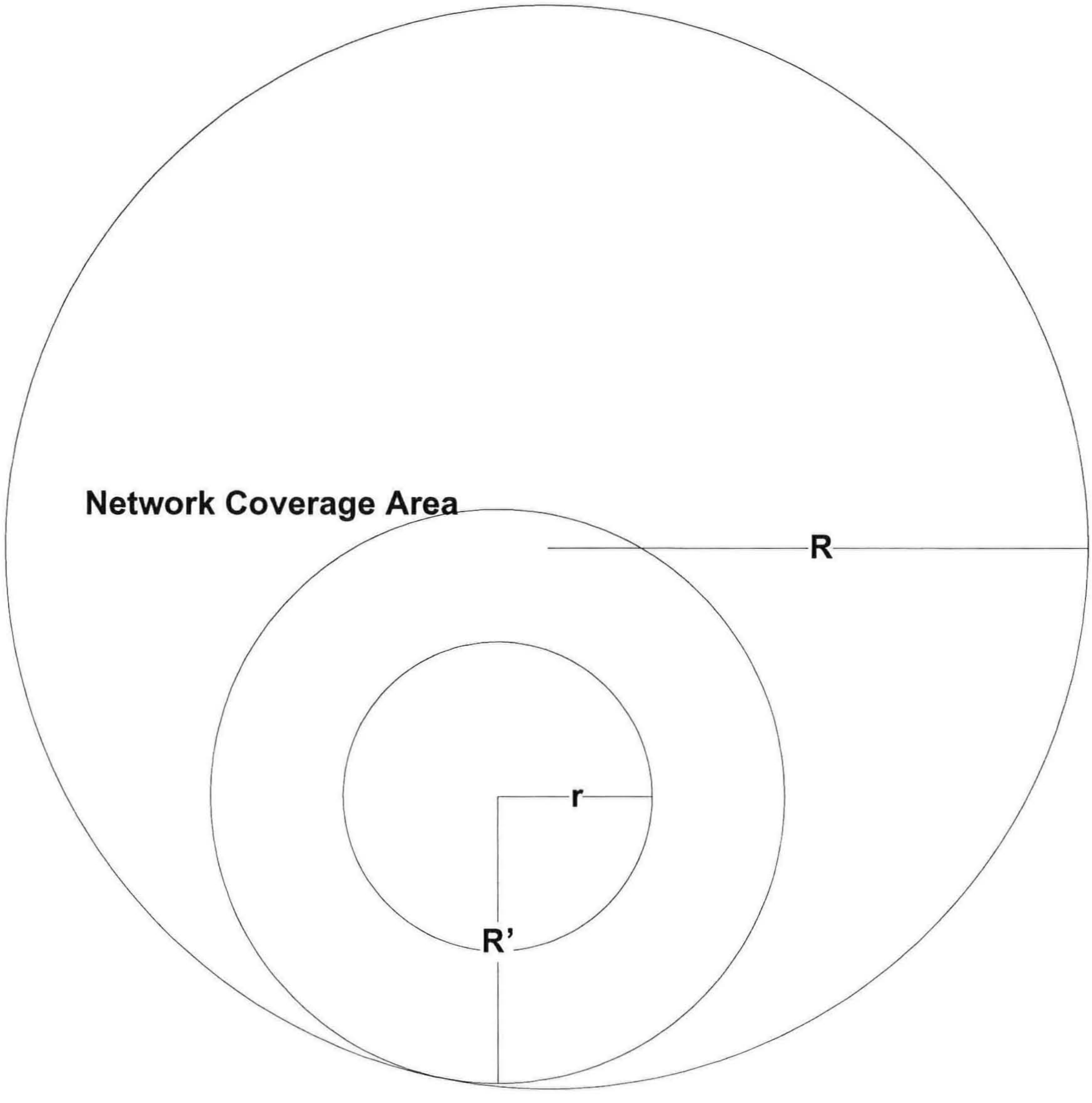


Figure 3-4 Effective coverage area of the network

Smaller networks could list as a network or have each of its access points listed separately based on their network topology and coverage area.

3.2.1.6 Query based on Mobile Node requirements

The user would be able to specify exactly what they are looking for and search the database more efficiently. Users would be able to query the database based on any combination of fields as listed below.

1) Interface – The terminal can specify the interface type in query.
2) QoS – The user/application can specify what QoS class it is looking for.
3) Security - The user/application can specify what security class it is looking for.
4) Price – The users can specify the price band for the service they desire to use.
5) Service description – Users can search for special service like the U2 concert live broadcast, etc.
6) Location – Smaller networks are searchable based on their location and network coverage area.

The users can search the database in any combination of fields but are discouraged to base their searches on QoS, security and price because if what user requires is not exactly matched by any network, the user's service selection program [Bansal, et al., 2004] might have to make some tradeoffs and hence would require details of the operators providing that service.

3.2.1.7 Performance Analysis of MOSSA Data Repository

The database advertisement mechanism would be affected by the delays introduced by the access mechanism employed by the terminal and the database query execution and results retrieval.

Database access Latency =

$AccessMechanismDelay$ + [Delays introduced by the access mechanism]

$DBExecuteQuerry$ + [Query execution delay]

$DBresultsRetrieval$ [Delay in retrieving the records from the database]

Equation 3-3 - Database access latency for the MOSSA Data Repository

Equation 2 gives the total delay introduced by the service discovery. The data repository does not have any control on the delays introduced by the access mechanism which includes transport latency. The database delays include the time taken to execute the query and do all the necessary calculations like location based information retrieval.

3.2.1.8 Test Scenarios

To test the efficiency of the database the following test scenarios have been devised –

1) Number of parameters against time – Since the user is allowed to narrow down his/her search based on their exact requirements; the queries would vary in terms of parameters. The total number of parameters on the MOSSA service advertisement database permitted for searching are 14. The first test is to identify delays introduced due to the change in the number of search parameters.
2) Location based search – The test here is to analyse the delays introduced by the location tracking feature when in use and when not in use.

A snapshot of the tested database (Appendix A) and the tested scenario search queries (Appendix B) is provided. To limit the influence of access mechanism factors in the evaluation, the test was carried out on one machine acting as both client and server. The configuration of the test machine was as follows –

1) Pentium IV 1.6 GHz
2) 256 MB RAM
3) Microsoft Access XP
4) Visual Studio .Net

Refer to Appendix A for sample data, queries and supporting programs.

3.2.1.9 Results

The effect of the number of search parameters is reflected in the Figure 3.5. The latency does not increase significantly with the increase in the number of search queries. Hence, the users should not be discouraged by having to limit the number of search parameters.

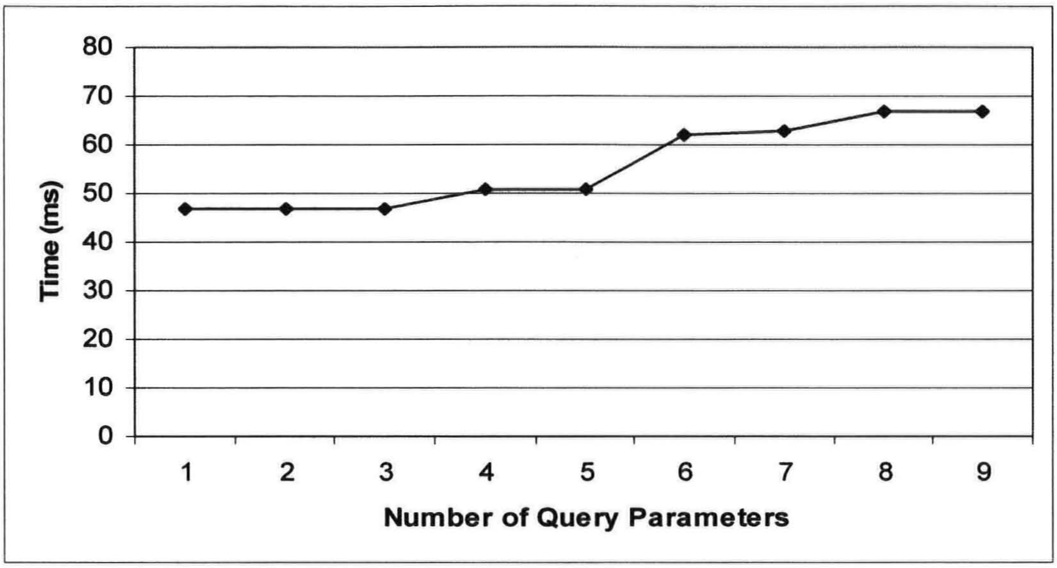


Figure 3-5 - Latency against the number of search parameters

The effect of location based search latency is illustrated in Figure 3.6. The results suggest that for every 10 operators it takes on an average 15 ms for the distance to be worked out between the terminal and the access point. The load of location based searches would increase with the number of users as well. The load of location calculation could be shared between the terminal and the data repository server. The graph is very linear and the latency does not increase as more and more distance calculations are performed.

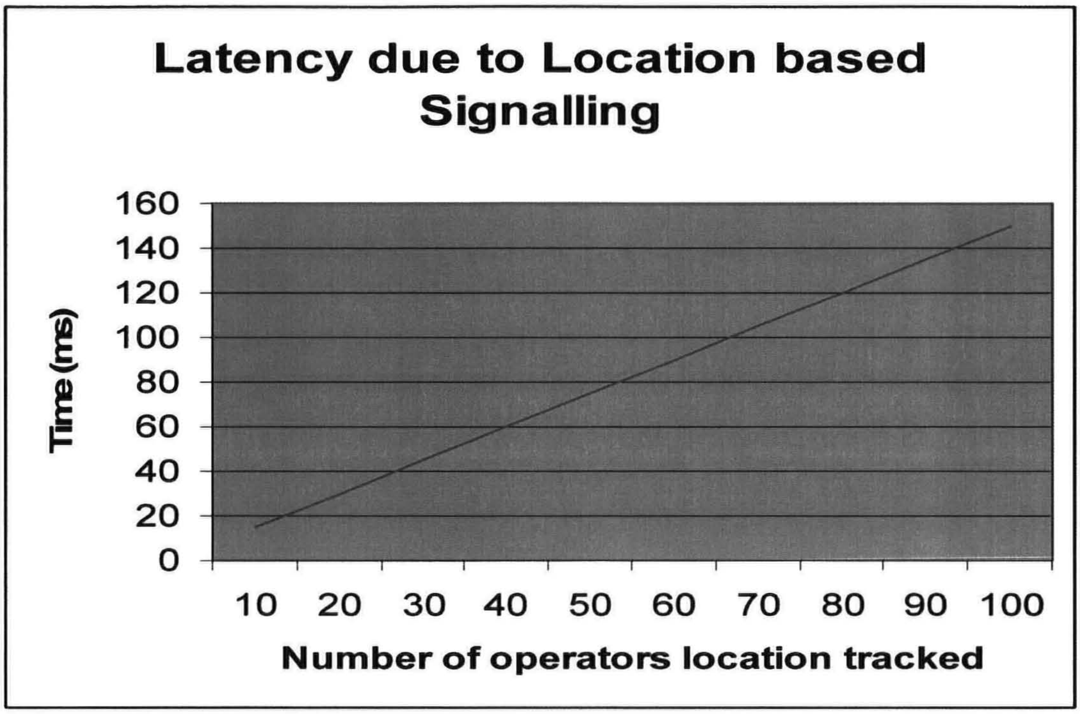


Figure 3-6 Latency due location based signalling

3.2.1.10 Drawbacks

1) The users should use the database reasonably to make sure everyone gets fast and efficient access. The users should generalise their searches to enable their selection algorithms [Bansal, et al., 2004] enough information to find suitable alternatives if users' requirements are not completely met by any operator. The database use might have to be charged to enforce users to generalise their search rather than having to make repeated searches.
2) The repository raises the issue of management. Operators should have editable access to their own records and the users should have a non-editable access to all records. The database would have to be managed by a neutral body to enable fair access to everyone.

3.2.2 Access Mechanism – The Frequency Band method

All the operators can advertise their services using a reserved frequency band. The service discovery frequency band would limit the range of frequencies to be scanned by the users. The reserved band would enable the operators to effectively deliver their service advertisements to the terminals without straining the resources of the terminal. The frequency band would have only a broadcast channel for mobiles to download the advertisements via the Router Advertisement (RA) packets. The frequency band method would not allow the terminals to access the data repository using its frequency. The frequency band could publicize the details of the frequency band method but would not allow the terminals to access the repository on its frequency.

All operators can reserve parts of the frequency band for their advertisements as in Figure 3.5. Some smaller operators might be discouraged by using the frequency band because of the cost involved. To provide consumers with a full range of choices, operators using the frequency band method would make available the location of the zonal data repository, thus making sure that the users can explore the services provided by the operators not using the advertisement broadcast channel via the repository. The service discovery access mechanism can be used in the absence of the repository but the two complement each other very well.

Table 3-5 - Service discovery frequency band allocations

BT	Starbucks	Vodafone	Orange	MMO2

3.2.2.1 Why just broadcast?

Unlike the MIRAI architecture, the proposed solution does not provide any common control and signalling architecture. It offers service providers a broadcast channel for service advertisements. Setting up a Basic Access Network (BAN) [Gang Wu, 2001] would be like setting up an access network in parallel to the already existing access networks as well as raising the issue of maintaining the operator independent BAN, which would eventually increase the level of complexity and the cost of ownership.

Using the broadcast frequency channel raises the issue of regulating the frequency but the issue was raised in the MIRAI architecture as well. The proposed solution passes on the burden of maintaining the service advertisement network on to the operators whilst still maintaining the operator independence. The solution utilises the already setup access networks rather than setting up a new one.

The proposed solution has the following advantages -

1) **Reduced cost of ownership**
2) **Reduced complexity**
3) **Use of existing access networks**

3.2.2.2 How does it work?

The service advertisement information periodically sent by the operator would contain the following information-

1) Zonal data repository
2) Operator data repository
3) Operator standard voice, video and data services along with their characteristics and access details.

Operators would publicise the details of their standard services along with the details of their access networks. Scanning the frequency band would enable the terminal to collect the details of standard services provided by most of the operators. The user can explore the specialised services of the operators by querying the data repository of that particular operator. The services provided by other operators can be explored via the zonal data repository which is also provided within the service advertisement.

The operators usually setup the broadcast channels networks in the same manner as they would setup the wireless access networks. The concepts of frequency re-use, hopping, etc., would all be utilised when setting up the broadcast networks. To reduce the cost, the operators could increase the coverage area of the broadcast stations to include more than one base station as shown in Figure 3.7.

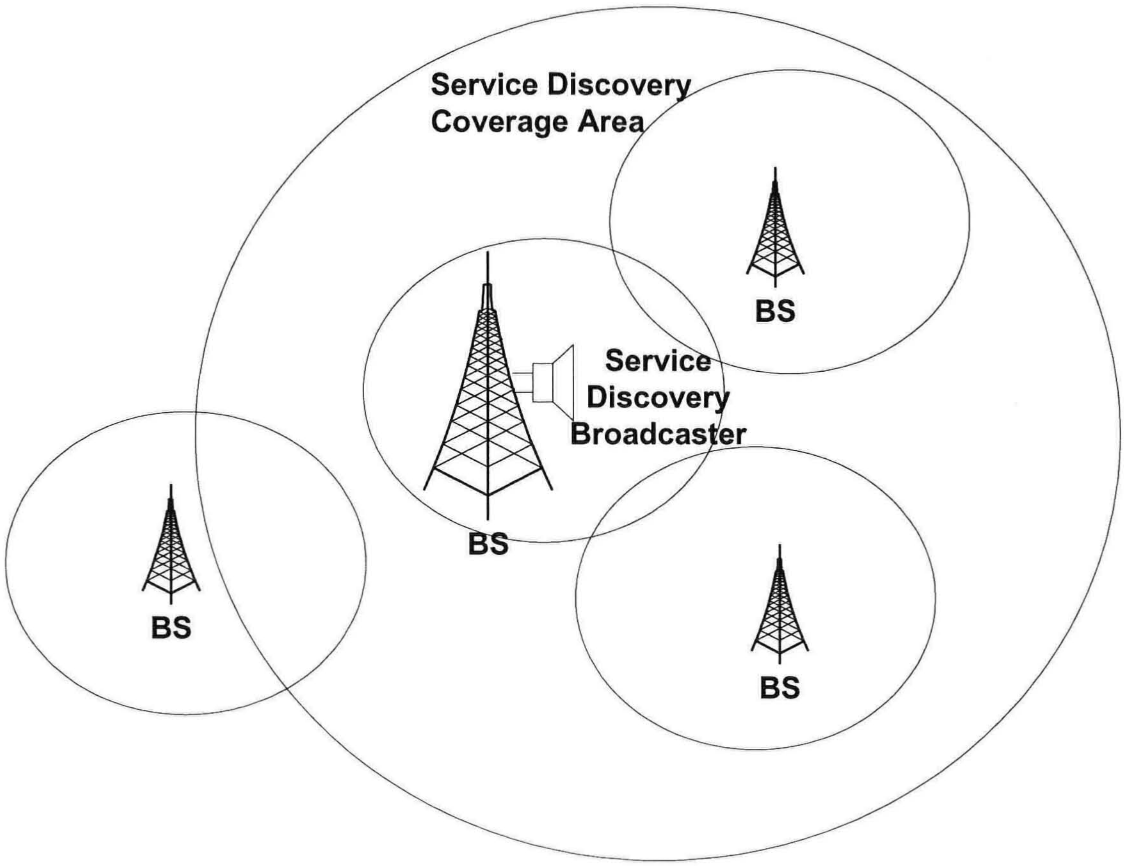


Figure 3-7 - Service Discovery broadcast network

The terminal can scan the frequency band in the absence of main service provider or if it cannot reach any access network.

3.2.2.3 Service advertisements via Router Advertisements

Routers advertise their presence together with various links and Internet parameters either periodically, or in response to a Router Solicitation message. Router Advertisements contain prefixes that are used for on-link determination and/or address configuration, a suggested hop limit value, etc. Service advertisements would be piggybacked on the router advertisements periodically sent by the routers. Since router advertisements are periodically sent by the routers, there wouldn't be major changes required in the Mobile IP operation and no need for extra control packets. The broadcast channels would periodically broadcast this information. The table below gives a layout of the router advertisement packet with most commonly used option prefix.

Table 3-6 - Router advertisements prefix option with Service Advertisements

8 bits	8 bits	16 bits				
Type	**Length**	**Prefix Length**	**L**	**A**	**R**	**Reserved**
Valid Lifetime						
Preferred Lifetime						
Reserved						
Prefix (128 bits)						
Service Discovery Payload						

Type	134
Code	0
Checksum	The ICMP checksum
Cur Hop Limit	8-bit unsigned integer specifying the maximum number of hops
M	1-bit "Managed address configuration" flag

O	1-bit "Other state full configuration" flag
Reserved	A 6-bit unused field. It must be initialized to zero by the sender and must be ignored by the receiver.
Router Lifetime	16-bit unsigned integer specifying the lifetime associated with the default router in units of seconds.
Reachable Time	32-bit unsigned integer specifying the time, in milliseconds, that a node assumes a neighbour is reachable after having received a reach-ability confirmation.
Retrans Timer	32-bit unsigned integer specifying the time, in milliseconds, between retransmitted Neighbour Solicitation messages.
Prefix Information	These options specify the prefixes that are on-link and/or are used for address auto-configuration.

The service providers would decide on the service advertisement payload with the maximum allowed payload size being 65535.

3.2.2.3.1 Performance Analysis

Scanning the single frequency band to listen to advertisements would be the delay introduced by the frequency band mechanism. The frequency band would work like a radio reception service where the terminal would tune into different radio frequencies for service advertisements.

Frequency band Handover latency = $(\sum_{Carrier=0}^{N} Downlink_{Delay})$

Equation 3-4 - Frequency band handover latency

Downlink delays represent the time taken by the terminal to tune into a frequency and download the advertisement.

3.2.2.3.2 Test Scenario

Due to the limitations imposed by the available equipment the maximum number of service providers was limited to 13. A 3Com 802.11b access point and a LAN PC card were used for the tests. A laptop with an 802.11b card was used to measure the time taken to scan the 13 channels.

3.2.2.3.3 Results

The graph below illustrates the latency introduced by scanning for available channels. The latency increases significantly with the number of channels scanned and so the frequency scans should be limited to scenarios like little or no network coverage and zonal movements. The subsequent sections discuss the ideal use of the frequency based advertisement broadcast system.

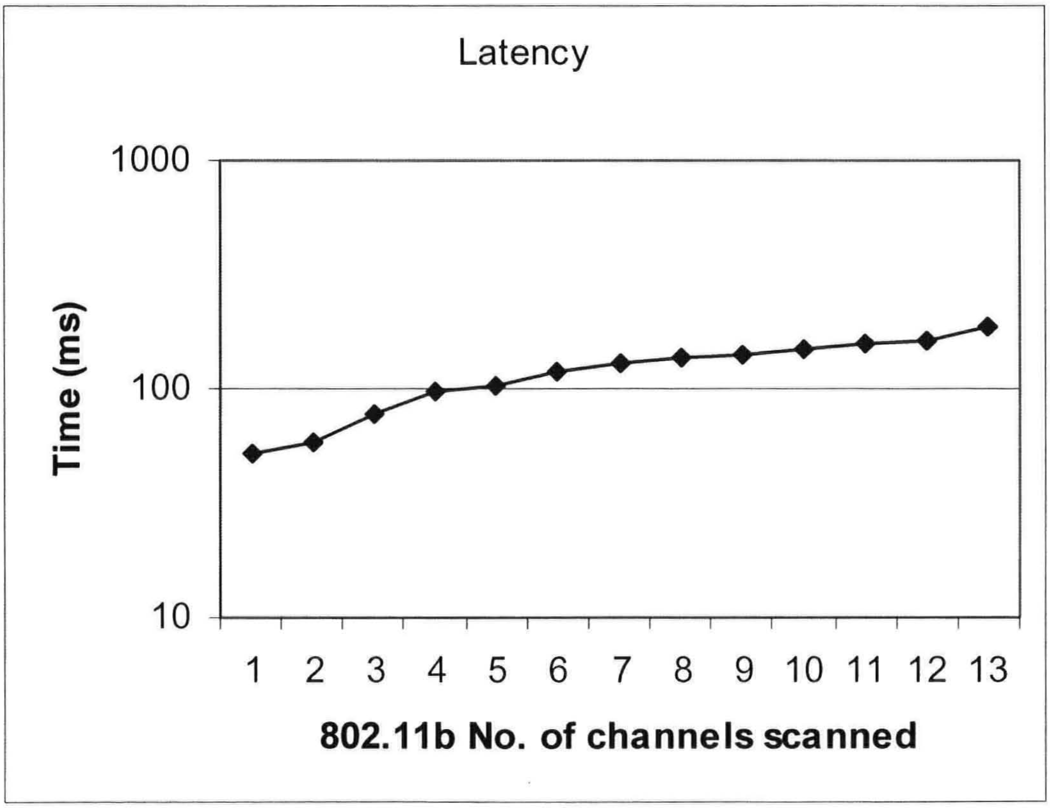


Figure 3-8 - Latency introduced by scanning for available channels

3.2.2.4 Critical Review of the MOSSA reserved frequency band access mechanism

- Transparency – Frequency band option is completely transparent and gives the control of advertisements to the operators. MIRAI provides a common signalling and paging platform transparent for all the operators.
- Regulation – The frequency band method raises an issue of regulating the frequency band and similar concern is raised in MIRAI. MIRAI proposes to install a common signalling and paging network accessible by terminals, thus the issue of regulating the common network and frequency is raised. The frequency band option does not propose a common network so the only issue to be tackled is the regulation of the frequency band.
- Cost of Ownership – Frequency band option would increase the cost for service providers who would have to maintain and install their own advertising equipment. However, the cost of ownership could be reduced by equipment sharing. The cost of ownership of the MIRAI would increase by having to install a common network, which is significantly higher than the broadcast equipment.
- Delays introduced – Frequency band operation introduces scanning delays of the whole band of frequency and then downloading the relevant advertisements. MIRAI would introduce delays by having the terminals access the BAN (Basic Access Network) for operator discovery.
- Data Repository – The frequency band option was designed to co-exist with the data repository whereas MIRAI does not explicitly specify any use of a data repository. However, the data repository should work with the both the access mechanisms.
- Management – The frequency band option passes the obligation of management on the operators whereas MIRAI requires a neutral independent body to maintain the common network to ensure fair access.
- Control messages – One of the main drawbacks of Mobile IP is the amount of control messages being sent around the network. The frequency band broadcast system does not increase that load significantly by just allowing broadcasts

whereas MIRAI increases the load significantly. MIRAI increases the signalling load by setting up a common network, which communicates with the access networks on a periodic basis.

3.3 Service Discovery using MOSAA approach

The frequency band broadcast system concludes that the delays introduced could be significant; hence it is proposed to use the broadcast system only when there is little or no network coverage and the terminal handovers within the zone. This would limit the effect of the latency introduced by the frequency band broadcast system. The flowchart in Figure 3-9 demonstrates the MOSAA approach.

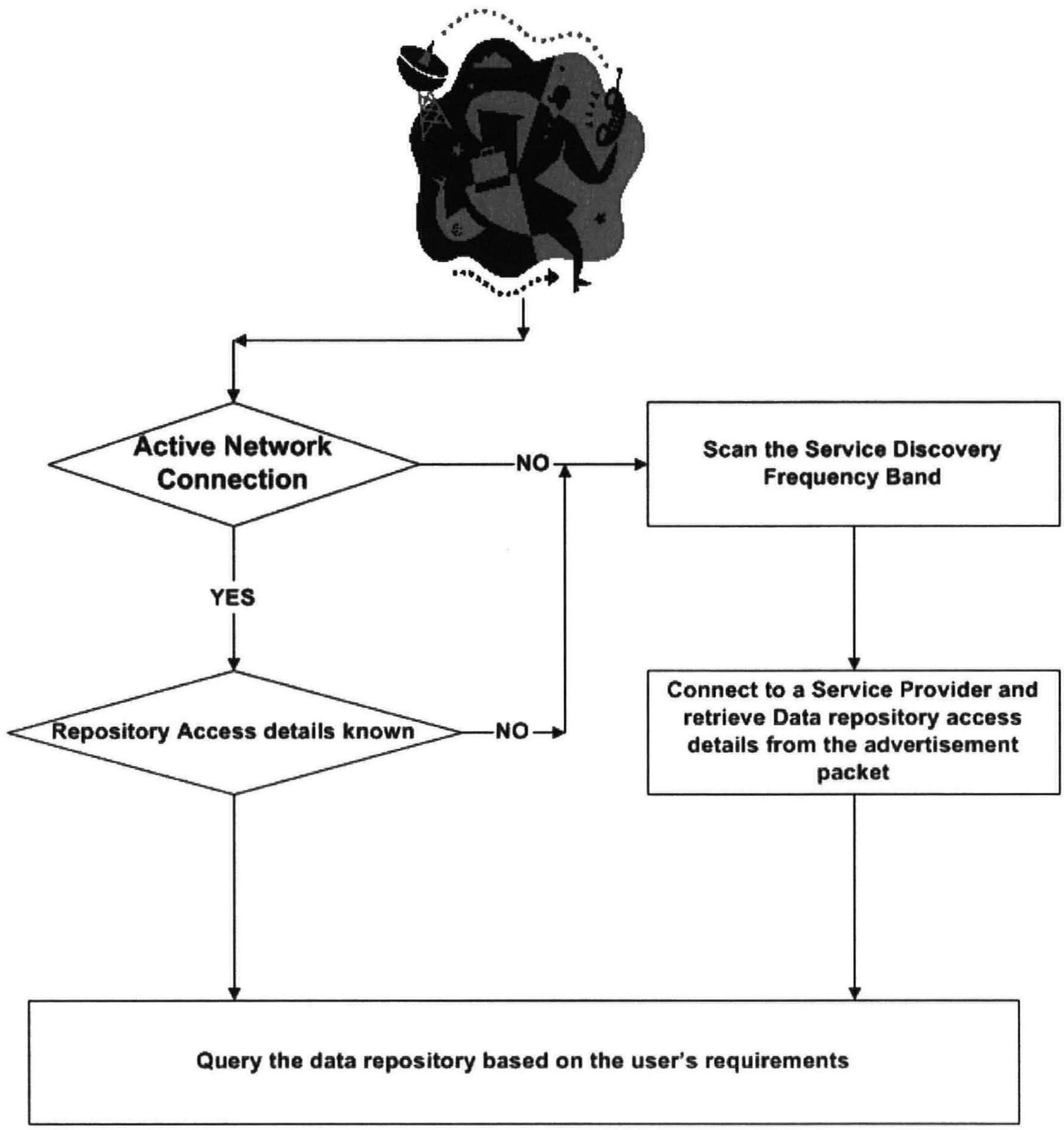


Figure 3-9 - Service advertisements using our approach

3.4 Summary

The service advertisement mechanism would form the basis of multi-operator multiple access type networks. This chapter presented unique service discovery architecture for the next generation heterogeneous networks. The architecture defined a data repository along with an access mechanism. These two mechanisms provided a platform that enabled terminals to discover operators and services within an area. The architecture was a novel idea that proposed location based service advertisements.

3.5 Contributions

1) Requirement analysis for the service advertisement mechanisms. Requirement analysis is the first of its type to be carried out.
2) Layout of the hierarchical service advertisement database system is the first of its kind.
3) Frequency based broadcast advertisement system is also an outcome of the study.
4) The study is the first one to propose a location based service description retrieval system.
5) The complete architecture for a service advertisement system is also a contribution of the study.

Chapter 4 Service Selections Expressing User's Preference

Heterogeneous access across networks based on IP transport technologies are the characteristics of the Fourth generation mobile networks [Mobile Info, 2001]. Network selection based on user preferences requires seamless handover protocols but also an efficient service advertisement and selection algorithm [Bansal, et al., 2004]. Supporting multiple operators would introduce further delays due to the increased scanning overhead [Bansal, et al., 2004]. To minimise the scanning overhead yet provide users with the freedom of choice, requires more efficient service advertisement and selection mechanisms [Bansal et al., 2004]. Upon gathering the service information from various service providers, the user ought to pick the network based on their preferences taking into account the services already in use.

Fourth Generation mobile telephony has promised seamless access across different types of networks [Clapton, 2001]. Different access technologies offer users greater choice in selecting services that can reflect differences in Quality of Service (QoS) support, business models and service providers. Wireless Local Area Networks (WLANs) could be a cost effective complement to 3G systems in urban blanket deployments for multimedia applications and provide users high speed access to the Internet in hotspot environments [Wisely, et al., 2003]. Various wireless standards would inter-work to provide users with a wide range of services delivered on a variety of networks [Clapton, 2001]. Real-time services with good quality should be generally accessible regardless of the technology and the access network, uninterrupted during a handover. Such a heterogeneous design poses challenges in the handover design, service advertisement and call setup. Integration of seamless handovers with QoS and AAAC (Authentication, Authorization, Accounting and Charging) coupled with the possibility of multi-operator scenarios add to the complexity.

Service advertisements discussed in Chapter 3 provided users with service characteristics to select operators. Once a list of available services has been gathered, users can employ the selection algorithm to make handover decisions. Service advertisements give users the flexibility of using gathered information about services offered by all the providers in the area; without applying a dynamic selection algorithm, all the information gathered would be of little use. A service selection algorithm would be deployed by the terminals during –

1) **Service initiation** – When the user initiates a service the terminal would use the algorithm to decide which service provider to use.
2) **Handover** – Handover refers to the process of a mobile node changing its point of association to the network. Handovers can be triggered due to a variety of reasons, as discussed in the subsequent sections.

4.1 Handover classification

Handovers are an integral part of the mobility management framework as discussed in Section 2.2. Handoff refers to the process of a mobile node (MN) changing its point of attachment. Handovers can be sub-classified based upon the reasons for a handover or the networks involved in the handover process.

4.1.1 Classification based on Networks

4.1.1.1 Horizontal handover

Handovers which only affect the link-layer without resulting in a change of IP are horizontal handovers. Horizontal handovers take place within a network with or without a change in the access technology.

4.1.1.2 Vertical Handover

Handovers which affect both the link and IP layers are vertical handovers. Vertical handovers which also result in a change of access mechanism are termed as diagonal handovers. Handovers across networks fall under vertical handovers. Diagonal handovers result from a change in networks and a change in access mechanism.

4.1.2 Classification based on handover cause

4.1.2.1 Operator induced

Operators may handover user services to other networks due to lack of capacity, technical problems, etc. Operator induced handovers offer limited choice or flexibility to terminals.

4.1.2.2 User preferences

Users may switch networks if their preferences are not satisfied by the current network or they find a better deal somewhere else. Handovers due to user preferences are termed as alternative handovers.

4.1.2.3 Imperative Handovers

Imperative handovers are processed when network properties (signal quality, bit error rate, network coverage) fall below an acceptable level. Acceptable level is the threshold below which the connection might become unstable and the terminal might loose the connection. Imperative handovers don't allow the flexibility of hunting around for the best network because the connection properties are already below an acceptable level, hence the terminal risks breaking the connection if it takes too long to initiate a handover.

4.1.2.4 Mobile assisted network controlled / Mobile Controlled

Mobile terminals in 2G and 3G networks send measurement reports to the network to enable the network to make handover decisions for the mobile node. Mobile assisted handovers work well in homogeneous networks but they have some disadvantages. Handovers within a domain do not require any security considerations whilst if the handover is cross-domain, the node would have to be authenticated in the new domain. The service profile of mobile nodes would vary from domain to domain. Accounting and authorisation information would have to be refreshed every time the node is handed over to a new administrative domain, which adds to the handover latency. Network controlled handovers would prevent a user from exercising their right of choice by looking elsewhere other than to the subscribed network. However, mobile controlled handovers require networks to disclose more capability information. Also, networks would need to make provision for service advertisements to provide nodes with a full list of available services.

4.2 MOSAA's elimination oriented policy based handoff's

The algorithms tackling alternative handovers tend to ignore imperative handovers. The proposed algorithm presents a service selection algorithm which differentiates between the requirements of imperative and alternative handovers. The algorithm was designed for mobile controlled handovers not mobile assisted handovers.

The Figure 4-1 [Bansal, et al., 2004] demonstrated the initiation of the handoff process. Handovers would be performed either to

1) maintain connection when network properties fall below an acceptable level, or
2) when a desired preference of the user was not satisfied.

The list of available networks for selection could be obtained from the relevant service advertisement mechanism. An alternative handoff process would only be initiated if the user preferences were not satisfied and there were other network choices available.

Imperative handoffs would only be processed when the network properties fall below an acceptable level.

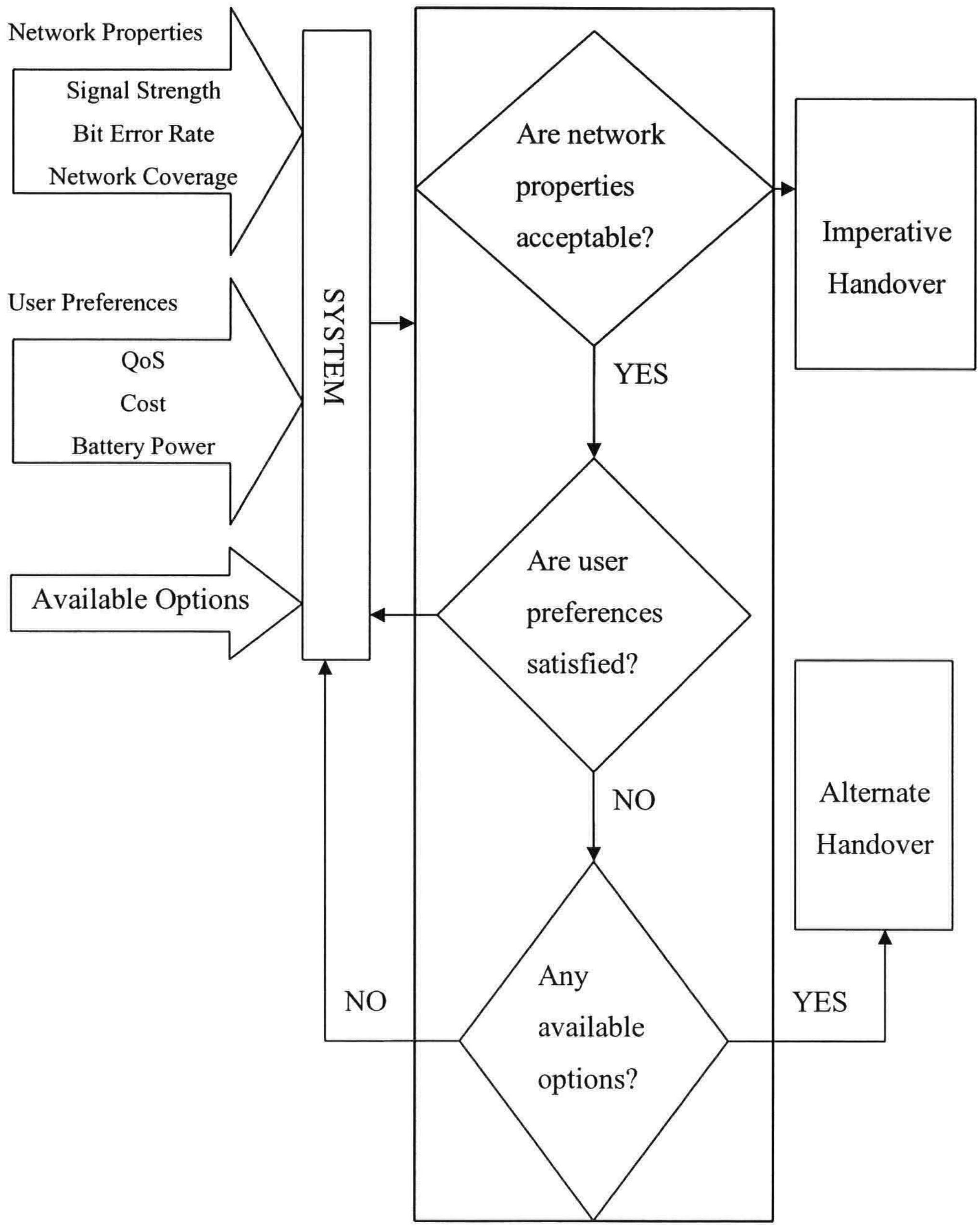


Figure 4-1 Handover Decision Engine

Alternative handovers allow users to exercise their right of choice but add to the handover latency. Handover latency was increased due to the time taken by the selection algorithm to look for an alternative. Alternative handover processing uses valuable resources of the terminal already constrained on resources. To minimise the handover latency and the effect of extra processing, it was proposed that an "Elimination oriented policy based handoffs" be employed. Network selection policies can be sub-categorised into the following-

1) Elimination policies – Selection criteria which cannot be negotiated upon or the user is not willing to negotiate on it would form the elimination policies. For example, Network A offers a high speed WLAN service but the terminal does not have that access interface then it would be pointless considering that network any further.
2) Progressive policies – Progressive policies would constitute on the selection criteria which can be negotiated upon. For example, Network A offers 2Mbps of bandwidth to the user but the user was looking for 3Mbps. Bandwidth could also be used as an elimination policy if the application's absolute requirement is not met by the operator.

Elimination policies would reduce the calculations by eradicating the networks which do not offer the minimum required and then the progressive policies could be employed to find the most suited network.

4.2.1 Imperative Handovers

Imperative handovers are processed when network properties fall below an acceptable level [Zhang, et al., 2003]. Imperative handovers have to be accomplished in a shorter timeframe otherwise the terminal might loose connection due to unacceptable network properties. The imperative handover processor, Figure 4-2 [Bansal, et al., 2004], selects the first network with acceptable network properties. Imperative handovers ignore user preferences when selecting the network, as user preferences can be taken into account once a stable network connection has been established. Imperative handovers are not affected by the user's elimination and progressive selection policies, since their primary objective is to maintain the connection.

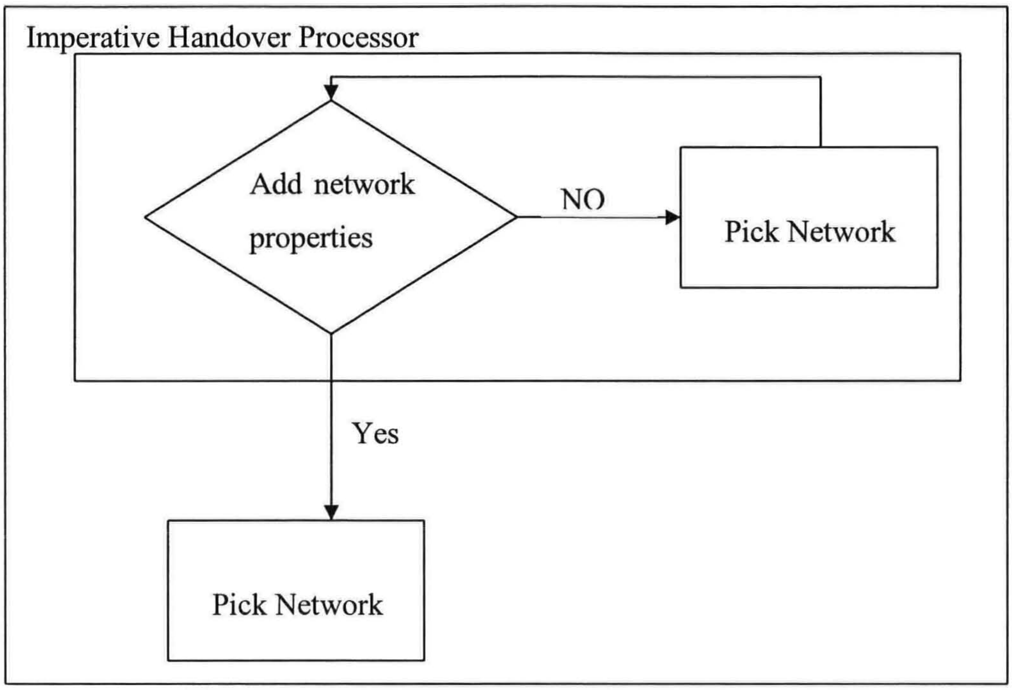


Figure 4-2 Imperative Handover Processor

4.2.2 Policy based alternate Handovers

Alternative handovers are processed when network properties are acceptable but user preferences are not satisfied. Alternative handovers are performed to exercise user's freedom of choice. Policies would be sub-categorized into

1) Elimination policies – Elimination policies would lead the specific network not considered for selection.
2) Selection policies – The selection policies would be used to compare networks based on user's requirements.

4.2.2.1 Elimination policies

4.2.2.1.1 Interface profiles

Various wireless technologies such as 3G are being deployed at an increasing rate. The terminals are often equipped with several network interfaces or are able to connect to these interfaces. The future wireless network interfaces would either –

1) Support one frequency, i.e. be able to connect to one type of network. An 802.11b interface will only connect to 802.11b access points.
2) Support multiple frequencies, i.e. be able to connect to more than one type of network. Most of the readily available 802.11g interfaces are able to connect to 802.11b networks. Future wireless network interfaces would be re-programmable to work at various frequencies and connect to different types of networks [Craninckx, Donnay, 2003]. This implies that even though the user's terminal is equipped with various interfaces, the user would not be able use all of them simultaneously.

It is proposed that interface profiles are used to tackle these issues. Interface profiles would enable the selection algorithm to identify the absolute and relative interfaces available for selection. The terminal would maintain an interface profile for every interface like the one in the table below. The profile would hold information on the various types the interface can be configured to work at, e.g. an interface can be programmed to work as both 802.11b and Bluetooth. The profile would also have details of the frequency for every interface type and the location of the driver file. The pointer to the driver file would be used for re-programming the interface to work as another type. The profile would also maintain service capability for every interface type. The service capability information would be used for selecting interfaces based on service type. For example, user requests for real-time voice service, the algorithm would be establish that interface I2.1 (table below) would be the only suitable interface for the requested service type.

Table 4-1 Interface profiles

Interface	Type	Driver	Frequency	Service capability
I1.1	802.11b	-	11MHz	Browsing
I1.2	Bluetooth	-	11MHz	Browsing, Peer to peer

Interface	Type	Driver	Frequency	Service capability
I2.1	**UMTS**	**-**	**Tbd**	**Real time voice, real time video, browsing**
-	**-**	**-**	**-**	**-**

Having various interfaces available for selection, increases the options of networks the terminals can connect to. However, since wireless interfaces share the common medium, air, for transmission of information, some wireless interfaces because of the shared medium are affected by certain wireless interfaces. The signal to noise ratio of the interfaces is affected when used simultaneously with other wireless devices in close proximity. It is proposed that an interface compatibility matrix is used which holds information relating to network interface type compatibility, like the one below. Referring to the table, the terminal (or selection algorithm) can identify which interfaces can be used simultaneously. Referring to the table below, it can be established that interface type I2.1 and I1.1 have a good compatibility with each other and hence can be used simultaneously.

Table 4-2 Interface compatibility matrix

Interface	I1.1	I1.2	I2.1
I1.1	**Good**	**Bad**	**Good**
I1.2	**Bad**	**Good**	**Good**
I2.1	**Good**	**Bad**	**OK**

Interface based policies would be used to eliminate networks available for selection.

4.2.2.1.2 Location and velocity

A provider network should have an optimum coverage area around the terminal thereby reducing the possibility of frequent handovers [Chan, et al., 2001]. The service advertisement data repository enables users to explore services based on their location.

The users can employ this information to select services. The users could also use the velocity information to search for services. For example, a user travelling at 80Km/h who wants to use browsing services would not be advised to use Wi-Fi services because they will be out of range of the access point very soon. The users can construct policies based on their location and relative speeds.

4.2.2.1.3 Multiple-Operator Support

The providers that do not support handovers to competitor networks would limit users in exercising their rights of choice. Users can construct policies giving preference to networks which support multiple operators.

4.2.2.1.4 Past Experience / Feedback

The user can keep a track of the past experiences with operators and can use these to construct policies. For example, A user can disregard all the operators with whom it had experienced poor performance. Since, most of the information in service advertisement database is not verified by any central authority, the past experience recording would discourage operators from making false claims about their services. Terminals can compare offered characteristics to the delivered characteristics and hold the information for future reference.

4.2.2.2 Network selection parameters

Network comparisons would be based on monetary cost, quality of service and battery power of the terminal. Cost is a primary concern when selecting services for most users [Wang, et al., 1999]. Quality of service would influence the users experience using the service and hence would be considered when selecting a service provider [Zhu, 2004]. Battery power on the terminals is a limited resource and should not be wasted [Wang, et al., 1999]. Hence, battery power qualifies as the network selection parameter. Network selection parameters could also be used for eliminations, as will be discussed in the subsequent sections.

Users would express their importance to each of these parameters for influencing network selection. The combined weighting of all the three parameters would be equal to 1 (or

100%). For example, if the user is looking for the cheapest network then the weighting associated with cost would be equal to 1 whereas QoS and battery power weighting would be equal to zero. The users could also define alternative weightings for each of these parameters. Alternative weightings would only be used when two networks have equal credentials on the primary weighting criteria.

Users would also specify the importance or weights for the power consumption (W_P), QoS (W_{QoS}) and cost (W_{COST}). The parameters which do not concern the user can be equated to zero. The weight parameter helps define the most suitable network for user's requirements. For example, if the batter of the mobile node is running out, W_P should increase dramatically to signal a change in the condition.

$$\text{Total Weight} = W_{COST} + W_{QoS} + W_P = 1$$

Equation 4-1 - Selection parameter importance factor

4.2.2.2.1 Monetary Cost

Cost would be a major consideration for most users in deciding which network to choose. Cost of using a service depends on the charge model employed by the operator. Users could define the maximum charge; they are prepared to pay for using the service along with the weighting as their selection criteria.

Users may have service agreements with some operators which would be taken into account when calculating the cost of the service. For example, a service contract with operator A entitles users on this contract to 200 free national voice minutes every month, which implies that even though there is a cost for using the service, it will be free for the first 200 minutes.

A network would not be considered if the cost of using the service is more than what the user is prepared to pay. The user would also specify importance or weighting for the cost parameter.

The algorithm would first filter the networks based on the user's elimination policy based on cost, Fig 4-3. The user would specify how much they are willing to pay for using the service (Cost_prepared_to_pay). The algorithm would eliminate all the networks over and above that price. The only exception would be made if the user has a special contract with an operator which entitles them to a lower price. In the stated scenario the users agreed price would be used as the price offered by the network.

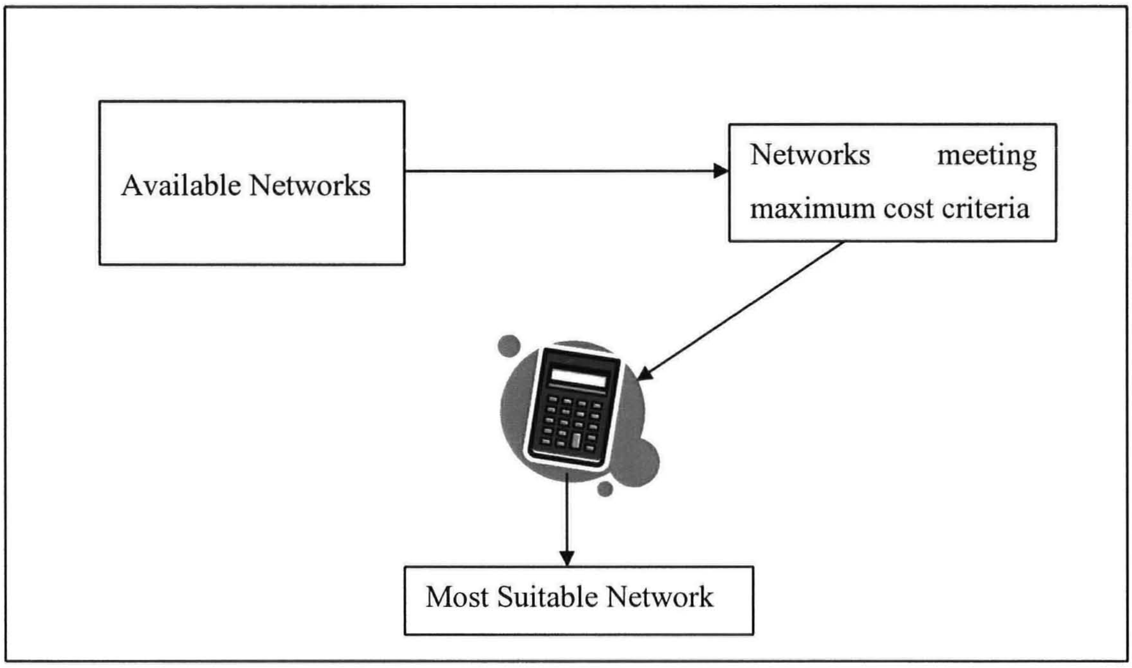


Figure 4-3 Network elimination based on cost

One problem with this approach is that operators might start using a sliding scale for prices to make their services look more attractive to the user. Sliding scale prices vary depending on the amount of time the services are used for [Virginmobile, 2007]. The user might get an attractive price to start with but if the price increases with usage the network would lose its cost benefit and may not satisfy user's requirements. To minimize the effect of marketing policies we propose ***"Service usability period"***. The user would specify for how long it intends to use the service and the algorithm would calculate the cost of using the service for every network over that period and use that for network selection.

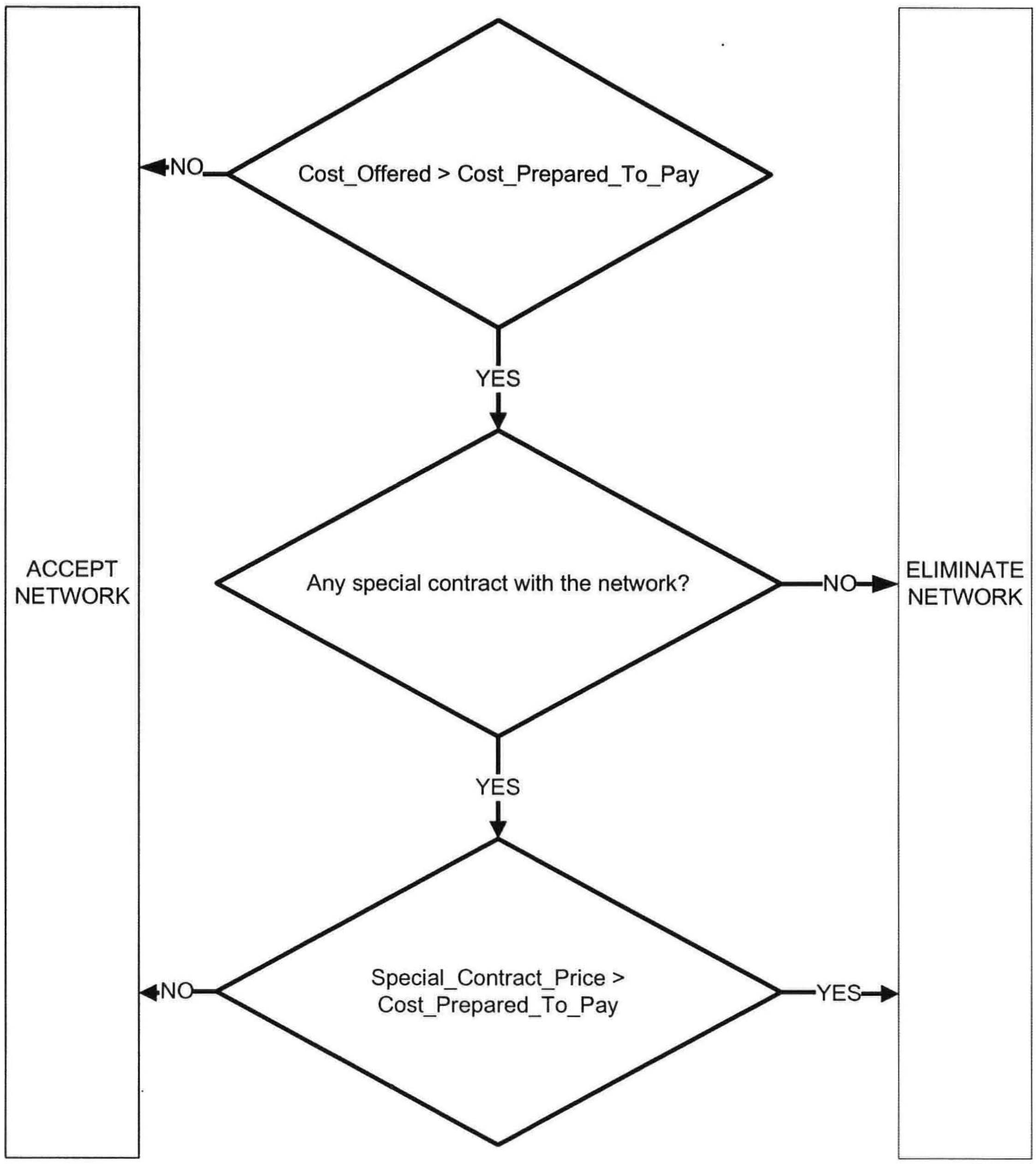


Figure 4-4 Cost Function

Cost offered, cost prepared to pay and special contract price would be worked out over the usability period in the above algorithm. When switching networks the algorithm would reduce the time by the amount of time a service has already been used for.

The cost of using a network, n Eq. 4-2, is a parameter of price of using the service over the estimated usability period. Once the network elimination policy has filtered the list of networks to meet the user's non-negotiable requirements, the algorithm would work out the cost function for every available network.

$$Function_{Cost_n} = f(C_n)$$

Where:

Cost Function $\propto$ 1/ Cost, since lower the cost is better for the user

$$Function_{Cost_n} = W_{COST} * \frac{1}{C_n}$$

Equation 4-2 - Most suitable network cost function

4.2.2.2.2 Quality of Service (QoS)

All the existing algorithms consider bandwidth as the only QoS factor to be considered in making handover decisions. As discussed in Chapter 3, the five attributes that affect most application types are Latency, Bandwidth, Jitter, Bit Error Rate and Packet Loss. Table 3-2 in Chapter 3 outlines the attribute sensitivity of different types of applications.

- Bandwidth – Bandwidth of the channel might be unsuitable for certain applications; therefore the bandwidth of the channel is an important constraint when selecting the channel.
- Latency – Voice and video applications have a low tolerance for high latency.
- Packet Loss – Data applications have a low tolerance for high packet loss.
- Jitter - Voice and video applications have a low tolerance for high jitter.
- Bit error rate – Similar to Received signal strength, if the Bit error rate is higher than optimum, the provider should not be considered for selection.

Hence, the QoS of a network, n, is a function of QoS parameters: bandwidth offered (B_n) Latency (L_n), Packet Loss (PL_n), Jitter (J_n) and Bit Error Rate (BER_n).

The user could use any combination of the five parameters for their QoS requirements. For example, real time voice applications have a delay sensitivity of 200-250ms, bandwidth required 64kbps for PSTN and jitter as low as possible. The user can define the parameters based on the performance they perceive.

The user/application would define their QoS requirements as -

1) Minimum bandwidth required
2) Maximum Packet drop tolerance
3) Maximum Latency tolerance
4) Maximum Jitter tolerance
5) Maximum Bit Error Rate tolerance

The algorithm would identify networks satisfying the minimum requirements of the user/application. Once the list of networks meeting the minimum criteria has been established, the algorithm calculates which network offers the best QoS over and above the minimum requirements.

The algorithm, Figure 4-5 below, demonstrates the network acceptability criterion. The user can define as many of the five QoS parameters as desired and the algorithm would filter out the networks based on the elimination algorithm. The Figure above represents the network selection procedure.

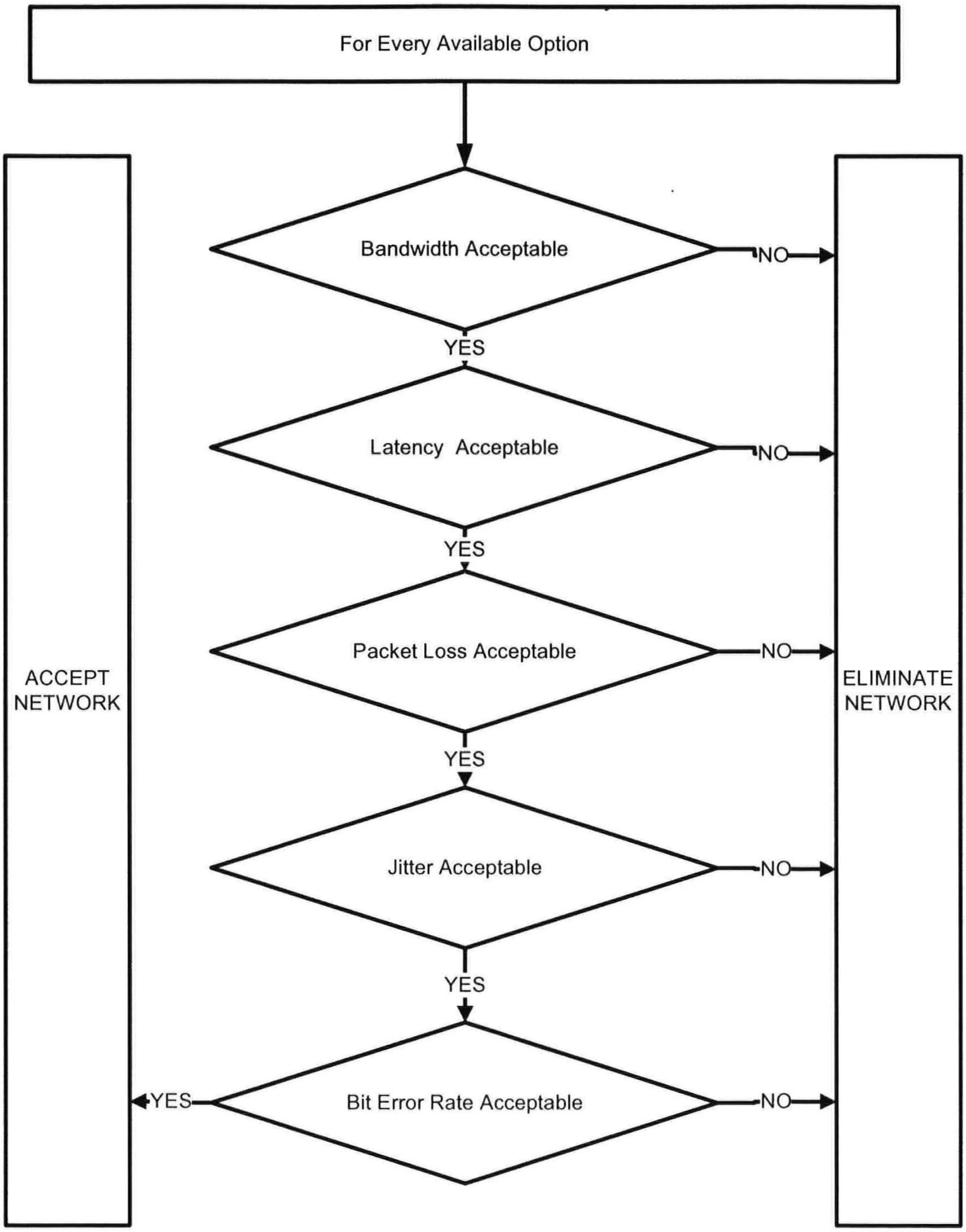


Figure 4-5 Quality of Service (QoS) function

The most suitable network would be calculated from the list of acceptable networks. The user would also specify the importance of the individual QoS sub-classifications along with the QoS weight. The advantage of doing this is that different applications have different requirements for the QoS parameters. The Table 3-2 outlines the main requirements for different types of applications.

Referring to Table 3-2, voice applications would have higher weights for packet loss, BER, latency and jitter. QoS parameter weights would be too complicated for the user to assign, so it should be assigned by the applications or be stored on the devices.

User/application assigned weights for bandwidth (W_B), latency (W_L), packet loss (W_{PL}), jitter (W_J) and bit error rate (W_{BER}) would be used to work out the most suitable network for the service the user wishes to user/handover. The parameters which do not concern the user can be equated to zero. The weight parameter helps identifying the most suitable network for user's requirements.

$$\text{Total Weight} = W_B + W_L + W_{PL} + W_J + W_{BER} = 1$$

Equation 4-3 - QoS parameter importance weights

Individual weights for the QoS parameter helps identify the networks capable of delivering best service.

The bandwidth, latency, packet loss, jitter and bit error rate would be gathered from the service advertisement mechanism or any other mechanism employed by the user/ network.

$$QoS_n = f(B_n, L_n, PL_n, J_n, BER_n)$$

Where:

Quality of Service (QoS) $\propto$ Bandwidth (B)

$\propto$ 1/ Latency (L)

$\propto$ 1/ Packet Loss (PL)

$\propto$ 1/ Jitter (J)

$\propto$ 1/ Bit Error Rate (BER)

$$=> \quad Function_{QoS_n} = W_{QoS}(W_B B_n + W_L \frac{1}{L_n} + W_{PL} \frac{1}{PL_n} + W_J \frac{1}{J_n} + W_{BER} \frac{1}{BER_n})$$

Equation 4-4 - Most suitable network QoS function

The QoS weight (W_{QoS}) signifies the significance of QoS to a user compared to other network selection parameters. The individual QoS parameter weightings enables the networks best capable of providing the service the user wishes to employ, to be identified.

4.2.2.2.3 Battery Power

When a terminal is running low on battery power, providers closest to the terminal would be preferred to reduce the possibility of the terminal running out of battery and thereby disconnecting all the active communication channels [Chan, 2001]. The power consumption is a parameter dependent on the mobile's battery life. The users can also use power consumption as an elimination criterion. For example, if the terminal battery power is low, it can eliminate all the networks which require a higher signal power output.

Since, the lower the power consumption of the phone the better for the user, the power function has an inversely proportional relationship with the power consumed.

$$\text{Power Function} \propto 1/\text{ Battery Power (P)}$$

The user/application could specify the maximum power output, which would be used as the elimination criteria. From the list of acceptable networks, the network with lowest power rating would be calculated.

$$Function_{Power} \propto \frac{1}{P_n}$$

Considering, the importance (or weight) assigned to the power rating of the network, the above relationship would be:

$$Function_{Power} = W_P * \frac{1}{P_n}$$

Equation 4-5 - Most suitable network power function

4.2.2.2.4 Most Suitable network

The most suitable network would be a function of cost, QoS and power rating. The algorithm would first filter out the networks, which do not meet the minimum criteria.

The list of acceptable networks would be used to determine the network most suitable for the user's requirements.

$$Function_{MostSuitableNetwork} = Function_{Cost} + Function_{QoS} + Function_{Power}$$

$$= (W_{COST} * \frac{1}{C_n}) \quad +$$

$$(W_{QoS}(W_B B_n + W_L \frac{1}{L_n} + W_{PL} \frac{1}{PL_n} + W_J \frac{1}{J_n} + W_{BER} \frac{1}{BER_n})) \quad +$$

$$(W_P * \frac{1}{P_n})$$

Equation 4-6 - Function most suitable network

Normalization would ensure the sum of the values in different units is meaningful. If a network offers twice as much bandwidth but is twice as expensive as the other network, then we consider that to be equally good (they have the same most suitable network function value). The property of logarithm $\log a - \log b = \log \frac{a}{b}$ can reflect this logic, and can serve as normalization [Wang, et al., 2001].

After normalization of the cost, QoS and power functions ->

$$Function_{Cost_n} = W_{COST} * \ln \frac{1}{C_n}$$

$$Function_{QoS_n} = W_{QoS}(W_B \ln B_n + W_L \ln \frac{1}{L_n} + W_{PL} \ln \frac{1}{PL_n} + W_J \ln \frac{1}{J_n} + W_{BER} \ln \frac{1}{BER_n})$$

$$Function_{Power} = W_P * \ln \frac{1}{P_n}$$

After normalization Eq 4-6 became –

$$Function_{MostSuitableNetwork} = (W_{COST} * \ln \frac{1}{C_n}) +$$

$$(W_{QoS}(W_B \ln B_n + W_L \ln \frac{1}{L_n} + W_{PL} \ln \frac{1}{PL_n} + W_J \ln \frac{1}{J_n} + W_{BER} \ln \frac{1}{BER_n})) +$$

$$(W_P * \ln \frac{1}{P_n})$$

Equation 4-7 - Normalized function for most suitable network

The above equation will be used to compute the most suitable network.

4.2.2.3 Cost function comparisons

The table below represents dummy records from the test database. Refer to Appendix B for sample test data.

Table 4-3 Dummy Records

Cost (Pence/ Minute)	Bandwidth (Mbps)	Latency (Seconds)	Packet Loss (Lost Packets/Million)	Jitter (UI)	BER (Erroneous bits/Million)	Power (W)
0.168475	73.97375	0.928437	0.38759366	0.769	0.967541	0.173289
0.163031	29.99431	0.868825	0.94098217	0.597	0.023186	0.952843

The cost function developed by Zhu and McNair requires active negotiation with the network, [Zhu, McNair, 2004]. The function calculates an elimination factor based on what the network can offer and the user's requirements. Due to the fact that the function takes the current network conditions, the algorithm would behave differently under different scenarios, which makes it difficult to compare.

The cost function developed would be compared with the Wang function [Wang, et al., 2001].

$$Wangf_n = W_B \cdot \ln \frac{1}{B_n} + W_P . \ln P_n + W_C \ln C_n (\sum W_i = 1)$$

Equation 4-8 - Wang most suitable network function

$$Function_{MostSuitableNetwork} = (W_{COST} * \ln \frac{1}{C_n}) +$$

$$(W_{QoS} (W_B \ln B_n + W_L \ln \frac{1}{L_n} + W_{PL} \ln \frac{1}{PL_n} + W_J \ln \frac{1}{J_n} + W_{BER} \ln \frac{1}{BER_n})) +$$

$$(W_P * \ln \frac{1}{P_n})$$

Equation 4-9 - MOSSA most suitable network function

Voice applications have a very low tolerance for packet loss, BER, latency, jitter and a medium to high tolerance to bandwidth; refer to Table 3.2 for more information. For real-time voice applications the typical QoS criteria weighting would be:

Bandwidth weight = $W_B = 0.1$

Latency Weight = $W_L = 0.225$

Packet Loss Weight = $W_{PL} = 0.225$

Jitter Weight = $W_J = 0.225$

Bit Error Rate Weight = $W_{BER} = 0.225$

Equal weights have been assigned to the attributes voice applications are most sensitive to and slightly less weighting to the other factors.

Assuming the user has ample battery power and is mainly concerned with cost and a reliable service, then:

Cost Weight = W_{Cost} = 0.4

QoS Weight = W_{QoS} = 0.5

Battery Power Weight = $W_{BatteryPower}$ = 0.1

The Wang [Wang, et al., 2001] and proposed functions for the dummy records would equate to:

Table 4-4 Cost Function comparison

Network	Wang (Cost Function)	Proposed (Cost Function)	Most Optimum Network
1	-3.039522045	1.251148615	According to Wang
2	-2.430860327	1.404544535	According to the proposed solution

The difference in result is because Wang's function takes bandwidth as the only QoS factor into account whereas the proposed function takes bandwidth, packet loss, latency, jitter and bit error rate into account. In the above example, network 2 is the better network because it is more reliable for voice service even though it is slightly higher in cost. The cost elimination algorithm would have already removed the network from consideration if the user was not prepared to pay 17 pence for using the service. The proposed algorithm reflects the user's requirements better than the Wang algorithm.

4.2.2.4 Case for eliminations

Handover latency is the most critical factor when evaluating the efficiency of a Mobile IP. In heterogeneous multiple operator scenarios, the selection algorithm adds to the handover latency. Refer to Appendix C for supporting test programs. The delay introduced due to the selection algorithm has to be kept low; otherwise the handover latency will increase significantly. As can be seen from the graph below, there is a linear increment in time as the number of service providers increases.

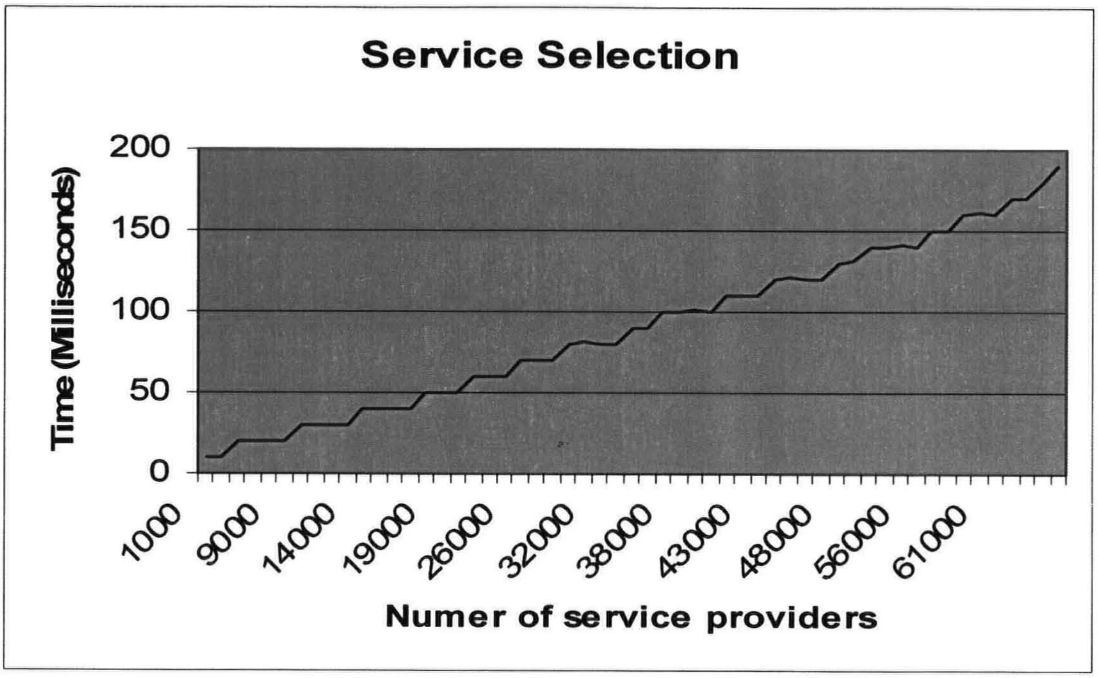


Figure 4-6 Service selection latency

The elimination algorithm reduces the workload of the network selection algorithm. The proposed elimination algorithm, shown in the Figure below, reduces the number of networks eligible for selection.

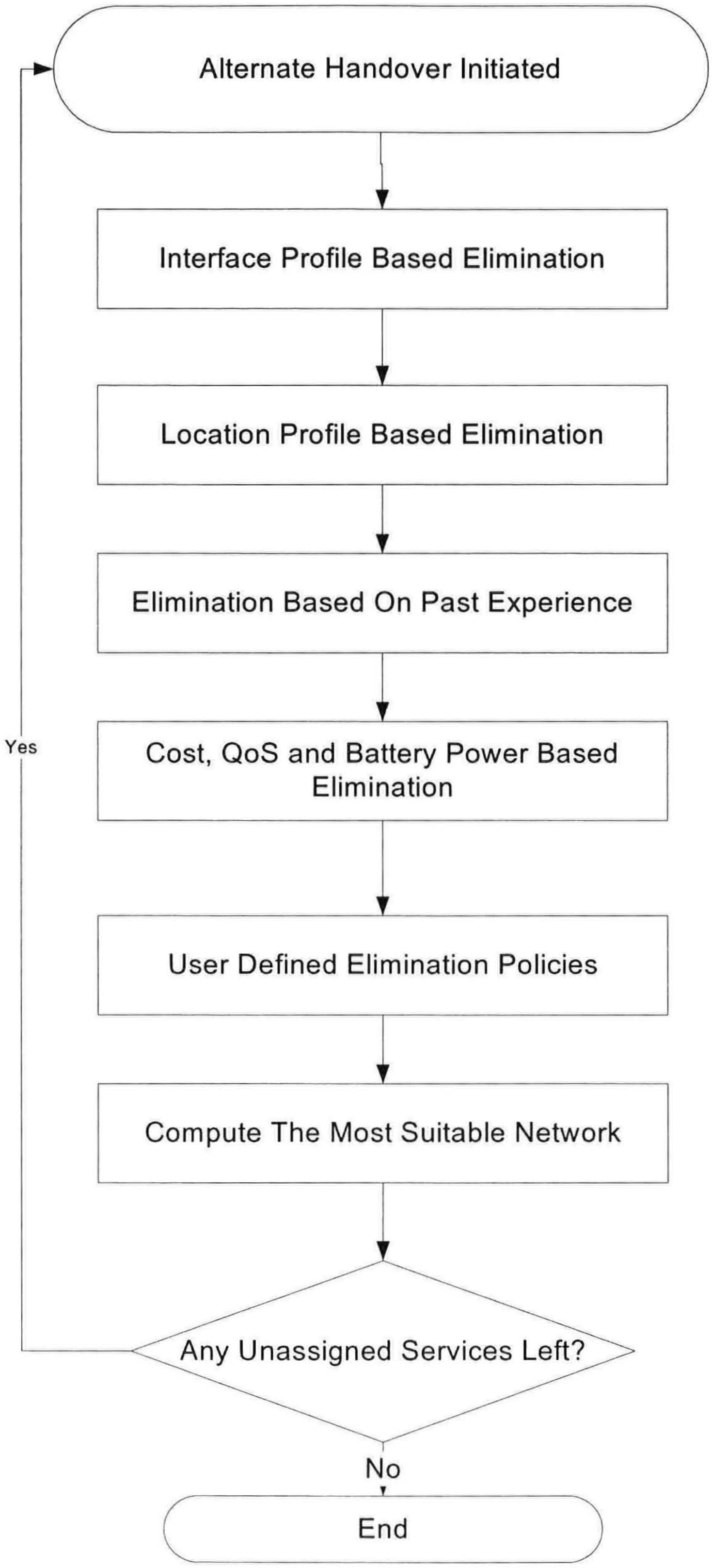


Figure 4-7 Algorithm for service selection

Figure 4.8 demonstrates the sharp reduction in delays when elimination policies are employed. When all the networks were found to be suitable, the selection algorithm took longer to compute the best network. The increased latency when all the networks were eligible for selection was due to the increased workload of the network selection algorithm.

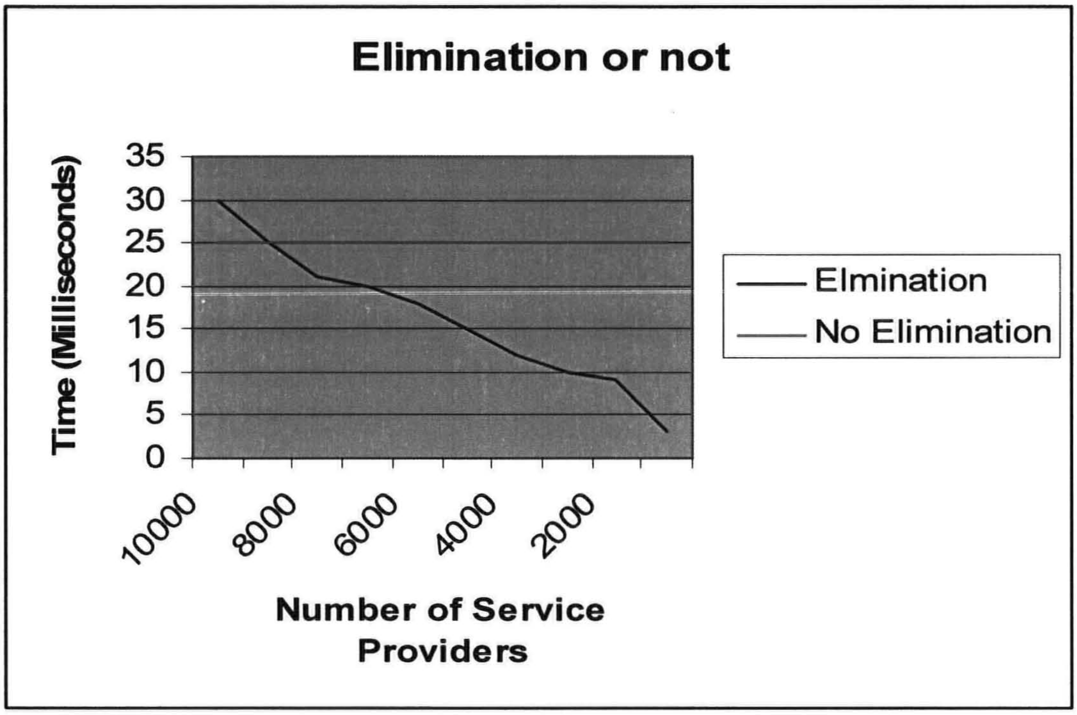


Figure 4-8 Effect of elimination on selection latency

The elimination reduces the workload of the algorithm and improves the handover latency. The improvements in efficiency along with elimination policies make the algorithm faster and more reliable.

4.3 Algorithm Overview

A single user may have multiple active sessions with various operators. When the user initiates new services the algorithm will evaluate the best network and hand over to the most suitable network for that service. If the user has been using more than one service from one operator, then all the services would be treated as a collection. The cost function would be computed for the collection and all the services would be handed off to the best network based on their collective cost function.

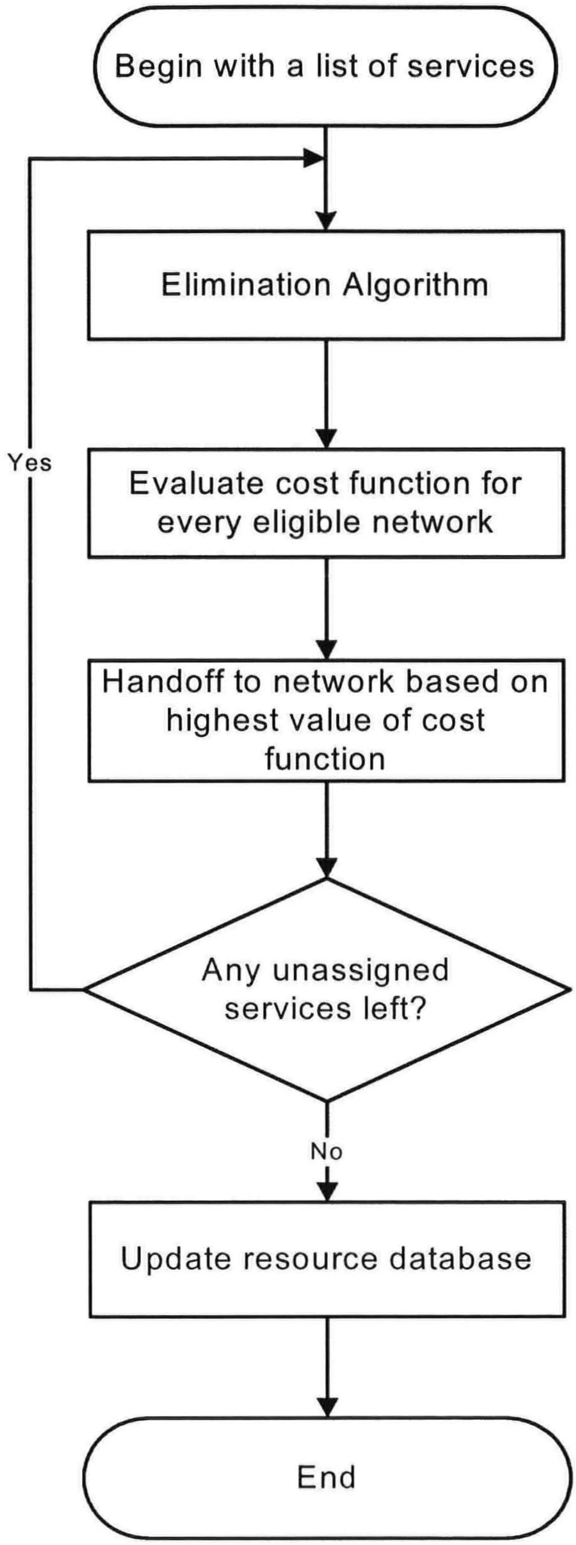


Figure 4-9 Selection procedure

The above flowchart demonstrates the outline process of the algorithm.

4.4 Summary

The service selection algorithm computes the best network based on user preferences without significantly affecting the handover latency. Service advertisement and selection mechanisms provide a perfect platform for multi-operator scenarios in the next generation networks. The algorithm computes the most suitable network based on user preferences and with the introduction of the elimination policies, it preformed faster than the existing methods.

4.5 Contributions

1) Service selection algorithm capable of handling both imperative and alternative handovers.
2) Service selection algorithm which reduces network selection latency by eliminating networks which are unusable or unsuitable based on user preferences.
3) Service selection algorithm which takes into account latency, jitter, BER and packet loss along with bandwidth as QoS selection criterion.
4) Service selection algorithm capable of working with a service advertisement mechanism.

Chapter 5 Mobility Management in Future Mobility Networks

5.1 Mobile IP simulation analysis

As discussed in the previous chapters, handover latency is one of the main drawbacks of Mobile IP. Mobile IP variants discussed above try to reduce handoff latency in their own unique ways. A simulation analysis was conducted to compare the effect on handoff latency of the different Mobile IP variants.

The Network Simulator version 2, NS-2, was used for evaluating Mobile IP variants. The standard NS-2 distribution version ns-allinone2.1b7a was patched with the NS wireless extension module [Widmer, 2001]. This was further extended by the Hierarchical and Fast Mobile IP extensions developed by Robert Hsieh [Hsieh, et al., 2003]. The extensions include a MAP entity, AR entity and Mobile IPv6 binding mechanism along with the HMIP and FMIP protocols.

5.1.1 Simulation Scenario

The goal of the simulation was to study the effects on an end-to-end TCP communication channel when hierarchical and fast Mobile IP are employed. In particular, the intention was to examine the effect on throughput under micro and macro mobility scenarios. The Figure below shows the network topology used for the simulations. The network topology consists of –

1) Correspondent node (CN) – Correspondent node is a wired node sending and receiving packets to/from the Mobile Node.
2) Node 1 (N1) – Node 1 is a wired node linking up correspondent node, home agent and mobility anchor point.
3) Node 2 (N2) – Node 2 is used as a link up node for mobility anchor point and access router 1.
4) Node 3 (N3) – Node 3 is used as a link up node for mobility anchor point and access router 2.

5) Access Router 1 (AR1) – Access router 1 is a wireless access point offering wireless connectivity to wireless devices.
6) Access Router 2 (AR2) – Access router 2 is a wireless access point offering wireless connectivity to wireless devices.
7) Mobile Node (MN) – Mobile node is a wireless device with no wired links and connecting to wireless access routers, when available, for connectivity.
8) Mobility Anchor Point (MAP) – Mobility anchor point serves as the MAP node of the HMIP, providing MN's a local home agent in the network.
9) Home Agent (HA) – Home agent serves as the home agent as described in Mobile IP.

The link characteristics namely bandwidth (megabits/sec) and delay (milliseconds), are shown beside the link. The access routers are set to be 70 meters with free space between them. The wireless coverage area of the access router is 40 meters approximately. The Lucent WaveLan card was simulated under NS. The Mobile Node's movement is constant at 1m/s.

The Correspondent Node creates a TCP communication channel with the Mobile Node. The TCP congestion window is set to 32 KB and maximum packet size is limited to 512 KB. The TCP application type simulated is FTP and is started at 5 simulator seconds and terminated at 80 seconds. The whole simulation run is for 75 simulator seconds.

The Mobile Node is originally positioned near HA and begins its journey towards access router 1 at 6 seconds. Then at 45 seconds makes its journey towards access router 2.

Figure 5.1 below gives a graphical representation of the simulation scenario. Along with each component, the X and Y co-ordinates have been listed in the parenthesis. Mobile Node's movements have been represented with the dotted lines and the type of handover is also mentioned.

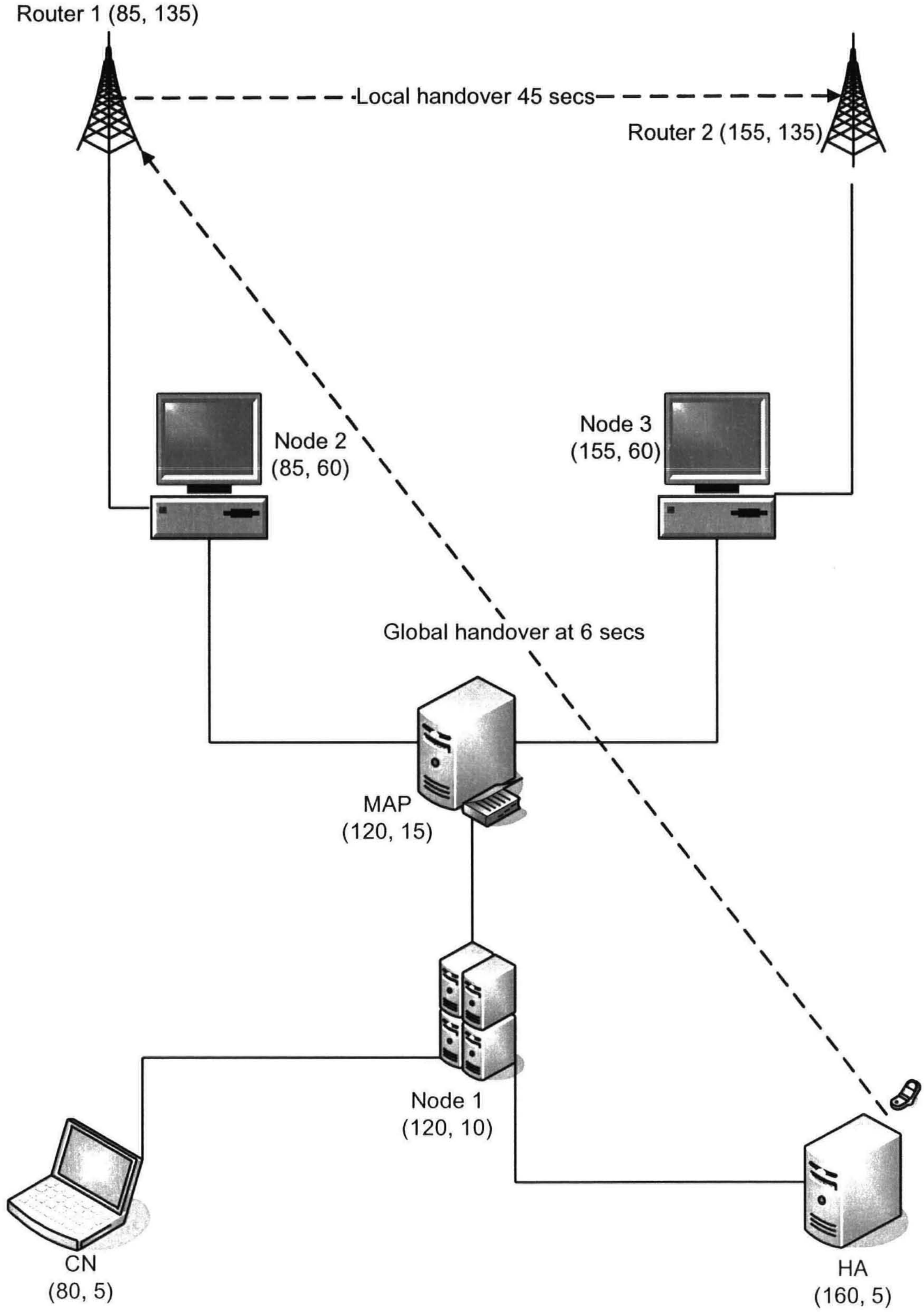


Figure 5-1 Simulation scenario

5.1.2 Simulation results

The overall throughput was observed to be best in the combination of fast and hierarchical followed by hierarchical, fast and Mobile IP respectively, as can be observed in Table 5.1 below.

Table 5-1Throughput comparison Mobile IP

	Fast Mobile IP	Hierarchical Mobile IP	Mobile IP	(Fast + Hierarchical) Mobile IP
Average Throughput in (Mbps)	37.42	38.43	36.90	38.64

The hierarchical and fast variants improve on the throughput of Mobile IP which can still further be improved by using the combination of the two. The results reaffirm what has been discussed earlier.

The plot below of throughput against time, provides a more detailed picture of when the node experiences a lower than average throughput. The node experiences a sharp drop in performance when it moves from the home agent coverage area to router 1's coverage area. The handover requires the node to move across domains and hence we notice a sharp drop in the throughput. However, when mobile node moves between Router 1 and Router 2 the drop in hierarchical and fast hierarchical is hardly noticeable whereas there is a sharp drop in fast and Mobile IP. Since the second handover is within the same domain, hierarchical Mobile IP deals with it locally.

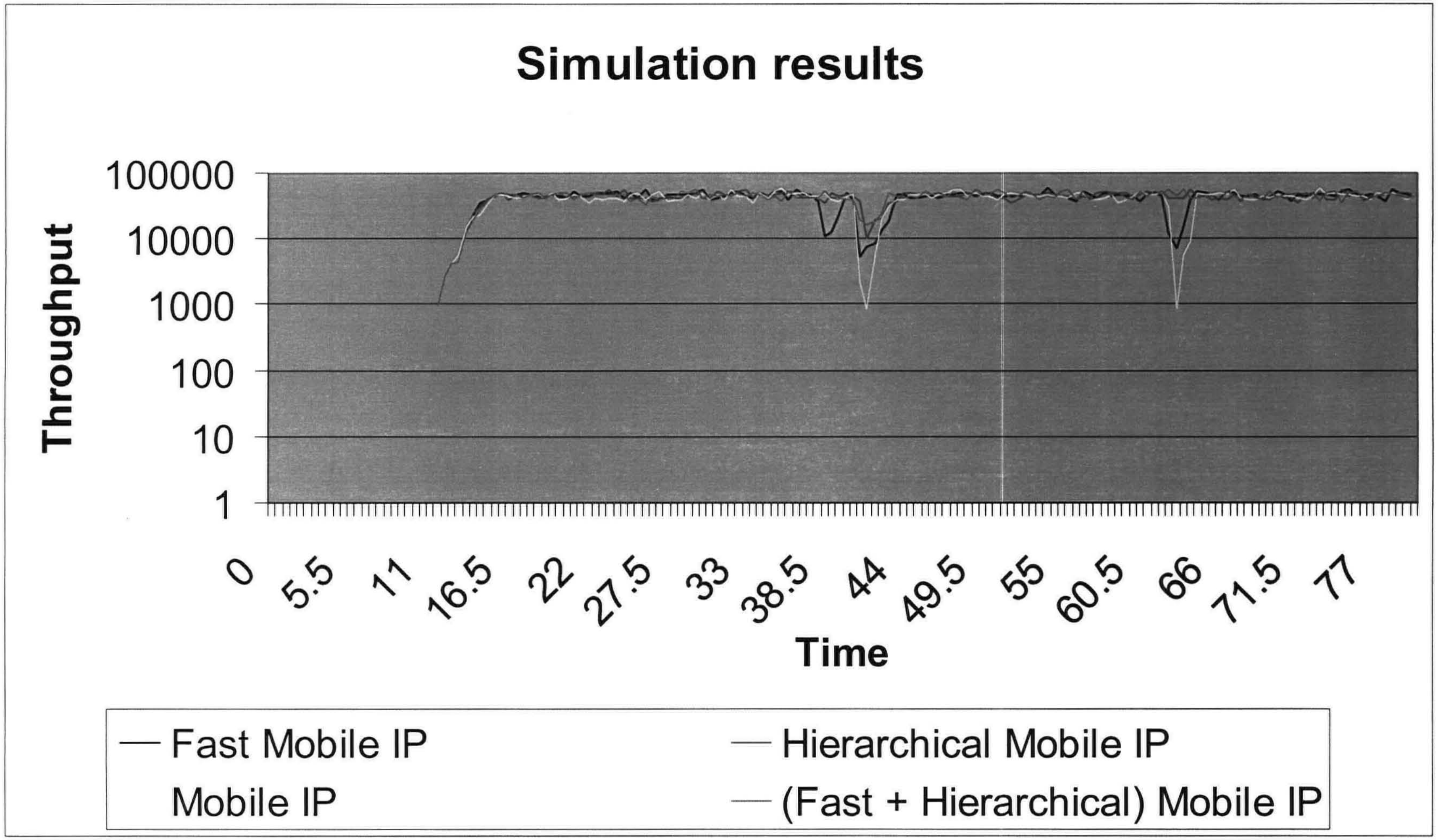


Figure 5-2 Simulation results Mobile IP

5.2 Mobility management

Hierarchical Mobile IP addresses the issue of registration latency but does not tackle the move detection latency. Fast Mobile IP on the other hand addresses move detection latency but ignores the registration latency. A combination of FMIP and HMIP reduces both move and registration latency as has been observed in simulations. Seamless Mobile IP builds upon a combination of fast and hierarchical but comprises the user's freedom of choice by handing over the control of move detection to the network. Future IP based networks promise users the flexibility of cross network mobility. Service advertisements and service selection algorithms discussed enable users to exercise their freedom of choice. Since, SMIP compromises with the user's freedom of choice, it will not be considered further.

SMIP does compromise handover transparency but highlights that a combination of HMIP and FMIP does reduce handoff latency, as has been observed in the above simulations. A combination of HMIP and FMIP will be used as the mobility management model for the remainder of this study.

5.2.1 Cross-network handovers

Simulation results indicate that handovers across domains lead to a sharp decrease in performance of all the mobility management architectures. Movement across domains is termed as cross-network (or inter-network) handovers. Handovers across domains are subject to accounting and security, not considered in the above simulations. Handovers between two networks could be made seamless by introducing coverage area overlaps [Bansal, et al., 2003].

5.2.1.1 Model Network

5.2.1.1.1 Coverage area overlaps and border base stations (BBS)

Coverage area overlaps, as shown in Figure 5-3, enable a smoother transition across domains. Border Base Stations (BBS) would serve as a bridge between two MAP domains. BBS adverts would hold information of both the domains and would enable MN to register with the adjacent network before even entering it, thus allowing a cushion by enabling pre-emptive handovers.

Large networks should be divided into smaller sub-networks to allow better load balancing. Its own MAP would serve each sub-network. Reducing the load on MAPs would enable them to assume more responsibilities like Session Initiation Protocol (SIP) proxy and redirect server [Handley, et al., 1999].

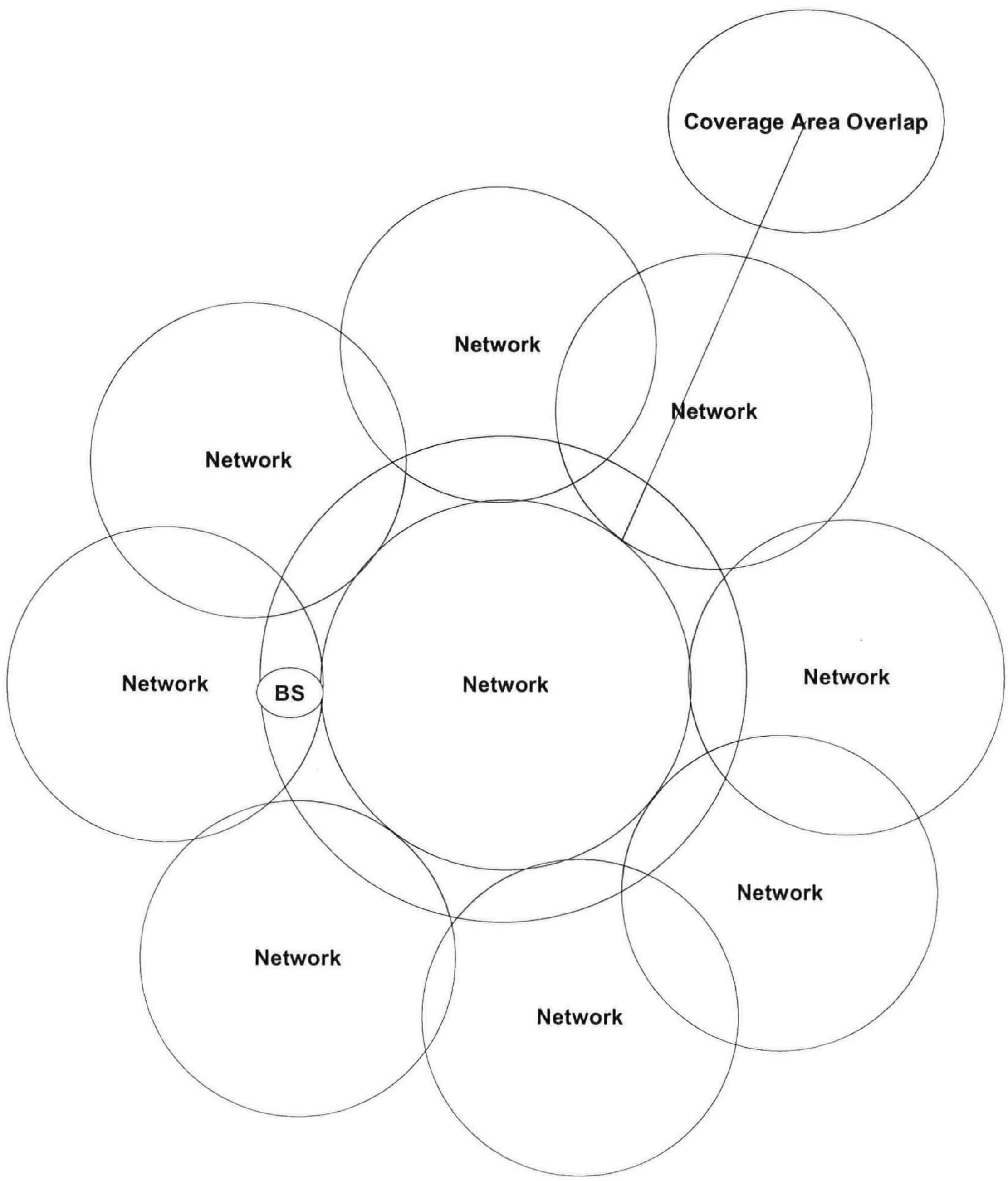


Figure 5-3 - Network divided in sub-networks with overlap area marked in grey

However, coverage area overlaps might not be very well appreciated between competitor networks but they would enable smooth cross-network handovers.

5.2.1.1.2 Registration process

HMIP allows MN's to hold multiple registrations but different billing policies might limit the registration capability of the MN. MN might not be able to register itself in a network because of a credit issue. To aid pre-handover selection and lower handoff latency it would help if the MN is able to register with all the possible options. Until the MN initiates a service, its registration would be in a soft state. Holding soft registration would enable mobiles to acquire candidate network capability information. Registration could be renewed or dropped after the validity period expires.

5.2.1.1.3 Sample network

A sample network, as shown in Figure 5-4, would be used to demonstrate the proposed coverage area overlaps.

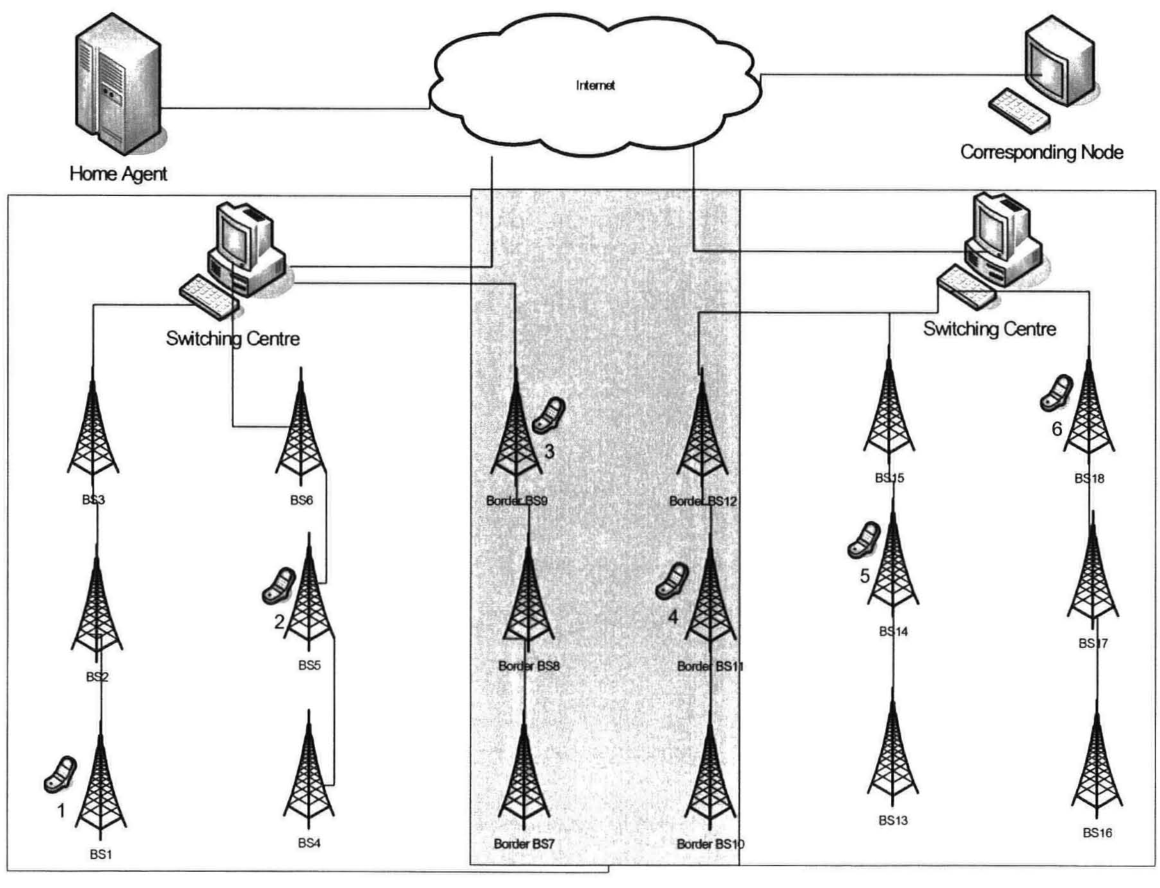


Figure 5-4 - HMIP model network

The network (Figure 5.4 above) comprises –

- Switching Centres (SC) – Switching centres would be responsible for providing regional registrations (RCoA) to the mobile nodes and maintain bindings for mobiles currently visiting the network. The switching centre also behaves as the border router for the site. Separate machines could perform all three tasks but because of the reduced load as compared with HMIPv6 all of them could be performed by one machine. A switching centre could also perform as a gateway to the Internet or could be linked up to the network's gateway. SC could also act as a Home Agent (HA) thus enabling distribution throughout the network. The area served by the SC as a MAP will be termed as a Switching Center Location Area (SCLA) in the rest of the thesis.
- Base Stations (BS) – BS would be a router capable of handling MNs in its Virtual Location Area (VLA). A VLA is the area range within which the MB is capable of handling an MN. BS should send advertisements from time to time and register new clients when requested. BS should be able to route MN's packets correctly and process handovers when requested by a node.
- Border Base Stations (BBS) – BBS would perform all the tasks of a BS but their advertisements would hold details of MAPs in adjacent networks. BBS would be wired to two networks forming a bridge. BBS serve in the border regions of two adjoining networks shown as the shaded region in Figure 5.10.
- Mobile Nodes (MN) - Since MN's are termed as devices with low power, every effort has been made to keep the processing by an MN down to a minimum. The MN would be effectively doing the same tasks as they have been doing. The common tasks include listening to router advertisements and registering with the foreign agents, to be able to de-capsulate packets sent, etc. No new tasks would be performed by the MN.
- Home Agent (HA) – Performs all duties as described in MIPv6 [Davidson, Arkko, 2003].
- Correspondent Node (CN) – CN could be a wired or wireless terminal communicating with MN in the same or different networks.

5.2.1.1.4 Protocol details

To describe the behaviour of operation performed by the network as MN travels through the network from stages 1 to 6 in the model network (Figure 5-4).

1) MN enters the VLA of BS1 and starts receiving router advertisements. MN forms a stateless link address (LCoA) and sends a BU to BS1. BS1 on receipt of BU updates its binding cache and sends a BAck to MN. MN forms a stateless site address (RCoA) and sends a BU to SC1. SC1 on receipt of BU updates its binding cache and sends a BAck to MN. MN sends a BU (RCoA) to its HA and CNs. HA (and CNs) on receipt of the BU updates its binding cache and sends a BAck to MN.
2) MN moves in the VLA of BS5 and starts receiving router advertisements. MN forms a LCoA and sends a BU to BS5. BS5 on receipt of BU updates its binding cache and sends a BAck to MN. MN sends a BU to SC1. SC1 updates its binding cache and sends a BAck to MN.
3) MN moves in the VLA of BS9 and starts receiving router advertisements. MN forms a stateless LCoA and sends a BU to BS9. BS9 on receipt of BU updates its binding cache and sends a BAck to MN. Router advertisements from BS9 contain details of MAPs in regions 1 and 2. Since, MN is already registered in region 1 it has a choice of also registering in region 2. MN forms a stateless secondary RCoA (SRCoA) for region 2 and sends a BU to SC2. SC2 on receipt of BU updates its binding cache and sends a BAck to MN. MN sends a BU (RCoA, SRCoA, LCoA) to SC1.
4) MN moves in the VLA of BS11 and starts receiving router advertisements. MN forms a stateless LCoA and sends a BU to BS11. BS11 on receipt of BU updates its binding cache and sends a BAck to MN. Since MN has moved in region 2 its SRCoA now becomes its RCoA and RCoA now becomes SRCoA. So MNs new binding is $RCoA_{NEW} \rightarrow SRCoA_{OLD}$, $SRCoA_{NEW} \rightarrow RCoA_{OLD}$, $LCoA_{NEW}$. MN sends a binding update to its home agent (RCoA). MN sends a binding update (RCoA, SRCoA, and LCoA) to SC1 and SC2.
5) MN moves in the VLA of BS14 and starts receiving router advertisements. MN forms a stateless LCoA and sends a BU to BS14. BS14 on receipt of BU updates its binding cache and sends a BAck to MN. MN drops its SRCoA and sends a binding update (RCoA, LCoA) to SC2.
6) MN moves in the VLA of BS16 and starts receiving router advertisements. MN forms an LCoA and sends a BU to BS16. BS16 on receipt of BU updates

its binding cache and sends a BAck to MN. MN sends a BU to SC1. SC2 updates its binding cache and sends a BAck to MN.

5.2.1.2 Comparison with HMIPv6

HMIP enables MNs to hold multiple site registrations at a given time. The framework discussed further enables smooth handovers across domains. MN would register itself in more than one site when in border areas and based on its future movements, handovers occur as discussed in the previous section. Whilst in border areas, MN will be served by SCs of both regions hence minimizing packet loss probability. The previous SC would hold the route to MNs new SC until MN moves out of the border areas or if it receives a tear down message from MN.

Table 5-2 HMIPv6 comparison with the proposed protocol

	HMIPv6	Proposed Protocol
Inter-site Mobility	MN gets a new LCoA on every link within the network and it sends a binding update to the site mobility server	Same as HMIPv6 but MN would need to register its secondary RCoA when it enters border areas.
Intra-site Mobility	MN gets a new RCoA and LCoA. MN sends binding updates to its home agent and the site mobility server.	MN just gets an LCoA and secondary RCoA becomes its RCoA. MN sends binding updates to its home agent and site mobility server.
Packet Delivery	Correspondent Node sends a packet to the MN using its RCoA. Site mobility, a server intercepts and tunnels the packet to the current LCoA of the mobile node and sends it.	Same as HMIPv6.

5.2.1.3 Performance evaluation of HMIP with Coverage Area Overlaps

5.2.1.3.1 Global handover latency

We assume that the handover is initiated by the Mobile Node. For the wireless delay we assume the same value for the uplink and downlink delay. We assume that stateless address formation processing time is negligible and there are no duplicate addresses formed.

5.2.1.3.1.1 HMIPv6

Consider the scenario of a handover between BS9 – BS11 (refer to Figure 5-4).

Figure 5-10, illustrates the interactions taking place in the event of a global handover. The global handover requires MN to send BUs to BS, MAP and HA.

Hence,

$$\text{Global Handover Latency} = BU_{BS} + BU_{MAP} + BU_{HA}$$

Equation 5-1 - HMIPv6 Global handover latency

Where,

BU_{BS} is the time taken by MN to send one or more BU to BS and get an acknowledgement

BU_{MAP} is the time taken by MN to send one or more BU to BS and get an acknowledgement

BU_{HA} is the time taken by MN to send one or more BU to BS and get an acknowledgement

A BU is confirmed when MN receives a BAck. Hence, the BU latencies of BS, MAP and HA will be

$$BU_{BS} = (\#BU_{BS} + 1)Wd$$

Equation 5-2 - Binding Update Latency

Where,

$(\#BU_{BS} + 1)$ is the number of binding updates sent to BS and we add 1 for the binding acknowledgement

Wd is the delay introduced by the wireless part of the network

The communication between BS and MN will be wireless so the round trip times will only be affected by wireless transmission delays. MN will send 1 or more BUs and wait for a BAck to confirm BU.

$$BU_{MAP} = ((\#BU_{MAP} + 1)(Wd + Ld_{BS-MAP}))$$

Equation 5-3 - Binding Update Latency Mobility Anchor Point

Where,

$(\#BU_{MAP} + 1)$ is the number of binding updates sent to MAP and we add 1 for binding acknowledgement

LD_{BS-MAP} latency introduced by the wired part between BS and MAP

The communication between BS and MAP will require packets to travel via air (wireless) to the BS and then travel the wired part between BS and MAP. MN will send 1 or more BU and wait for a BAck to confirm BU.

$$BU_{HA} = ((\#BU_{HA} + 1)(Wd + Ld_{BS-HA}))$$

Equation 5-4 - Binding Update Latency Home Agent

Where,

$(\#BU_{HA} + 1)$ is the number of binding updates sent to the Home Agent and we add 1 for binding acknowledgement

LD_{BS-HA} latency introduced by the wired part between Bs and HA

The communication between BS and HA will require packets to travel via air (wireless) to the BS and then travel the wired part between BS and HA. MN will send 1 or more BU and wait for a BAck to confirm BU.

So,

$$\text{Global Handover Latency} = BU_{BS} + BU_{MAP} + BU_{HA}$$
$$= (\#BU_{BS} + 1)Wd +$$
$$((\#BU_{MAP} + 1)(Wd + Ld_{BS-MAP})) +$$
$$((\#BU_{HA} + 1)Wd + Ld_{BS-HA})$$

Equation 5-5 - Global Handober Latency

MN will also need to notify correspondent nodes (CN) of its site movement but that can be done simultaneously with Home Agent BU. We assume BU_{HA} takes less time than BU_{CN}

5.2.1.3.1.2 Proposed protocol

In this proposal, MN will initiate the registration process once it enters the BBS VLA. So by the time MN reaches BS11 we assume the BU_{MAP} phase would have been completed. Hence, we equate BU_{MAP} to '0'.

$$\text{Global Handover Latency} = BU_{BS} + BU_{MAP} + BU_{HA}$$

Equation 5-6 - Global Handover Latency from Eq. 5-1

Where $BU_{MAP} = 0$ since it has already been processed.

Therefore,

$$\text{Global Handover Latency} = BU_{BS} + BU_{HA}$$

Equation 5-7 - Global Handover Latency proposed protocol

BU_{HA} will be processed whilst MN is in border areas thus effectively enabling a seamless global handover.

The MNs primary site, RCoA, will be registered with the HA based on MNs movements within the border areas. Border areas enable a smooth transition from one network to another and in most cases should provide a lossless handover.

MN will get enough time to process BU_{HA} before leaving the border areas.

5.2.1.3.2 Local Handover latency

The local handover latency will not be affected by this proposal. The local signalling load will be increased when the MN is in the border areas because MN will have to notify MAPs of both the regions about its movements.

5.2.2 WLAN Handovers

Providing users with freedom to choose has to be backed with the freedom of movement. If the users have to restrict their movements or reinitiate their channels upon reattachment, then it would be pointless to provide the freedom of selection. However, in some scenarios freedom of choice may have to be sacrificed for the sake maintaining a connection. For example, a user is downloading the latest HP drivers for his printers sitting in a Starbucks coffee shop. After finishing his coffee he leaves the shop without realising the download still has not completed. Due to the short range of the WLAN hotspots the user is quickly out of reach with the download still half complete. Handover now has to be performed very quickly in order to maintain connection with the web server. Since, web servers do not support route optimisation the task becomes even harder.

5.2.2.1 WLAN Binding information

Hotspot binding tables would have to be changed to provide seamless handovers between WLAN and HMIP networks. Upon registration in a hotspot the MN would be provided with virtual and private bindings. If the hotspot supports the extended service set. To provide seamless handovers, hotspots will have to store a mobile node's home address, refer to table 5-3. Since hotspots operate in very small areas, much smaller than a standard mobile telephone cell, the MN could use its HMIP VCoA as the HA for the hotspot.

Table 5-3 WLAN binding table

MAC	PCoA	VCoA	Credit	Seamless	HA
00-13-20-8B-9F-A0	192.168.0.3	66.94.234.13	$5	Y	64.4.33.7

5.2.2.2 WLAN to HMIP handover

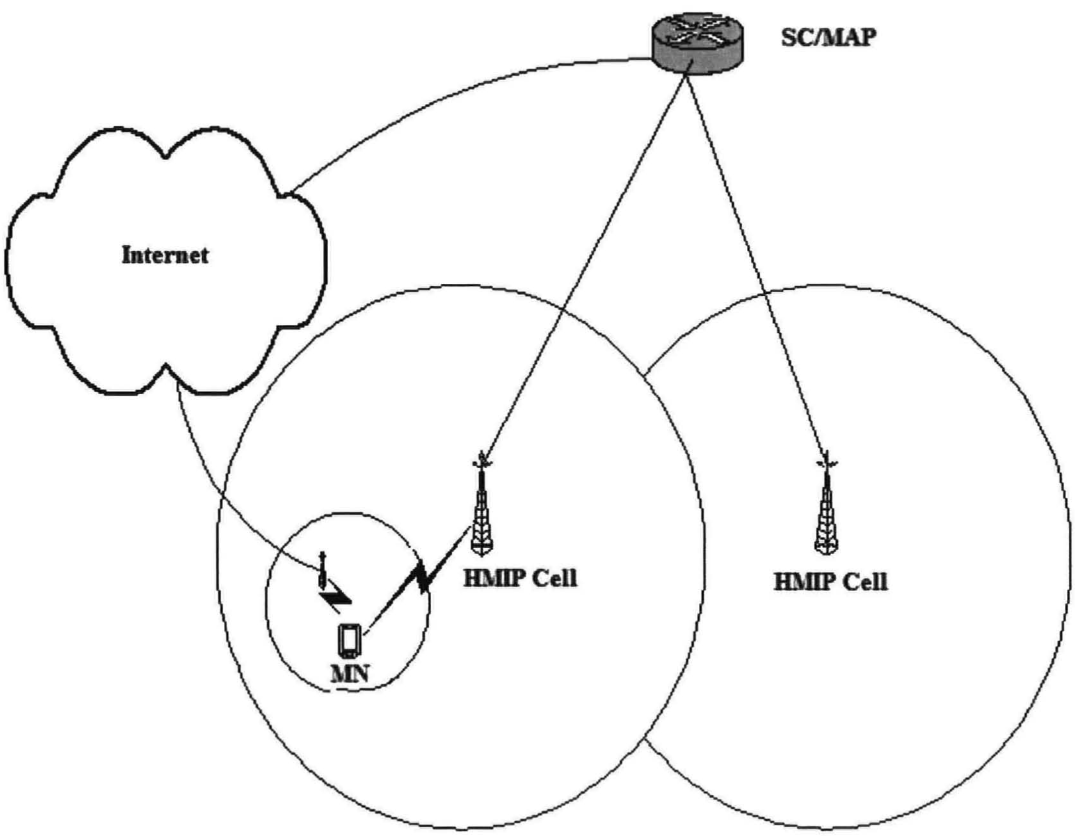


Figure 5-5 - WLAN-HMIP handover

Consider the following scenario (refer to figure 5-5) – MN moves into the hotspot and wants to enjoy high speed wireless access provided by the hotspot whilst still being connected to the HMIP network for voice and other services. MN initiated a data channel but moves out of the hotspot coverage. If the hotspot does not support cross-network seamless handover MNs call will be dropped or the MN will have to move back within the hotspot coverage are, hence restricting the movements of the MN.

Seamless handover could happen in the following ways –

- CN supports route optimization – If the CN (Correspondent Node) supports route optimization, MN could send it a binding update so that the CN could start transmitting the packets to its new link attachment. Packets already in transit will be rerouted by the hotspot access router to the MNs home address (which could be its VCoA).
- CN does not support route optimization – If the CN does not support route optimization then all the future packets will be rerouted to the HA of the MN where they will be tunnelled to MNs current link attachment.

5.2.2.3 HMIP to WLAN handover

These handovers will only be useful if the CN supports route optimisation. Since Web services do not support route optimisation, packets belonging to the MN will have to be re-routed by the MAP to the hotspot whereby the packets will be travelling through the global Internet domain at the best effort service.

Packet latency HMIP to WLAN $= HMIP_{(Wired+wireless)delay} + BE + Hotspot_{(wired+wireless)delay}$

Equation 5-8 - Packet Latency HMIP to WLAN

Packet Latency HMIP = $= HMIP_{(Wired+wireless)delay}$

Equation 5-9 - Packet Latency HMIP

In absence of route optimisation, packets cannot be routed to the MN directly. Hence, this scenario is only useful if the CN supports route optimization or both the cells are a part of the same network. Route optimisation would enable updating MNs current location, so that the CN can send the packets to MNs current location rather that the packets having to be re-routed to MNs current location from its previous location.

5.3 Call setup and Handover procedure

Service advertisements and selections should be operating as an integral part of the mobility management architecture. The flowchart, figure 5.6, demonstrates the operation of the proposed mechanisms in the greater mobility management domain. The terminal starts by scanning for the main network. The main network of the user is the subscribed to network which will provide Home Agent service to the network. The terminal will register with the network if it can connect to it and then access the service advertisement information.

The terminal will scan for the single frequency band, if the terminal cannot connect to its main network. The terminal will only scan the single frequency band if it does not have any network connections available to it. Once a connection with an operator is established the terminal can collect service advertisements and the user can start using the services of the connected network. The services initiated whilst the service advertisement information is being accumulated will not benefit from the service selection algorithm, as the algorithm will be working with limited information.

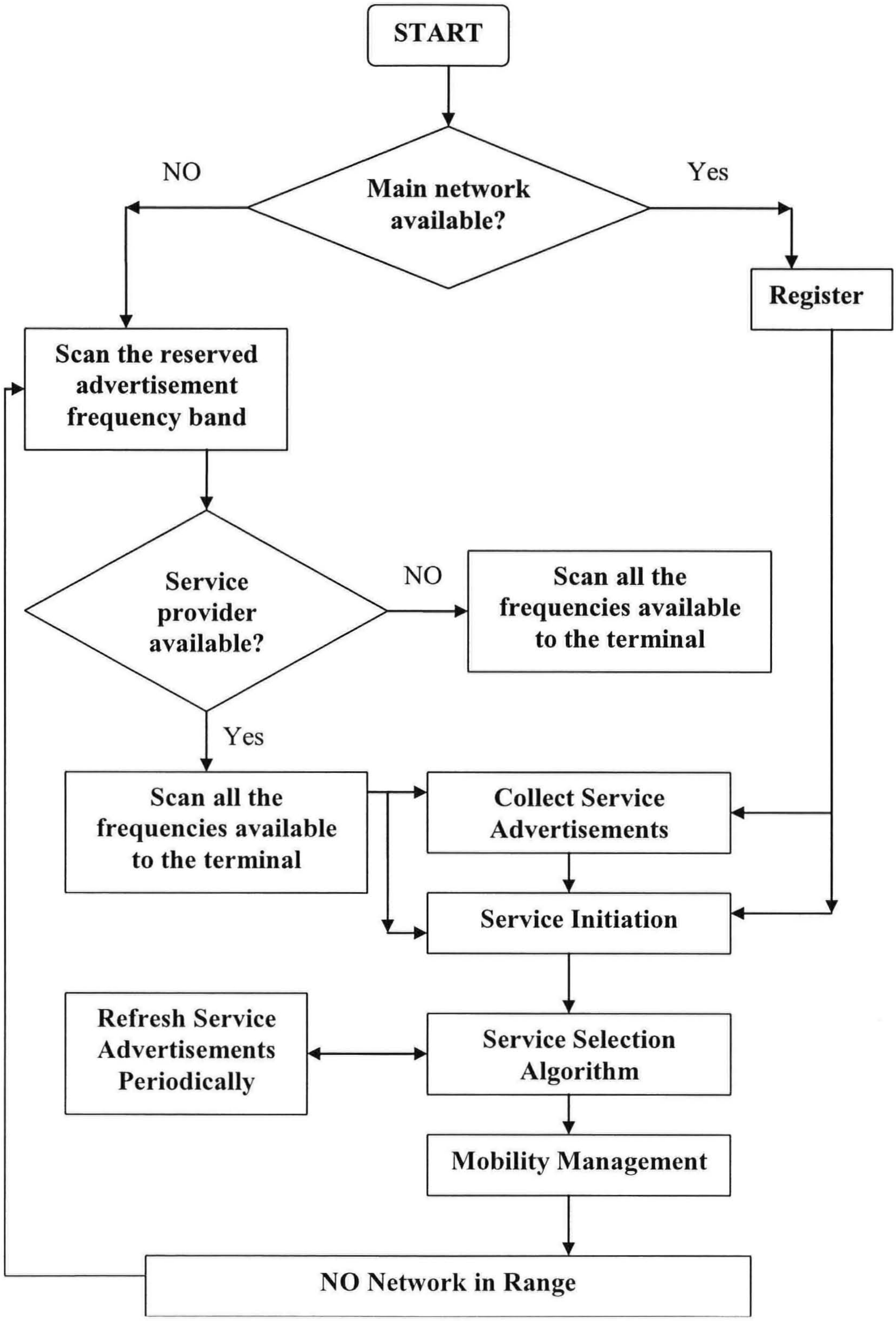


Figure 5-6 Service Selection Algorithm

Once the service advertisements have been gathered, the service selection algorithm selects the most suitable network for the user's needs. Since the users enjoy freedom of mobility along with the freedom of choice, the service advertisements have to be refreshed periodically. Until the mobile unit maintains an active network connection, the mobility management would be oscillating within advertisements, selection and handover management as shown in the Figure 5-7.

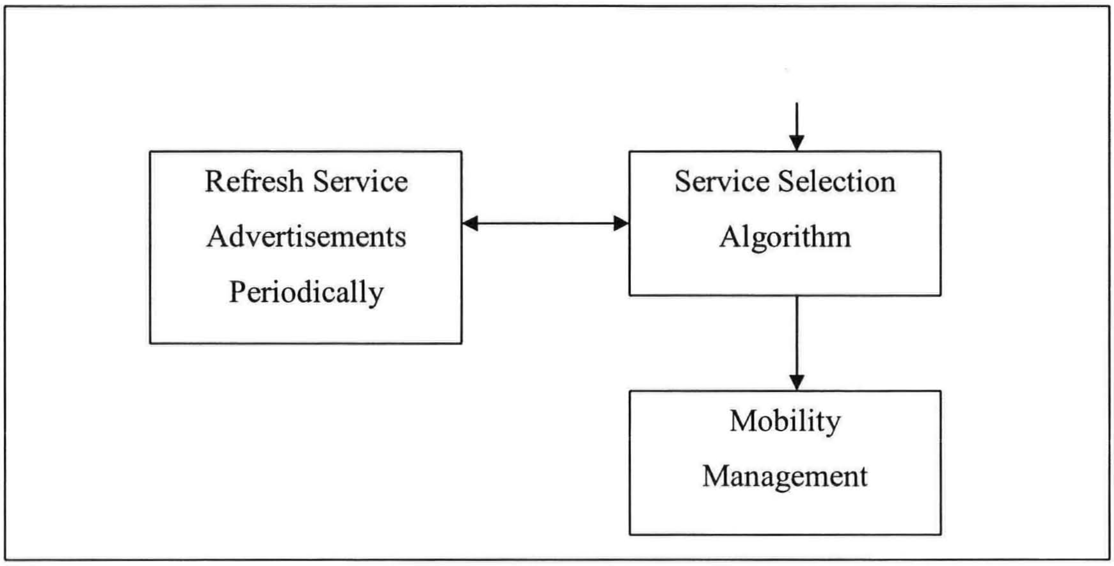


Figure 5-7 Maintaining active network extract of figure 5-6

In the event of the terminal losing the active network connection and not being able to establish a connection with any other network, the terminal will scan the single frequency band. If the single frequency band is unable to locate a service provider then the terminal will scan all the frequencies available to itself to try and establish a connection. Once a connection has been re-established then the terminal can acquire service advertisement information to discover other service providers in the area.

5.4 Analysis of the architecture

The proposed mobility management architecture reduces the handover latency by using a mobility management architecture based on HMIP and FMIP. The simulations highlighted the weakness in all the mobility management architectures when undergoing global handovers. The proposed inter-domain handovers reduce the global handover latency and enable seamless handovers across domains. The modifications to WLAN binding tables enable smoother handovers between WLANs and other networks. Service advertisements and service selections provide added value to the mobility management architecture.

5.5 Contributions

⇒ Introduced Coverage Area Overlaps to reduce HMIP global handover latency

⇒ Introduced Home Address in WLAN binding table, which enabled seamless handoffs between WLAN and HMIP networks.

Chapter 6 Conclusions and Further Work

6.1 Conclusions

The thesis presented a user-centric mobility management architecture for IP based wireless networks. Future generations of wireless technology will be designed to provide heterogeneous access to users. Heterogeneous access networks provide users with flexibility in selecting services but pose great challenges. This study addressed the challenges caused by service advertisements and service selections whilst trying to minimise the effects on handover latency.

Heterogeneous networks empower users to move across domains rather that just being confined to one region, but the possible lack of knowledge of networks, may lead to the terminals making bad choices which lead to poor performance. The rapid growth in the wireless network technology industry has led to the development of various types of wireless networks, ranging from 802.11b to WiMAX. A multitude of wireless access technologies add weight to the problem since the terminals would had to scan all those frequencies to establish which networks it had access to. The problem is intensified with the terminal's movement, as the list of accessible networks will have to be refreshed periodically. The scanning overhead requirement to exercise the freedom of choice will have negative consequences on the terminal's limited resources. Service advertisements provide an alternative to the overhead of scanning to explore the available networks in the vicinity.

6.1.1 Service Advertisement Data Repository

To establish the requirements of service advertisements, a scenario based study was conducted. Service advertisements were sub-classified into an access mechanism and a data repository. Classification of service advertisements enabled the study to break it up in two separate entities capable of working with each other but also could be employed in other architectures with ease. The data repository outlined the information stored in the database whereas the frequency band access mechanism defined an operator independent access mechanism.

The hierarchical architecture proposed for the data repository enabled segregation and the local information to be more accessible to the user. The proposed hierarchy facilitates repositories catering for local services on a zonal basis. Due to the segregation, the data repository is able to allow location based searches, enabling a terminal to explore local services with a limited coverage. The repository held network capability information enables the user's selection algorithm to identify networks capable of delivering the services required. The QoS is specified in measurable QoS metrics: latency, jitter, bandwidth, packet loss and bit error rate. The measurements are provided by the networks and hence the repository relies on the honesty of the networks to provide reliable information. Cost features are included in the repository such as the charge model, cost and usability period. Usability periods have been introduced to shield the user from sliding scale prices. The coverage area of smaller networks also features in the database to enable the terminals to establish the coverage areas of the service. The coverage area is the radius of the network with the access point as the centre. The modified radius gives the terminal a clearer picture of the extent of the coverage.

It was established that the data repository's response was not much affected as the number of query parameters were increased whereas it showed a linear increment with the location based searches. The data repository raises the issue of management and misuse. The issue of misuse could be handled by charging the user per use. The data repository discussed provides an access independent mechanism to retrieve network availability and capability information.

6.1.2 Service Advertisement Single Frequency Band Access Mechanism

The main challenge facing the access mechanism was availability in the absence of the main service provider. Since the repository did not put a restriction on the type of access mechanism, the study focussed on developing a mechanism accessible when the terminal's main service provider was absent. The proposed mechanism was a single frequency band reserved for service advertisements. All the operators would be able to broadcast service advertisement information piggybacked on a router advertisement packet. The architecture does not define any types of packets but just uses the resources available thus incurring a very small implementation overhead. The service advertisement information piggybacked on a router advertisement includes links to the major service providers in the region along with links to an operators own repository and its access information. Scanning just the single frequency will reduce the scanning overhead.

The performance test indicated a latency increase with every channel to be scanned and thus scanning should be limited to scenarios when the terminal has no network coverage. If network coverage is available the terminal should access the data repository via existing access mechanisms. Such access mechanisms complement the data repository thereby providing a perfect platform to deliver service advertisements.

The study contributed requirements analysis, hierarchical architecture, location based retrieval and single frequency based-access mechanism to service advertisements. The QoS attribute and cost breakdown layout in the repository is also a contribution of this study. The combination of access mechanism and data repository not only provide a transparent service but also reduced the scanning overhead. The service advertisements must be accessible even when the main service provider of the user is absent.

6.1.3 Service Selection Algorithm

The information retrieved from the service advertisements feed into the service selection algorithm, which picks the most suitable network based on the user's requirements. Service advertisements will not be of much use unless they are assisted by the selection algorithms. Service selection algorithms will be used to initialise new services and during handovers. The selection algorithm takes into account imperative and alternative handovers.

Imperative handovers were processed when network properties fall below an acceptable level and do not enjoy the luxury of time when selecting services. The algorithm picks the first network with acceptable network properties and once a connection has been established the algorithm picks a network based on the user's preferences.

Handovers triggered due to user's preferences not being satisfied by the network are alternative handovers. Alternative handovers are characterized by the freedom of choice of the user to look elsewhere if their requirements are not met by the network. The study proposes an elimination oriented policy based handoff strategy. The algorithm lays great importance on the workload of the terminal by reducing the number of networks available for selection by way of elimination.

Eliminations are also performed based on the interfaces available to the terminal and also the interfaces already in use. Compatibility issues between various interface types are looked after by the proposed interface compatibility matrix. The matrix eliminates the interfaces that have compatibility issues with other interfaces already in use and also suggests the interface combinations which could be used. Eliminations are performed based on the interfaces not available to the terminal.

The elimination policy can be constructed by the user based on location, velocity, multiple operator support and feedback. The algorithm does not impose limitations on the type of elimination policies. The user (and/ or terminal) can add to the policies as and when required. One of the drawbacks with service advertisements (discussed above) was that it relies on the operator's honesty when it comes to the network capability information. The user can store feedback on the terminal and eliminate (or give low priority to) networks, all operators with whom it had experienced poor performance. The eliminations pave the way for selection algorithms to compare networks based on their capability.

The network selection parameters employed by the algorithm were: cost, QoS and source power. QoS provides the capability information of the network whereas battery power establishes the strain on a terminal for using the service. Cost will be the most important criteria for most users when selecting a service. The user will associate an importance with all the parameters based on their requirements. For example, if the user requires a cheap service that is not too costly in terms of battery power and the mobile device is running low on power but does not really care about the quality of the service; a higher importance will be assigned to cost and battery power while assigning minimum importance to QoS.

The cost of using a service has been defined as cost, unit and usability period, as discussed above. The user will define the maximum they are prepared to pay for using the service, all the networks maximum cost requirement and above will be eliminated. QoS as discussed above will also be measured in terms of jitter, latency, bandwidth, packet loss and bit error rate. The user will define the maximum tolerable for all the network selection parameters such as cost, which will enable the algorithm to eliminate networks.

The remaining networks after all the eliminations will be available for selection. The selection function is directly proportional to the bandwidth whilst it has an indirect proportional relationship with cost, latency, jitter, packet loss, bit error rate and battery power. The relationship equation is normalised using logarithms and leads to the following network selection function –

$$Function_{MostSuitableNetwork} = (W_{COST} * \ln \frac{1}{C_n}) +$$

$$(W_{QoS}(W_B \ln B_n + W_L \ln \frac{1}{L_n} + W_{PL} \ln \frac{1}{PL_n} + W_J \ln \frac{1}{J_n} + W_{BER} \ln \frac{1}{BER_n})) +$$

$$(W_P * \ln \frac{1}{P_n})$$

Equation 6-1 Most Suitable Network Function

Network selection was the most expansive and closely matched of the requirements of the user, and the success of the proposed algorithm has been demonstrated in the results obtained from various measurements. The elimination based policies reduced the burden on terminals, thereby simplifying the selection of the network that most closely matches the user's requirements.

6.1.4 Mobility Management

The service selection and advertisement mechanisms serve as a perfect accompaniment to any mobility management architecture. Mobile IP is the most commonly available mobility management architecture. When a terminal changes its point of attachment to the network it suffers from handoff latency which can be significant enough to affect communication channels. Many variants of Mobile IP reduce handoff latency. Based on the simulation analysis presented in Section 5.1.1, it was recommended that mobility management architecture be based on a combination of fast and hierarchical Mobile IP. To further reduce the handoff latency, the author introduced coverage area overlaps and border base stations to reduce the global handover latency, as simulations pointed out Section 5.1.5 that global handover latency was still a significant problem. The proposed solution further reduced the handover latency and was backwards compatible with hierarchical Mobile IP and Mobile IP. Analysis proves inter-domain handovers can be performed seamlessly by employing proposed modifications. The mobility management architecture also includes modifications to WLAN binding tables enabling smoother handoffs between WLAN and Mobile IP based networks.

The mobility management architecture along with the service advertisement and service selections provides the basis for future heterogeneous mobile networks. It is therefore concluded that the service advertisements and selections enable users to exercise their freedom of choice as is promised for future mobile networks.

6.2 Further Work

6.1.5 6.2.1Security

The selection algorithm computes the most suitable network based on cost, battery power and QoS but users are getting increasingly concerned about the security of information flowing through the networks. The terminals are getting faster and able to support various application types and soon users will start accessing mission critical information on their wireless devices. The algorithm should be able to identify the most secure network (or the network offering most stringent security considerations) when selecting networks for sensitive applications. The algorithm does not consider security as a concern but to cater for future requirements, security will have to be added as a criterion when selecting networks.

6.1.6 6.2.2 Quality of service (QoS) Attributes

QoS attributes form an integral part of the selection algorithm and are the key differentiators between networks. Different application types have varying sensitivity towards various QoS attributes. Table 3-2 attempts to differentiate between various QoS attributes based on application type. Since QoS attributes are an essential part of the selection algorithm, further research on the subject will improve the efficiency of the algorithm. QoS attributes also form a part of the service advertisement data repository, hence application level constraints of QoS attributes will enable more competent use of the information available to the terminal for service selections.

6.1.7 6.2.3 Frequency Band

The MOSSA frequency band access mechanism stops short of listing the available frequencies that can be made available. Competent research in evaluating non-overlapping frequencies which do not interfere with existing (and future) wireless communication frequencies will facilitate more well-organised use of the spectrum and a more proficient use of the MOSSA service discovery mechanism.

References

M. Annoni, R. Hancock, T. Paila, E. Scarrone, R. Tonjes, L. Dell'Uomo, D. Wisely, and R. Mort, "Radio access networks beyond 3G: a first comparison of architectures," 2001, pp. G-133-G-140 vol.2

S. Bansal, "Quality of Service over Wireless Multimedia Networks," Middlesex University 2002

Bansal S, Lasebae A, and Comley R, "Global handover latency minimisation in HMIPv6," *WSEAS TRANSACTIONS on CIRCUITS and Systems,* vol. 2, pp. 58-65, January 2003 2003,1109-2734,

Bansal S., Lasebae A., and Comley R., " Freedom to choose along-with freedom of mobility," in *Information and Communication Technologies: From Theory to Applications, 2004*, Damascus, 2004, p. 191,0-7803-8482-2

M. Birchler, P. P. Smyth, G. Martinez, and M. Baker, "Future of Mobile and Wireless Communications." vol. 21: Kluwer Academic Publishers, 2003, pp. 11-21,1358-3948,

N. Blefari-Melazzi, D. Di Sorte, M. Femminella, L. Piacentini, and G. Reali, "Resources discovery and selection in wireless networks," 2004, pp. 3901-3905 Vol.7,0-7803-8533-0

L. Burness, D. Higgins, A. Sago, and P. Thorpe, "Wireless LANs - Present and Future," *BT Technology Journal,* vol. 21, pp. 32-47, 2003,1358-3948

C. Castelluccia, "A hierarchical mobility management scheme for IPv6," in *Computers and Communications, 1998. ISCC '98. Proceedings. Third IEEE Symposium*, 1998, pp. 305-309

P. M. L. Chan, R. E. Sheriff, Y. F. Hu, P. Conforto, and C. Tocci, "Mobility management incorporating fuzzy logic for heterogeneous a," *Communications Magazine, IEEE,* vol. 39, pp. 42-51, 20010163-6804

H. Chaskar, "Requirements of a Quality of Service (QoS) Solution for Mobile IP," *IETF,* 2002

A. Clapton, *Future Mobile Networks* First ed.: Institution Electrical Engineers, 2001,978-0852969830

J. Craninckx and S. Donnay, "4G terminals: how are we going to design them?," 2003, pp. 79-84

N. Damberg, D. G. Ternero, M. Jangelind, B. Jonsson, and M. Lind, "Handover extensions in Mobile IPv6," *Communication System Design (2G1319, 2G1705) at IMIT-KTH, Stockholm, Sweden, summer 2002,* 2002

C. Dannewitz, S. Berg, and H. Karl, "An IEEE 802.21-based Universal Information Service," *Proc. of the Wireless World Research Forum Meeting* 2006

G. Dommety, A. Yegin, C. Perkins, G. Tsirtsis, K. El-Malki, and M. Khalil, "Fasthandovers for mobile IPv6," *IETF,* 2001

A. Dutta, S. Das, D. Famolari, Y. Ohba, K. Taniuchi, T. Kodama, and H. Schulzrinne, "Seamless Mobility Across Heterogeneous Access Networks – An 802.21 Centric Approach," *WPMC,* 2005

H. Einsiedler, R. L. Aguiar, J. Jähnert, K. Jonas, M. Liebsch, R. Schmitz, P. Pacyna, J. Gozdecki, Z. Papir, J. I. Moreno, and I. Soto, "THE MOBY DICK PROJECT: A MOBILE HETEROGENEOUS ALL-IP ARCHITECTURE ". vol. 2007, 2001http://www-ks.rus.uni-stuttgart.de/Publications/MobyDickATAMS_final.pdf

S. M. Faccin, "Mechanism to discover 802.21 remote events and information services," USPTO, Ed. US: Nokia Corporation, 2006

F. Fang and D. S. Reeves, "Explicit proactive handoff with motion prediction for mobile IP," 2004, pp. 855-860 Vol.2,1525-3511

H. Feyrer, "Introduction to IPv6," *O'Reilly Onlamp.com,* 24 May 2001

W. Gang, P. J. M. Havinga, and M. Mizuno, "Wireless Internet over heterogeneous wireless networks," 2001, pp. 1759-1765 vol.3

W. Gang, M. Mizuno, and P. J. M. Havinga, "MIRAI architecture for heterogeneous network," *Communications Magazine, IEEE,* vol. 40, pp. 126-134, 20020163-6804

C. D. Gavrilovich, Jr., "Broadband communication on the highways of tomorrow," *Communications Magazine, IEEE,* vol. 39, pp. 146-154, 20010163-6804

R. Hsieh, A. Seneviratne, H. Soliman, and K. El-Malki, "Performance analysis on hierarchical Mobile IPv6 with fast-handoff over end-to-end TCP," 2002, pp. 2488-2492 vol.3,0-7803-7632-3

R. Hsieh, Z. G. Zhou, and A. Seneviratne, "S-MIP: a seamless handoff architecture for mobile IP," 2003, pp. 1774-1784 vol.3,0743-166X

D. Johnson, C. Perkins, and J. Arkko, "Mobility Support in IPv6, Internet proposed standard RFC 3775," *IETF,* 2004

M. d. Jonge, "Benchmarking Various Untrained Localization Algorithms," in *3rd Twente Student Conference on IT* Enschede, 2005

Kan Zhigang, Jian MA, Luo Jun, and Hu Jianping, "Mobile IPv6 and some issues for QoS," http://www.isoc.org/inet2001/CD_proceedings/T28/T28.htm

P. Kankaanpää, "Candidate Access Router Discovery," *Helsinki University of Technology Telecommunications software and multimedia laboratory,* 2003

G. Karagiannis, "Mobile IP: State of the Art Report," Ericsson 1999-07 1999,

G. Karagiannis and V. Rexhepi, "A Framework for QoS&Mobility in the Internet Next Generation," 20007 April 2007,http://citeseer.ist.psu.edu/440150.html

R. Koodli, "Fast Handovers for Mobile IPv6," *IETF,* 2005,rfc4068,

J. Korhonen, *Introduction to 3G Mobile Communications*, Second ed.: Artech House Publishers, 2003,978-1580535076

Y. Kun, T. Chris, and O. Shumao, "Model-based service discovery for future generation mobile systems," in *Proceeding of the 2006 international conference on Communications and mobile computing* Vancouver, British Columbia, Canada: ACM Press, 2006

H. Lacohée, N. Wakeford, and I. Pearson, "A Social History of the Mobile Telephone with a View of its Future," *BT Technology Journal,* vol. 21, pp. 203-211, 2003,1358-3948

A. Lasebae, S. Bansal, and R. Comley, "Service Advertisements for future mobile communication networks," in *IEEE GCC*, Kingdom of Bahrain, 2004,0-7803-9147-0

A. Lasebae and S. Bansal, "Service Advertisements for Fourth Generation Mobile Networks," 2006, pp. 3245-3250

S.-J. Lee, J.-J. Son, P.-Y. Joo, J.-Y. Jung, C.-K. Kim, and J. Yoo, "Apparatus and method for implementing handoff between heterogeneous networks in a wireless communication system ", USPTO, Ed. US: SAMSUNG ELECTRONICS CO., Ltd., 2007

P. Leo, M. Glenford, and H. Andy, "An end-system approach to mobility management for 4G networks and its application to thin-client computing." vol. 10: ACM Press, 2006, pp. 13-33,1559-1662,

X. Lin, R. Tonjes, T. Paila, W. Hansmann, M. Frank, and M. Albrecht, "DRiVE-ing to the Internet: Dynamic Radio for IP services in Vehicular Environments," 2000, pp. 281-289

S. Losi, "3G: Hard News or Hype?," Mobile Tech Today, 2001http://www.wirelessnewsfactor.com/perl/story/6325.html

O. W. Lourens and P. S. Kritzinger, "Cellular networks: past, present and future." vol. 7: ACM Press, 2000, pp. 4-ff35

K. Mahmud, W. Gang, M. Inoue, H. Murakami, and M. Mizuno, "Basic access network - the signaling-only network for power-efficient mobile multiservice terminals in MIRAI architecture," 2002, pp. 2385-2389 vol.4,1090-3038

N. McKay, "4G – turn-on, tune-in, no dropout." vol. 2006, 2002www.thefeature.com/article?articleide=15014

J. McNair and Z. Fang, "Vertical handoffs in fourth-generation multinetwork environments," *Wireless Communications, IEEE [see also IEEE Personal Communications],* vol. 11, pp. 8-15, 20041536-1284

S. Milroy, "3G Wireless Networks: Hype and Reality." vol. 2007: Mobile & Wireless Perspectives, 2002http://www.windowsitpro.com/Articles/ArticleID/23640/23640.html?Ad=1

mobileinfo.com, "4G - Beyond 2.5G and 3G Wireless Networks,"

N. Montavont, T. Noel, and M. Kassi-Lahlou, "Description and evaluation of mobile IPv6 for interfaces," 2004, pp. 144-148 Vol.1,1525-3511

S. Naghian, "Hybrid predictive handover in mobile networks," 2003, pp. 1918-1922 Vol.3,1090-3038

A. Neumann and A. Festag, "Hierarchical Mobile IPv6 Implementation Experiences," in *TKN, TU-Berlin, Siemens IPv6 Workshop* Berlin: TKN, 2002http://www.tkn.tu-berlin.de/publications/papers/ipv6_wrkshp_siemens_tkn_neumann_festag.ps

Ofcom, "Cellular & 3G Telephony." vol. 2007http://www.ofcom.org.uk/static/archive/ra/topics/pmc/document/licencetypes/cellularinfo.htm

O. Ormond, G. M. Muntean, and J. Murphy, "Economic Model for Cost Effective Network Selection Strategy in Service Oriented Heterogeneous Wireless Network Environment," 2006, pp. 1-4,1542-1201

O. Ormond, J. Murphy, and G. M. Muntean, "Utility-based Intelligent Network Selection in Beyond 3G Systems," 2006, pp. 1831-1836,8164-9547

C. Perkins, " IP Mobility Support - RFC2002," *IETF,* 1996

C. E. Perkins, *Mobile IP Design Principles and Practices*: Prentice Hall PTR, 1998,978-0201634693

C. Perkins, "IP Mobility Support for Ipv4 - RFC3220," *IETF,* 2002

Ramjee Prasad and Luis Munoz, *WLANs and WPANs Towards 4G Wireless*: Artech House Publishers, 2003,9781580530903

P. Reinbold and O. Bonaventure, "A Survey of IP micro-mobility protocols." vol. 2007, 2002http://citeseer.ist.psu.edu/rd/41876892%2C504352%2C1%2C0.25%2CDownload/http://citeseer.ist.psu.edu/cache/papers/cs/25932/http:zSzzSzwww.infonet.fundp.ac.bezSzdoczSzreportszSzCSTRei21Mar02.pdf/reinbold02survey.pdf

H. Sasaoka, *Mobile Communications* vol. Wave Summit Course: IOS Press, 2000,978-1-58603-100-8

J. Schiller, *Mobile Communications*, 2 ed.: Addison Wesley, 2003,978-0321123817

T. C. Schmidt and M. Wählisch, "Predictive versus Reactive—Analysis of Handover Performance and Its Implications on IPv6 and Multicast Mobility " *Telecommunication Systems Springer Netherlands,* vol. 30, 2005,1018-4864

H. Soliman, C. Castelluccia, K. E. Malki, and L. Bellier, "Hierarchical Mobile IPv6 Mobility Management (HMIPv6), Internet experimental RFC 4140," *IETF,* 2005

K. Suh, "Regional Mobile IPv6 mobility management," *IETF,* 2002

T. Paila, S. Alladin, M. Frank, T. Goransson, W. Hansmann, T. Lohmar, R. Toenjes, and L. Xu, "Flexible Network Architecture for Future Hybrid Wireless Systems," in *Mobile Summit 2001*, Barcelona, 2001

Tero Ojanpera and Ramjee Prasad, *WCDMA: Towards IP Mobility and Mobile Internet*: Artech House Publishers, 2001,978-1580531801

U. Varshney and R. Jain, "Issues in emerging 4G wireless networks," *Computer,* vol. 34, pp. 94-96, 20010018-9162

P. Vidales, R. Chakravorty, and C. Policroniades, "PROTON: A Policy-based Solution for Future 4G devices," in *Proceedings of 5th International Workshop on Policies for Distributed Systems and Networks*, 2004

P. Vidales, L. Patanapongpibul, G. Mapp, and A. Hopper, "Experiences with Heterogeneous Wireless Networks – Unveiling the Challenges," *Second International Working Conference on Performance Modeling and Evaluation of Heterogeneous Networks (HET-NETs),* 2004

Virginmobile, "OUR TARIFF." vol. 2007http://www.virginmobile.com/vm/genericContent.do?contentId=payasyougo.tariff.howdoi.sm131

H. J. Wang, R. H. Katz, and J. Giese, "Policy-enabled handoffs across heterogeneous wireless networks," 1999, pp. 51-60

J. Wang, "Broadband Wireless Communications - 3G, 4G and Wireless LAN," *The Kluwer International Series in Engineering and Computer Science,* vol. 620, 2001,079237391X,

Z. Wenhui, J. Jaehnert, and K. Dolzer, "Design and evaluation of a handover decision strategy for 4th generation mobile networks," 2003, pp. 1969-1973 vol.3,1090-3038

Z. Wenhui, "Handover decision using fuzzy MADM in heterogeneous networks," 2004, pp. 653-658 Vol.2,1525-3511

D. Wisely, *IP for 3G: Networking Technologies for Mobile Communications*: Wiley, 2002,978-0471486978

D. Wisely and E. Mitjana, "Evolving Systems Beyond 3G — the IST BRAIN and MIND Projects," *BT Technology Journal,* vol. 21, pp. 102-121, 200310.1023/A:1025115300186

X. P. Costa, R. Schmitz, H. Hartenstein, and M. Liebsch, "A MIPv6, FMIPv6 and HMIPv6 handover latency study: analytical approach," in *IST Mobile & Wireless Telecommunications Summit 2002* Thessaloniki, Greece, 2002http://newton.ee.auth.gr/ist2002/papers/SessionM4/0403181.pdf

G. Xia, W. Gang, and T. Miki, "End-to-end QoS provisioning in mobile heterogeneous networks," 2003, pp. 1361-1366 vol.2,1525-3511

R. Yaqub and S. Madhani, "IP NETWORK INFORMATION DATABASE IN MOBILE DEVICES FOR USE WITH MEDIA INDEPENDENT INFORMATION SERVER " in *USPTO*, USPTO, Ed. US: TOSHIBA AMERICA RESEARCH, INC., 2007

F. Zhu and J. McNair, "Optimizations for vertical handoff decision algorithms," in *Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE*, 2004, pp. 867- 872,0-7803-8344-3

Appendix A. Database Sample data, queries and program to compute time taken to execute a query

The table below provides a sample of the data used for testing MOSAA service discover architecture. The table is followed by the queries executed on the sample data. The program used to execute queries and compute response times is also included.

ID	Provider	Provider_ID	Frequency (MHz)	Service	Service Description	Bandwidth (Mbps)	QoS class	Security class	Interface	Price	Location Coordinates (X, Y, Z)	Repository Link	Network Coverage Area (Metres)	MAP IP
1	Vodafone	1	3000	Voice	Real time Voice	22	1	1	3G	£0.2		repository.vodafone.co.uk		10.11.24.56
4	Vodafone	1	3000	Data	Browsing	22	1	1	3G	£0.05		repository.vodafone.co.uk		10.11.24.56
5	Vodafone	1	3000	Video	Video	22	1	1	3G	£0.5		repository.vodafone.co.uk		10.11.24.56
6	BT	2	3000	Voice	Real time Voice	22	1	1	3G	£0.1		repository.bt.co.uk		11.10.24.01
7	BT	2	3000	Data	Browsing	22	1	1	3G	£0.08		repository.bt.co.uk		11.10.24.01
8	BT	2	3000	Video	Video	22	1	1	3G	£0.7		repository.bt.co.uk		11.10.24.01
9	Starbucks	15	11	Data	Browsing	11	0	0	802.11b	£0.02	101,500,650		50	156.50.75.89
10	Brighton	63	11	Data	Browsing	11	0	0	802.11b	£0	125,750,690		50	201.20.35.63

Query 1 Parameter

```
SELECT *
FROM Service
WHERE Service.Provider_id=2;
```

Query 2 Parameters

```
SELECT *
FROM Service
WHERE Service.Provider_id=2 And Frequency="3000";
```

Query 3 Parameters

```
SELECT *
FROM Service
WHERE Service.Provider_id=2 And Frequency="3000" And Service="data";
```

Query 4 Parameters

```
SELECT *
FROM Service
WHERE Service.Provider_id=2 And Frequency="3000" And Service="data" And
Bandwidth>=10;
```

Query 5 Parameters

```
SELECT *
FROM Service
WHERE Service.Provider_id=2 And Frequency="3000" And Service="data" And
Bandwidth>=10 and QoS>=0;
```

Query 6 Parameters

```
SELECT *
FROM Service
WHERE Service.Provider_id=2 And Frequency="3000" And Service="data" And
Bandwidth>=10 And QoS>=0 And Security>=0;
```

Query 7 Parameters

SELECT *

FROM Service

WHERE Service.Provider_id=2 And Frequency="3000" And Service="data" And Bandwidth>=10 And QoS>=0 And Security>=0 And Interface Like "3G";

Query 8 Parameters

SELECT *

FROM Service

WHERE Service.Provider_id=2 And Frequency="3000" And Service="data" And Bandwidth>=10 And QoS>=0 And Security>=0 And Interface Like "3G" And Price<=0.15;

Query 9 Parameters

SELECT *

FROM Service

WHERE Service.Provider_id=2 And Frequency="3000" And Service="data" And Bandwidth>=10 And QoS>=0 And Security>=0 And Interface Like "3G" And Price<=0.15 And (Network_Coverage Is Null Or Network_Coverage>40);

```
/*
 * Author:  Saurabh Bansal (s.bansal@mdx.ac.uk)
 * Purpose:    Visual basic program to compute time taken to execute the queries
mentioned in
 *          Appendix A
 * Copyright (2007) by the author, all rights reserved.
 */
Option Compare Database
Private Sub Command1_Click()
On Error GoTo Err_Command1_Click


    Screen.PreviousControl.SetFocus
    DoCmd.DoMenuItem acFormBar, acEditMenu, 10, , acMenuVer70

Exit_Command1_Click:
    Exit Sub

Err_Command1_Click:
    MsgBox Err.Description
    Resume Exit_Command1_Click

End Sub
Private Sub Command2_Click()
On Error GoTo Err_Command2_Click

Dim StartTime As Long, EndTime As Long

Dim ElapsedTime1 As Long
Dim ElapsedTime2 As Long
Dim ElapsedTime3 As Long
Dim ElapsedTime4 As Long
Dim ElapsedTime5 As Long
Dim ElapsedTime6 As Long
```

```
Dim ElapsedTime7 As Long
Dim ElapsedTime8 As Long
Dim ElapsedTime9 As Long

Dim stDocName As String

StartTime = Timer * 1000
   stDocName = "Query1"
   DoCmd.OpenQuery stDocName, acNormal, acEdit
EndTime = Timer * 1000
'MsgBox "The elapsed time is " & CStr((EndTime - StartTime))
ElapsedTime1 = EndTime - StartTime
StartTime = 0
EndTime = 0
'----Query 2

StartTime = Timer * 1000
   stDocName = "Query2"
   DoCmd.OpenQuery stDocName, acNormal, acEdit
EndTime = Timer * 1000
ElapsedTime2 = EndTime - StartTime
StartTime = 0
EndTime = 0


'----Query 3
StartTime = Timer * 1000
   stDocName = "Query3"
   DoCmd.OpenQuery stDocName, acNormal, acEdit
EndTime = Timer * 1000
ElapsedTime3 = EndTime - StartTime
StartTime = 0
EndTime = 0
```

```
'----Query 4
StartTime = Timer * 1000
   stDocName = "Query4"
   DoCmd.OpenQuery stDocName, acNormal, acEdit
EndTime = Timer * 1000
ElapsedTime4 = EndTime - StartTime
StartTime = 0
EndTime = 0


'----Query 5
StartTime = Timer * 1000
   stDocName = "Query5"
   DoCmd.OpenQuery stDocName, acNormal, acEdit
EndTime = Timer * 1000
ElapsedTime5 = EndTime - StartTime
StartTime = 0
EndTime = 0


'----Query 6
StartTime = Timer * 1000
   stDocName = "Query6"
   DoCmd.OpenQuery stDocName, acNormal, acEdit
EndTime = Timer * 1000
ElapsedTime6 = EndTime - StartTime
StartTime = 0
EndTime = 0


'----Query 7
StartTime = Timer * 1000
   stDocName = "Query7"
```

```
   DoCmd.OpenQuery stDocName, acNormal, acEdit
EndTime = Timer * 1000
ElapsedTime7 = EndTime - StartTime
StartTime = 0
EndTime = 0


'----Query 8
StartTime = Timer * 1000
   stDocName = "Query8"
   DoCmd.OpenQuery stDocName, acNormal, acEdit
EndTime = Timer * 1000
ElapsedTime8 = EndTime - StartTime
StartTime = 0
EndTime = 0


'----Query 9
StartTime = Timer * 1000
   stDocName = "Query9"
   DoCmd.OpenQuery stDocName, acNormal, acEdit
EndTime = Timer * 1000
ElapsedTime9 = EndTime - StartTime
StartTime = 0
EndTime = 0


MsgBox "The elapsed time is " & CStr(ElapsedTime1)
MsgBox "The elapsed time is " & CStr(ElapsedTime2)
MsgBox "The elapsed time is " & CStr(ElapsedTime3)
MsgBox "The elapsed time is " & CStr(ElapsedTime4)
MsgBox "The elapsed time is " & CStr(ElapsedTime5)
MsgBox "The elapsed time is " & CStr(ElapsedTime6)
MsgBox "The elapsed time is " & CStr(ElapsedTime7)
```

```
MsgBox "The elapsed time is " & CStr(ElapsedTime8)
MsgBox "The elapsed time is " & CStr(ElapsedTime9)


Exit_Command2_Click:
   Exit Sub

Err_Command2_Click:
   MsgBox Err.Description
   Resume Exit_Command2_Click

End Sub
```

Appendix B. Selection algorithm test data

The table below provides a sample of data used to evaluate the MOSAA selection algorithm. The parameters used for evaluation include cost, bandwidth, latency, jitter, BER, packet loss and power.

Cost (Pence/ Minute)	Bandwidth (Mbps)	Latency (Seconds)	Packet Loss (Lost Packets/Million)	Jitter (UI)	BER (Erroneous bits/Million)	Power (W)
0.016235	93.30346	0.543959	0.43351696	0.781	0.476275	0.828719
0.354227	64.12062	0.497526	0.41334303	0.653	0.116566	0.702435
1.267372	64.15847	0.215143	0.12789259	0.244	0.83484	0.454281
1.425167	49.33125	0.033175	0.02269763	0.125	0.578294	0.056638
0.472655	18.41649	0.894871	0.74221861	0.54	0.793962	0.795423
1.942768	17.45801	0.762976	0.40101707	0.6	0.258836	0.4269
1.981939	52.26542	0.198604	0.70519496	0.931	0.002317	0.265779
0.893234	43.55552	0.799054	0.7254005	0.37	0.002116	0.75838
1.3113	35.24186	0.314523	0.1412861	0.782	0.94392	0.125501
1.229609	20.5151	0.999466	0.97022785	0.819	0.742943	0.721303
0.826071	63.47232	0.828085	0.82947289	0.465	0.322513	0.617301
1.877214	26.8311	0.293242	0.13824245	0.89	0.322798	0.216313
1.51824	53.08662	0.848593	0.24162182	0.179	0.98755	0.382752
0.004511	36.0531	0.986512	0.7499556	0.525	0.213069	0.324424
0.589381	36.45874	0.823823	0.98036443	0.37	0.44362	0.923237
1.473128	0.800352	0.813936	0.8745598	0.263	0.707472	0.566104
1.329102	76.6602	0.010019	0.60457742	0.766	0.466605	0.39521
0.912336	23.96453	0.768035	0.08545237	0.439	0.579111	0.735789
0.231437	68.08653	0.989622	0.2910919	0.025	0.073745	0.167246
0.164074	89.60763	0.576463	0.65377394	0.925	0.40056	0.314424
1.643211	19.84384	0.075099	0.7580572	0.091	0.146713	0.048071
0.523354	13.90059	0.388652	0.16246171	0.748	0.215926	0.093558
0.471805	1.004491	0.862363	0.47855502	0.1	0.698814	0.245482
0.611115	9.186944	0.461052	0.20272451	0.169	0.179821	0.832302
1.077535	44.92966	0.752862	0.07316712	0.786	0.308736	0.909441

Appendix C. Selection Algorithm Implementation

The Appendix includes programs used for evaluating the MOSAA selection algorithm. The first program evaluated the interface compatibility of the MOSAA selection algorithm. The second program evaluated the affect of eliminations on the MOSSA selection algorithm latency.

```
/*
 * File: elim1.java
 * Author:  Saurabh Bansal
 * Purpose: The program evaluates the selection algorithm developed under the
MOSSA architecture
 * Copyright (2007) by the author, all rights reserved.
 */
import java.util.Random;
import java.io.*;
public class elim1 {
 public static void main(String args[]) throws IOException {
        BufferedReader stdin = new BufferedReader( new InputStreamReader(
System.in ) );

//variables to store time
long elp1, elp2, elp;
 double func;
  int eliminated=0;

  Random generator = new Random();
 int number_of_interfaces =0;
 int[] int_in_use = new int[10];
 int[][] int_available =  new int[10][10];
 int[] good_int = new int[10];
  String input;
        //initialize interfaces
  System.out.println("Int  0  1  2  3  4  5  6  7  8  9");
  for (int i=0;i<10;i++){
```

```
            System.out.print(i+"     ");
            for (int j=0;j<10;j++){
                  int_available[i][j]= generator.nextInt(3); //0=bad,1=ok,2=good
                  System.out.print(int_available[i][j] + "  ");
                  }
            System.out.print("\n");
      }
      System.out.print("\n");
      System.out.print("\n");
      System.out.print("\n");

      //Interface compatibility based eliminations
      System.out.print ("Enter the no. of interfaces in use:");
      input = stdin.readLine();
      number_of_interfaces = Integer.parseInt( input );
      System.out.print("\n");
      for (int i=0;i<number_of_interfaces;i++){
            System.out.print ("Interface " + i + " in use : ");
            input = stdin.readLine();
            int_in_use[i]= Integer.parseInt( input );
            System.out.print("\n");
      }

      System.out.print("\n");
      System.out.print("\n");
      System.out.print("\n");

      elp1 = System.currentTimeMillis();
      for (int i=0;i<10;i++){
            for (int j=0;j<number_of_interfaces;j++){
                  int temp1 =int_in_use[j];
                  if (int_available[i][temp1]==2){
                        good_int[i]=2;
                  } else if (int_available[i][temp1]==1){
```

```
                        good_int[i]=1;
                }else {
                        good_int[i]=0;
                        break;
                }
        }
}
//System.out.println ("Total time for processing \t" + elp1 + "\t" + elp2 + "\t" + elp);

for (int i =10000; i>=0;i=i-1000){
        //elp1 = System.currentTimeMillis();
        int service_int=generator.nextInt(9);

        if (good_int[service_int]==1){
        func = (.33 * Math.log(1/.35));
        func = func + (.33 * ((.2 * Math.log (10)) + (.2 * Math.log (1/10)) + (.2 *
Math.log( 1/100)) + (.2 * Math.log(1/.45)) + (.2 * Math.log (1/1000000))));

        func = func+ (.33 * Math.log (1/100));
        eliminated=eliminated+1;

        System.out.println(eliminated);
}
}
elp2 = System.currentTimeMillis();

elp = elp2-elp1;

System.out.println (eliminated +"\t" + elp1 + "\t" + elp2 + "\t" + elp);
}
}
```

```
/*
 * File: elim2.java
 * Author:  Saurabh Bansal
 * Purpose: The program evaluates the selection algorithm developed under the
MOSSA architecture
 * Copyright (2007) by the author, all rights reserved.
 */
public class elim2 {
public static void main(String args[]) {

        long elp1, elp2,elp3,elp4,elp0 ,elp;
        double func;

        elp1 = System.currentTimeMillis();

        func = (.33 * Math.log(1/.35));

        func = func + (.33 * ((.2 * Math.log (10)) + (.2 * Math.log (1/10)) + (.2 *
Math.log( 1/100)) + (.2 * Math.log(1/.45)) + (.2 * Math.log (1/1000000))));

        func = func+ (.33 * Math.log (1/100));
        elp2 = System.currentTimeMillis();
        elp0 = elp2-elp1;
        System.out.println ("Without elimination - " + "\t" + elp1 + "\t" + elp2 + "\t" +
elp0);

        for (int i =10000; i>=0;i=i-1000){
                int j =0;
                elp1 = System.currentTimeMillis();
                for (j = 0; j<i; j++){
                        func = (.33 * Math.log(1/.35));
```

```
                func = func + (.33 * ((.2 * Math.log (10)) + (.2 * Math.log
(1/10)) + (.2 * Math.log( 1/100)) + (.2 * Math.log(1/.45)) + (.2 * Math.log
(1/1000000))));

                func = func+ (.33 * Math.log (1/100));
            }
        elp2 = System.currentTimeMillis();

      elp = elp2-elp1;
      System.out.println (i + "\t" + j + "\t" + elp1 + "\t" + elp2 + "\t" + elp);

    }
  }
}
```

Appendix D. Published Papers

1) Bansal S, Lasebae A, and Comley R, "Global handover latency minimisation in HMIPv6," *WSEAS TRANSACTIONS on CIRCUITS and Systems,* vol. 2, pp. 58-65, January 2003 2003,1109-2734

2) Bansal S., Lasebae A., and Comley R., " Freedom to choose along-with freedom of mobility," in *Information and Communication Technologies: From Theory to Applications, 2004*, Damascus, 2004, p. 191,0-7803-8482-2

3) A. Lasebae, S. Bansal, and R. Comley, "Service Advertisements for future mobile communication networks," in *IEEE GCC*, Kingdom of Bahrain, 2004,0-7803-9147-0

4) A. Lasebae and S. Bansal, "Service Advertisements for Fourth Generation Mobile Networks," 2006, pp. 3245-3250

Global Handover Latency Minimisation in HMIPv6

SAURABH BANSAL, ABOUBAKER LASEBAE, RICHARD COMLEY
School of Computer Science
Middlesex University
Bounds Green Road, London N112NQ
UK
{s.bansal, a.lasebae, r.comley}@mdx.ac.uk
http://www.cs.mdx.ac.uk/research/PhDArea/research_students/Saurabh/sb_profile.html

Abstract –

Numbers of mobile users are growing at an exponential rate and the future looks bright [11]. Mobile IP was introduced to satisfy the growing demand for IP based services by the mobile users. The growing demand for delay sensitive applications [4], e.g. audio streaming, required Mobile IP to support low latency seamless handoffs. Hierarchical Mobile IP (HMIP) was introduced to reduce average handover latency by differentiating between local and global handoffs [3]. HMIP has attracted a lot of attention from the research community. The main problem with Hierarchical Mobile IP is when a mobile node undergoes a global handover it undergoes a triple binding update, which adds to handover latency. We propose a seamless handover architecture that builds on top of hierarchical Mobile IP to facilitate global seamless handover whilst reducing the handover latency. We propose coverage area overlaps between domains to facilitate seamless transition from one domain to another. Using mathematical analysis, we argue that our proposal effectively reduces global handover latency. Lower handover latency along with seamless intra-domain transition further reduces packet loss. Our proposal is fully compatible with Hierarchical Mobile IP and Mobile IPv6.

Key-Words: - Mobile IP, Hierarchical Mobile IP, Seamless Handoff

Introduction

Although the Internet offers access to information sources worldwide, typically we do not expect to benefit from that access until we arrive at some familiar point whether home, office, or school. However, the increasing variety of wireless devices offering IP connectivity, such as PDA's, handhelds, and digital cellular phones, is beginning to change our perceptions of the Internet. By the use of the conventional network configuration, a computer must be shut down and modify its network settings when moving or connecting to a different network. This causes much more inconvenience for the end users. The situation becomes even worse when the computer keeps roaming and switching to different networks. To solve these problems and to provide real time mobility to the mobile users Mobile IP was introduced. Mobile IP is a modification of IP that allows nodes to continue to receive datagram's (i.e. UDP packets) no matter where they happen to be attached to the Internet. It involves some additional control messages that allow the IP nodes involved to manage their IP routing tables reliably [10]. Mobile IP deploys a Home Agent (HA) which intercepts packets bound for the Mobile Node (MN) and tunnels them to it's current point of attachment to the network via a foreign agent. Tunnelling (or triangle routing) ensures that MN continues to receive it's packets but MN has to ensure that HA knows about its current point of attachment. MN notifies it's HA when changing in between networks but this indirection requires a registration process and address resolution procedure. The indirection procedure has been shown to result in long handover latencies, which often leads to packet loss and performance degradation [7].

Hierarchical Mobile IP

Macro mobility schemes overcome this problem by differentiating between inter and intra domain movement. Hierarchical Mobile IP (HMIP) provides a scheme for performing registrations locally in the foreign network, thereby reducing the number of signalling messages forwarded to the home network as well as lowering the signalling latency that occurs when a mobile node moves from one Base Station (BS) to another. HMIP introduces a conceptual entity, Mobility Anchor Point (MAP), to separate local mobility from global mobility. MAP acts as a Mobile Node's (MN) Home Agent (HA) in the foreign network. MAP intercepts all packets destined for MN and tunnels them to its on-link CoA (LCoA).

In HMIP, MN attaches itself to two Care of Addresses (CoA), a regional (or site) address and an on-link (or point of attachment) address. Figure 1 shows a typical HMIP network architecture. The regional and local CoA's of the MN are formed in a stateless manner, i.e. MN forms these addresses based on the site (and local) prefix in router advertisements. To avoid duplicate addresses all the nodes run a duplicate address detection algorithm.

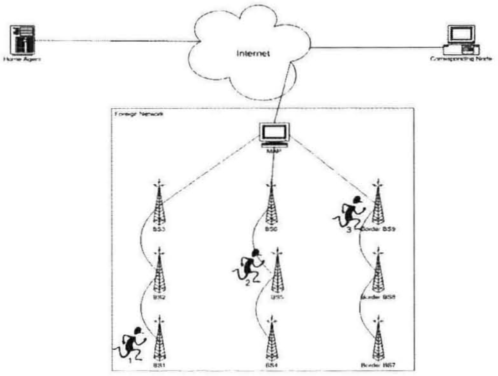


Figure 1 Hierarchical Mobile IP network architecture

Global Handover HMIP

When MN enters a new site, it goes through a global handover. A global handover consists of these stages –

- MN forms a new stateless address based on the BS prefix and sends a Binding Update (BU) to the BS.
- BS updates it's binding cache and sends a Binding Acknowledgement (BAck) to the MN.
- MN forms a new stateless site address based on the MAP address in the router advertisements and sends a BU to MAP.
- MAP updates it's binding cache and sends a BAck to the MN.
- MN sends a BU to it's HA (and CN's) notifying it of it's site movement.
- HA updates it's binding cache and sends a BAck to the MN.

Figure 2 shows the sequence of message transfers between MN, BS, MAP and HA during the handover. In the process, we have assumed that there were no duplicate addresses detected.

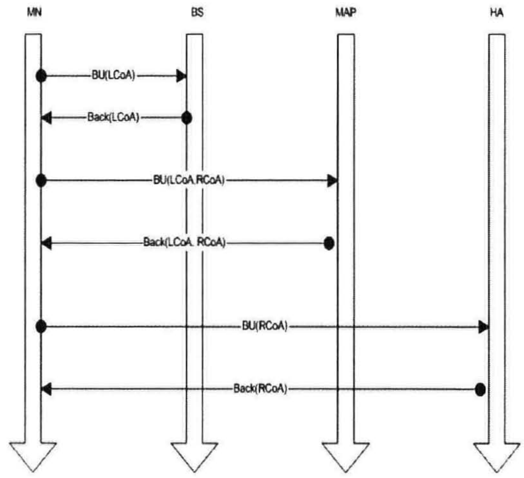


Figure 2 Interaction between MN and BS, MAP, HA during a global handover

Local Handover HMIP

Local handover (handover within a MAP domain) consists of the following stages –

- MN forms a new stateless local address based on the router advertisements and sends a BU to BS.
- BS updates it's binding cache and sends a BAck to the MN.
- MN sends a BU to the MAP notifying it of it's site movement.
- MAP updates it's binding cache and sends a BAck to the MN.

The main differences between local and global handover is that MN does not need to notify HA of it's movements within a site thus reducing handover latencies.

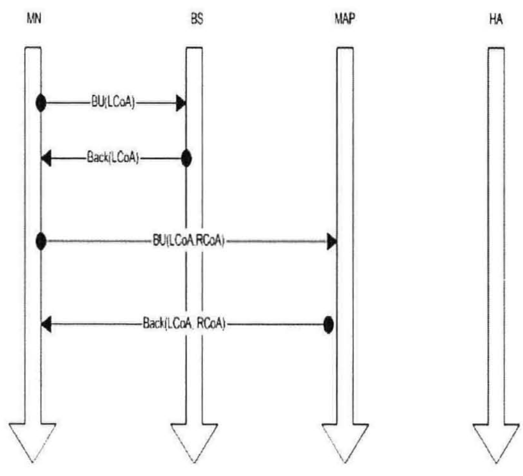


Figure 3 Interaction between MN, BS, MAP and HA during a local handover

Mobility Management Architecture

Hierarchical mobile IP reduces the signalling load on the internet by differentiating between local and global mobility. Global (or intra-MAP) handovers requires MN to undergo three BU's, thus increasing the global handover latency.

In large networks with more than one MAP when an MN moves out of the scope of one MAP it would undergo a global handover. We propose to introduce Border Base Stations (BBS), which would serve as a bridge between two MAP domains. BBS adverts would hold information of both the domains and would enable MN to register with the adjacent network before even entering it, thus allowing a cushion by enabling pre-emptive handovers.

Large networks should be divided into smaller sub-networks to allow better load balancing. Its own MAP would serve each sub-network. Reducing the load on MAP's would enable them to assume more responsibilities like Session Initiation Protocol (SIP) proxy and redirect server [5]. Enabling SIP on top of our architecture will be the future direction taken by this project.

Figure 4 shows a large network being divided into sub-networks.

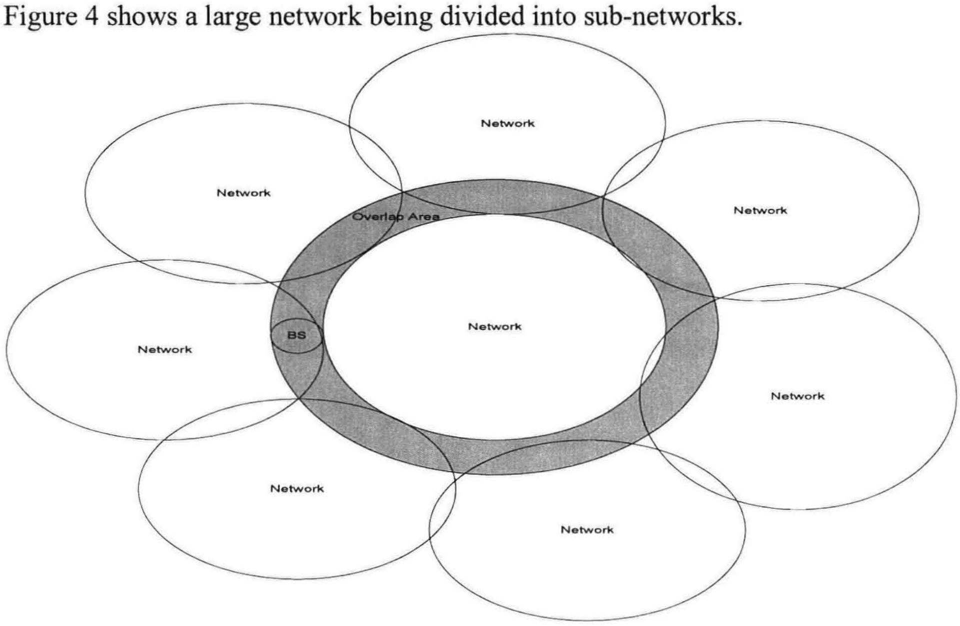


Figure 4: Network divided in sub-networks with overlap area marked in grey

We propose a hierarchical architecture, which would increase the global handoff performance.

Protocol Overview

HMIPv6 differentiates between inter-site and intra-site mobility [3]. HMIPv6 completely hides intra-site mobility from a communicating host. When a mobile node goes through an inter-site, handover it's regional and private bindings would have to be updated. Triple binding updates could lead to delays hence packet losses and in extreme cases loss of communication channel. To provide better Quality of Service (QoS) handover latency has to be minimized.

Figure 5 illustrates movement of a typical MN with its movement marked as stages 1 to 6. To describe the behaviour protocol, operations performed in different stages will be listed.

Model network

The Home Agent (HA) acts as the Home Agent of the Mobile node. Mobile node communicates with the correspondent node. The network (figure 2) comprises –

- Switching Centres (SC) – Switching centres would be responsible for providing regional registrations (RCoA) to the mobile nodes and maintain bindings for mobiles currently visiting the network. The switching centre also behaves as the border router for the site. Separate machines could perform all the three tasks but because of the reduced load as compared HMIPv6 all of them could be preformed by one machine. A switching centre could also perform as a gateway to the internet or could be linked up to the networks gateway. SC could also act as a Home Agent (HA) thus enabling distribution throughout the network. The area served by the SC as a MAP would be termed as a Switching Center Location Area (SCLA) in the rest of the document.
- Base Stations (BS) – BS would be a router capable of handling MNs in its Virtual Location Area (VLA). A VLA is the area range within which the MB is capable of handling an MN. BS should send advertisements from time to time and register new clients when requested. BS should be able to route MN's packets correctly and process handovers when requested by a node.
- Border Base Stations (BBS) – BBS would perform all the tasks of a BS but their advertisements would hold details of MAP's in adjacent networks. BBS would be wired to two networks forming a bridge. BBS serve in the border regions of two adjoining networks shown as the shaded region in figure 4.
- Mobile Nodes (MN) - Since MN's are termed as devices with low power, all the efforts have been made to keep the processing by an MN down to minimum. The MN would be effectively doing the same tasks as they have been doing. The common tasks include listen to router advertisements and register with the foreign agent, to be able to de-capsulate packets sent, etc. No new tasks would be performed by the MN.
- Home Agent (HA) – Performs all duties as described in MIPv6 [9].
- Correspondent Node (CN) – CN could be a wired or wireless terminal communicating with MN in the same or different networks.

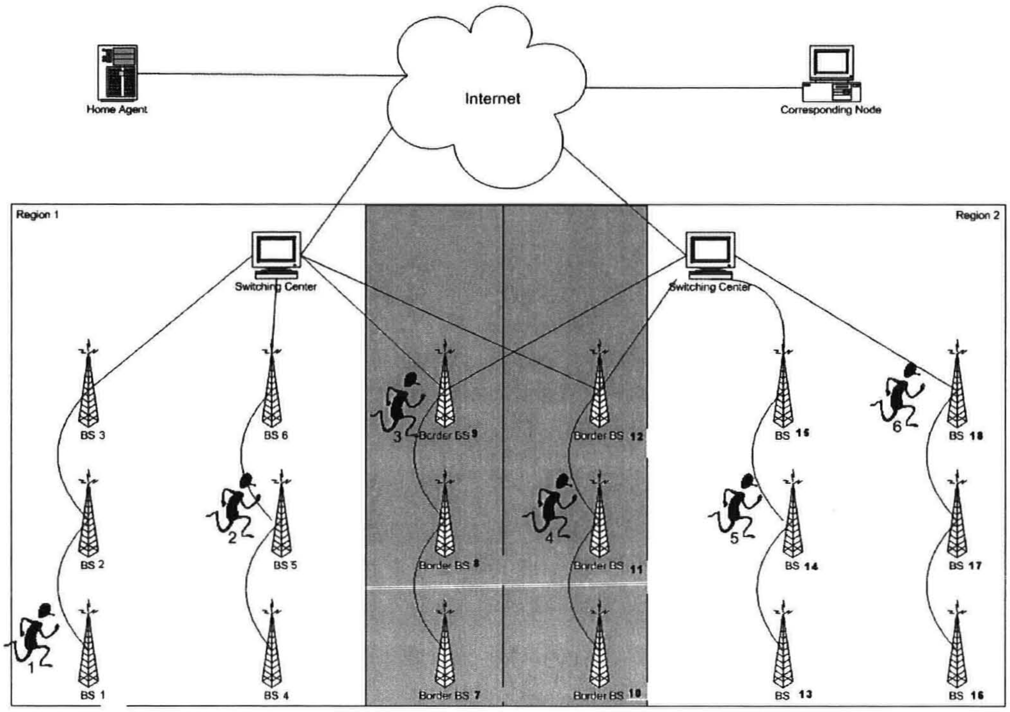


Figure 5 - Sample Network

Protocol Details

To describe the behavior of operation performed by the network as MN travels through the network from stages 1 to 6 in the model network (figure 5).

Stage 1. MN enters the VLA of BS1 and starts receiving router advertisements. MN forms a stateless link address (LCoA) and sends a BU to BS1. BS1 on receipt of BU updates it's binding cache and sends a BAck to MN. MN forms a stateless site address (RCoA) and sends a BU to SC1. SC1 on receipt of BU updates it's binding cache and sends a BAck to MN. MN sends a BU (RCoA) to it's HA and CN's. HA (and CN's) on receipt of the BU updates it's binding cache and sends a BAck to MN.

Stage 2. MN moves in the VLA of BS5 and starts receiving router advertisements. MN forms a LCoA and sends a BU to BS5. BS5 on receipt of BU updates it's binding cache and sends a BAck to MN. MN sends a BU to SC1. SC1 updates it's binding cache and sends a BAck to MN.

Stage 3. MN moves in the VLA of BS9 and starts receiving router advertisements. MN forms a stateless LCoA and sends a BU to BS9. BS9 on receipt of BU updates it's binding cache and sends a BAck to MN. Router advertisements from BS9 would contain details of MAP's in regions 1 and 2. Since, MN is already registered in region 1 it would have a choice of also registering in region 2. MN forms a stateless secondary RCoA (SRCoA) for region 2 and send a BU to SC2. SC2 on receipt of BU updates it's binding cache and send a BAck to MN. MN sends a BU (RCoA, SRCoA, LCoA) to SC1.

Stage 4. MN moves in the VLA of BS11 and starts receiving router advertisements. MN forms a stateless LCoA and send a BU to BS11. BS11 on receipt of BU updates it's binding cache and sends a BAck to MN. Since MN has moved in region 2 it's SRCoA now becomes it's RCoA and RCoA now becomes SRCoA. So MN's new binding is $RCoA_{NEW} \rightarrow SRCoA_{OLD}$, $SRCoA_{NEW} \rightarrow RCoA_{OLD}$, $LCoA_{NEW}$. MN sends a binding update to it's home agent (RCoA). MN sends a binding update (RCoA, SRCoA, LCoA) to SC1 and SC2.

Stage 5. MN moves in the VLA of BS14 and starts receiving router advertisements. MN forms a stateless LCoA and sends a BU to BS14. BS14 on receipt of BU updates it's binding cache and sends a BAck to MN. MN drops it's SRCoA and sends a binding update (RCoA, LCoA) to SC2.

Stage 6. MN moves in the VLA of BS16 and starts receiving router advertisements. MN forms a LCoA and sends a BU to BS16. BS16 on receipt of BU updates it's binding cache and sends a BAck to MN. MN sends a BU to SC1. SC2 updates it's binding cache and sends a BAck to MN.

Comparison with HMIPv6

HMIP allows an MN to hold more than 1 site registrations at a given time but we provide a framework for intra-domain handovers to progress as seamlessly as possible. MN would registers itself in more than one site when in border areas and based on it's future movements handovers occur as discussed in section 4.2. Whilst in border areas MN would be served by SC's of both regions hence minimizing packet loss probability. Previous SC would hold the route to MN's new SC until MN moves out of the border areas or if it receives a tear down message from MN.

	HMIPv6	Proposed Protocol
Inter-site Mobility	MN gets a new LCoA on every link within the network and it sends a binding update to the site mobility server	Same as HMIPv6 but MN would need to register it's secondary RCoA when it enters border areas.
Intra-site Mobility	MN gets a new RCoA and LCoA. MN sends binding updates to it's home agent and the site mobility server.	MN just gets a LCoA and secondary RCoA becomes it's RCoA. MN sends binding updates to it's home agent and site mobility server.
Packet Delivery	Correspondent Node sends a packet to the MN using its RCoA. Site mobility, a server intercepts and tunnels the packet to the current LCoA of the mobile node and sends it.	Same as HMIPv6.

Table 1 Comparison with HMIPv6

Performance evaluation

Global handover latency

We assume that the handover is initiated by the Mobile Node. For the wireless delay we assume the same value for the uplink and downlink delay. We assume that stateless address formation processing time is negligible and there are no duplicate address formed.

HMIPv6

Consider the scenario of handover between BS9 – BS11.

Figure 2 illustrates the interactions taking place in event of a global handover. The global handover requires MN to send BU's to BS, MAP and HA.

Hence,

Global Handover Latency $= BU_{BS} + BU_{MAP} + BU_{HA}$

Where –

BU_{BS} is the time taken by MN to send one or more BU to BS and get an acknowledgement

BU_{MAP} is the time taken by MN to send one or more BU to BS and get an acknowledgement

BU_{HA} is the time taken by MN to send one or more BU to BS and get an acknowledgement

A BU is confirmed when MN receives a BAck. Hence, the BU latencies of BS, MAP and HA would be

$$BU_{BS} = (\#BU_{BS} + 1)Wd \quad [12]$$

Where –

$(\#BU_{BS} + 1)$ is the number of binding updates sent to BS and we add 1 for binding acknowledgement

Wd is the delay introduced by the wireless part of the network

The communication between BS and MN would be wireless so the round trip times would only be affected by wireless delays. MN would send 1 or more BU's and wait for a BAck to confirm BU.

$$BU_{MAP} = ((\#BU_{MAP} + 1)Wd + Ld_{BS-MAP})$$

Where –

$(\#BU_{MAP} + 1)$ is the number of binding updates sent to MAP and we add 1 for binding acknowledgement

LD_{BS-MAP} latency introduced by the wired part between BS and MAP

The communication between BS and MAP would require packets to travel via air (wireless) to the BS and then travel the wired part between BS and MAP. MN would send 1 or more BU and wait for a BAck to confirm BU.

$$BU_{HA} = ((\#BU_{HA} + 1)Wd + Ld_{BS-HA})$$

Where –

($\#BU_{HA}$+1) is the number of binding updates sent to the Home Agent and we add 1 for binding acknowledgement

LD_{BS-HA} latency introduced by the wired part between Bs and HA

The communication between BS and HA would require packets to travel via air (wireless) to the BS and then travel the wired part between BS and HA. MN would send 1 or more BU and wait for a BAck to confirm BU.

So –

$$\text{Global Handover Latency} = BU_{BS} + BU_{MAP} + BU_{HA}$$
$$= (\#BU_{BS} + 1)Wd +$$
$$((\#BU_{MAP} + 1)Wd + Ld_{BS-MAP}) +$$
$$((\#BU_{HA} + 1)Wd + Ld_{BS-HA})$$

MN would also need to notify correspondent nodes (CN) of it's site movement but that can be done simultaneously with Home Agent BU. We assume BU_{HA} takes less time than BU_{CN}

Our proposal

In our proposal MN would initiate the registration process once it enters the BBS VLA. So by the time MN reaches BS11 we assume the BU_{MAP} phase would have been completed. Hence, we equate BU_{MAP} to '0'.

$$\text{Global Handover Latency} = BU_{BS} + BU_{MAP} + BU_{HA}$$

Where $BU_{MAP} = 0$ since it has already been processed.

Therefore

$$\text{Global Handover Latency} = BU_{BS} + BU_{HA}$$

BU_{HA} would be processed whilst MN is in border areas thus effectively enabling a seamless global handover.

The MN's primary site, RCoA, would be registered with the HA based on MN's movements within the border areas. Border areas enable a smooth transition from one network to another and in most cases should provide a lossless handover.

MN would get enough time to process BU_{HA} before leaving the border areas.

Local Handover latency

The local handover latency would not be affected by our proposal. The local signaling load would be increased when the MN is in the border areas because MN would have to notify MAP's of both the regions about it's movements.

Conclusion

HMIPv6 reduces the signaling load by differentiating between local and global mobility but when performing a global update MN has to go through triple binding updates. Analysis has shown that it is further possible to optimize the handover performance of Mobile IP [7]. Tests have shown that HMIPv6 outperforms Mobile IPv6 in local handovers but global handovers take more time [2].

Our proposal reduces the registration processing time because MN would be able acquire regional registration even before entering the site. Triple binding updates could lead to MN's packets being lost but in our scenario MN would have to cross two VLA's before the binding update is received by the correspondent nodes and home agent. Hence MN is provided ample time to get regional registrations and send BU's. Our proposal does not affect local mobility but rather tries to make global mobility as seamless as possible. We have shown that our proposal further reduces global handover latency by introducing intra-domain coverage area overlaps (BBS). BBS would serve as a bridge between adjoining domains thus enabling a smooth handover. Our proposal does not affect the scaling quality of HMIP. By reducing the effective global handover latency we reduce the number of packet losses.

Reduced handover latencies and packet losses would enable better quality of service. Our proposal not only reduces global handover latencies but is fully compatible with hierarchical Mobile IP and Mobile IPv6.

In future work, we plan to implement our model and observe behavior in complex multi connection scenarios. We also plan to incorporate session initiation protocol (SIP), for QoS handling, in our model.

References

[1-12]

[1] *Requirements on Handover for 3G.* 1999. http://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_02/Docs/Tpdfs/R4-99068.pdf

[2] Axel Neumann, A.F., *Hierarchical Mobile IP Implementation Experiences*. 2002, Siemens IPv6 Workshop: Berlin.

[3] Castelluccia, C., *A Heirarchical Mobile IPv6 Proposal.* INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE, 1998. http://www.inrialpes.fr/planete/people/bellier/hmip.pdf

[4] Glen, R., *RADVISION Announces 3G-324M Protocol Toolkit for 3G Wireless Multimedia Application Development*, in *Wireless Developer Network*. 2002. http://www.wirelessdevnet.com/news/2002/275/news6.html

[5] Handley.M, S.H., Schooler.E, Rosenberg.J, *SIP: Session Initiation Protocol.* 1999. http://www.ietf.org/rfc/rfc2543.txt?number=2543

[6] Hesham Soliman, C.C., Karim El-Malki, Ludovic Bellier, *Hierarchical Mobile IPv6 mobility management (HMIPv6).* IETF, 2002. http://www.ietf.org/internet-drafts/draft-ietf-mobileip-hmipv6-07.txt

[7] Hsieh Robert, S.A., Soliman Hesham, El-Malki Karim. *Performance analysis on Hierarchical Mobile IPv6 with Fast-handoff over End-to-End TCP*. in *GLOBECOM*. 2002. Taipei, Taiwan.
[8] Hsieh Robert, Z.Z., Seneviratne Aruna. *S-MIP: A Seamless Handoff Architecture for Mobile IP*. in *INFOCOM*. 2003. USA.
[9] J Davidson, C.P., J Arkko, *Mobility Support in IPv6*. IETF, 2003.
[10] Perkins, C., *Mobile IP Design Principles and Practices*. 1998: Addison Welsey.
[11] Swain, K., *Global mobile forecasts to 2010*. 3rd ed ed. Baskerville strategic research. 2002, London: Baskerville 2002. xiv, 468.
[12] Xavier Pérez Costa, R.S., Hannes Hartenstein, Marco Liebsch, *A MIPv6, FMIPv6 and HMIPv6 handover latency study: analytical approach.*

Freedom to Choose along with Freedom of Mobility

Saurabh Bansal, Aboubaker Lasebae, Richard Comley
School of Computing Science, Middlesex University, London UK
{s.bansal, a.lasebae, r.comley}@mdx.ac.uk

Abstract

Fourth generation mobile communication networks are characterized by heterogeneous access networks and IP based transport technologies. Realization of handovers to the best network section while considering QoS and Authentication, Authorization, Accounting and Charging (AAAC) calls not only for seamless handover protocols, but also intelligent handover decision strategies. To realise the goal of providing the user with the freedom and flexibility to choose services, there have to be means for service advertisements, service selection and seamless handovers across different types of access networks as well a service selection algorithm, a handover strategy to accomplish seamless cross-network handovers supporting WLAN and HMIP cells, as well as all the necessary modifications to the WLAN binding table to facilitate handovers with HMIP cells. Our presentation of the above solutions and strategies is designed to be fully compatible with Hierarchical Mobile IP and Mobile IPv6.

1. Introduction

Fourth Generation is not yet a global standard but more of a concept that will enable devices to switch between different types of networks [1]. Though there are many conflicting views about 4G, e.g. Japanese view of network based on orthogonal frequency division multiplexing (OFDM) offering data rates up to 100Mbits/sec, it is widely believed that IP will be the means to integrate access networks of all technologies [5]. The switch from traditional circuit switched technology to packet switched would provide users great flexibility but it will complicate network and mobile terminal design [12].

Mobile IP was designed to provide mobile devices freedom to move from one network to another while maintaining their permanent IP address. Mobile IP is a modification of IP that allows nodes to continue to receive datagrams (i.e. UDP packets) no matter where they happen to be attached to the Internet. It involves some additional control messages that allow the IP nodes involved to manage their IP routing tables reliably [10]. Mobile IP deploys a Home Agent (HA) that intercepts packets bound for the Mobile Node (MN) and tunnels them to its current point of attachment to the network via a foreign agent. Tunnelling (or triangle routing) ensures that MN continues to receive it's packets but MN has to ensure that HA knows about its current point of attachment. MN notifies its HA when changing in between networks but this indirection requires a registration process and address resolution procedure. The indirection procedure has been shown to result in long handover latencies, which often leads to packet loss and performance degradation [7].

Many variants of mobile IP have been developed to counter the problems faced by mobile IP and to integrate access networks of different technologies. Of the many variants, Hierarchical Mobile IP (HMIP) provides a scheme for performing registrations locally in the foreign network, thereby reducing the number of signalling messages forwarded to the home network as well as lowering the signalling latency that occurs when a mobile node moves from one Base Station (BS) to another. HMIP introduces a conceptual entity, Mobility Anchor Point (MAP), to separate local mobility from global mobility. MAP acts as a MN's Home Agent (HA) in the foreign network. MAP intercepts all packets destined for MN and tunnels them to its on-link CoA (Link Care of Address). In HMIP, MN attaches itself to two Care of Addresses (CoA), a regional (or site) address and an on-link (or point of attachment) address. Figure 1 shows a typical HMIP network architecture.

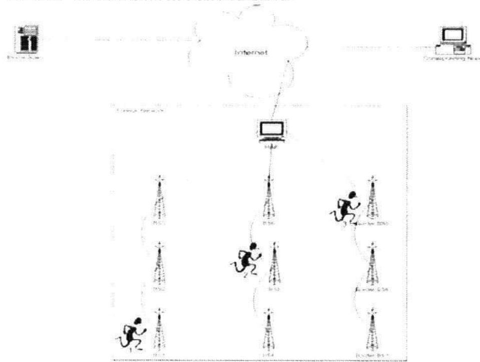


Figure 1 Hierarchical Mobile IP network architecture

Wireless Local Area Networks (WLAN) could be a cost effective complement to 3G systems in urban blanket deployments for multimedia applications [10] and provide users high speed access to the Internet in hotspot environments. WLAN's are short range broadband radio systems. Long-term applications range from networks on trains to ad-hoc networks. IEEE 802.11 is a series of WLAN standards ranging from a to m, mainly differing at the physical layer. HIPERLAN/2 (or HiSWANa) is a solution for the 5GHz band offering a number of air interfaces up to

54Mbits/s, mobility and QoS for applications such as multimedia, VoIP and real-time video.

All the different wireless standards are expected to inter-work to provide users with a wide range of services at a higher grade in the fourth generation systems. Real-time services with quality comparable to traditional cellular networks should be generally accessible regardless of the technology and the access network and uninterrupted during a handover. Such a heterogeneous design poses challenges in the handover design, service advertisement and call setup. Integration of seamless handovers with QoS and AAAC coupled with the possibility of a large number of networks being involved make it even more complicated. This paper aims to provide a handover decision strategy and network selection algorithm to facilitate seamless handovers.

2. Handover Classification

Mobile terminals in 2G and 3G networks send measurement reports to the network to enable the network to make handover decisions for the mobile node. Mobile assisted handovers work well in non-heterogeneous networks within the administrative domain, whereas in heterogeneous networks they have some disadvantages. Handovers within a domain don't require any security considerations whilst if the handover is cross-domain the node would have to be authenticated in the new domain. Service profile of mobile nodes would vary from domain to domain whereas it stays the same within a single domain. Accounting and authorization information would have to be refreshed every time the node is handed over to a new administrative domain, which adds to the handover latency. Network controlled handovers would prevent a user from exercising their right of choice by looking elsewhere other than the subscribed to network. To enable mobile nodes to make informed decisions about network associations, in this paper we only considered mobile-controlled handovers. However, mobile controlled handovers require networks to disclose more capability information. Also, networks would need to make provisions for service advertisements to provide nodes with a full list of services available.

Handovers in 4G would not only be concerned to maintain connection but also to provide better services and to meet individual requirements. The mobile terminal should be able to make the handover decision based on pre-selected user preferences with the option of allowing the user to manually change the settings for a specific channel. Handovers can be categorized as imperative and alternative. Imperative handovers are a priority because their selection and execution has to be done quickly in order to maintain connection of on-going channels. Alternative handovers take place to provide users with better performance or to meet a particular preference. Alternative handovers can tolerate longer handover latency compared with imperative handovers.

To enable the mobile node to exercise its right of freedom it needs to know the list of services available in the region, not just those from its service provider.

2.1. Service Advertisement

In 2G and 3G networks, handover measurements are controlled by the network. In UMTS systems, prior to handover decisions, neighbouring cell parameters, such as frequencies and scrambling codes are sent to the mobile node by the UTRAN/RNC. In WLAN 802.11b wireless nodes have to scan all the 13 channels to find the corresponding access point, which can take up to 400ms [12]. Large scale scanning adds to the L2 handovers. Though scanning a wide band of frequencies adds to the handover latency, without doing so the mobile node cannot make a pre-selection handover decision from all the available options. Mobile nodes could have a service contract with some network operators but there could be other operators available who may offer services as a pay-per-use basis. This however, raises the issue of billing and security. Billing and security are beyond the scope of this paper. The ways in which channel specification could be delivered to the mobile node without the overhead of scanning –

1. Neighbouring cell specification could be delivered by the MN's current point of attachment upon authentication. MN's current AP could send it a list of neighbouring cells [12].
2. All the service advertisements for WLAN and mobile telephony could be made on a single frequency band, so the mobile node only has to scan for one frequency band to find out all the available options. This raises the issue of regulating the frequency band. This would enable mobile nodes to connect to competitor networks in areas where its subscribed network does not provide any service. The feasibility of the method is beyond the scope of this paper.
3. A central database enabling the network operators to maintain their zonal pricing policies. Network operators could allocate a small band of frequency from their allocated band to enable the mobile nodes to query the database during call initiation and handovers. This would enable network operators to have a zonal pricing policy and reduce the amount of cell specification data transferred to the mobile node. Querying the database would however increase the call setup time. Users should be able to use the database whenever they desire but could also wish to ignore it. Feasibility of the database method is beyond the scope of this paper.

2.2. Service selection

The information required by the mobile node could either be collected from the network itself once the mobile node knows who's around or could be made a part of the service advertisement. To measure the capability of candidate networks mobile nodes could employ CARD protocol [9].

2.2.1. Service characteristics. The mobile node could be informed about certain characteristics of the service through the means of advertisements and the node could acquire the rest of the information through the network itself. The advertisement would contain information such as the provider, the frequency at which they operate, services offered, price band, QoS class offered and support for seamless handovers. Then the mobile node could measure the received signal strength, signal interference, bit error rate and network coverage. The node could then register with the network and acquire capability information of the network.

Table 1. Service advertisement information[1]

Provider	Frequency	Service	Price	QoS	Seamless	Carrier
A		Voice	Low	BE	Yes	0
B		Voice	Medium	Pri Que	Yes	1
C		Data	Free	BE	Yes	1
A		Data	Medium	BE	Yes	0

2.2.2. Registration. HMIP allows MN's to hold multiple registrations but different billing policies might limit the registration capability of the MN. MN might not be able to register itself in a network because of a credit issue. To aid pre-handover selection and lower handoff latency it would help if the MN is able to register with all the possible options. Until the MN initiates a service, MN's registration would be in a soft state. Holding soft registration would enable mobiles to acquire candidate network capability information. Registration could be renewed or dropped after the validity period expires.

2.2.3. Algorithm for selecting channel. A possible algorithm for selecting a channel once the candidate networks capability information has been gathered. Suppose a user wants to use Application 'A' using service 'S'.

Users pre-defined values –
QoS_expected = qe
Cost_expected = ce
Security_expected = se
} User expectations

Network properties -
QoS_offered = qo
Cost_offered = co
Security_offered = so
} Network capability

User importance -
User_qos_weighting = qw
User_cost_weighting = cw
User_security_weighting = sw
} qw+cw +sw = 1

Figure 2. Handover Decision Engine[2]

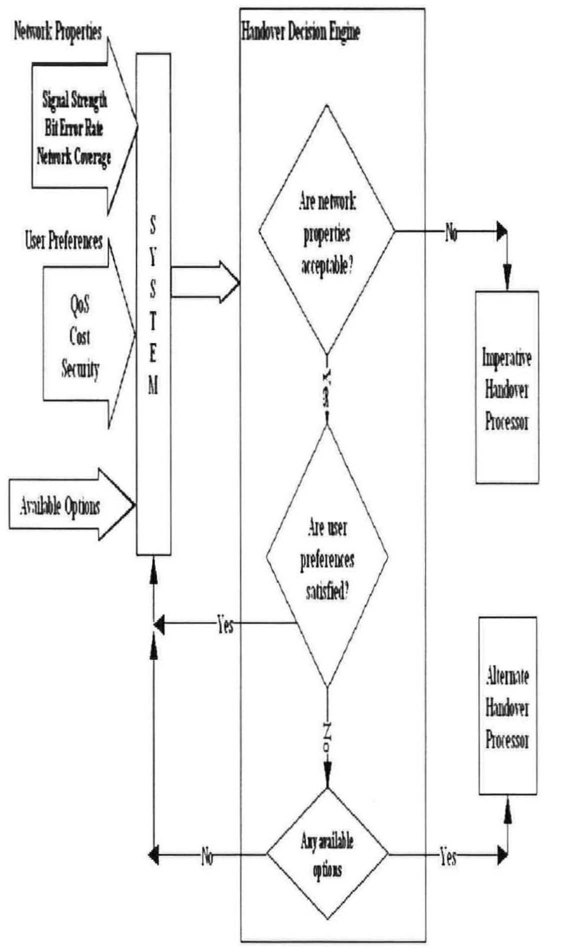


Figure 3. Imperative Handover Processor

[1] Refer to Appendix B for enlarged table

[2] Refer to Appendix A for enlarged figure

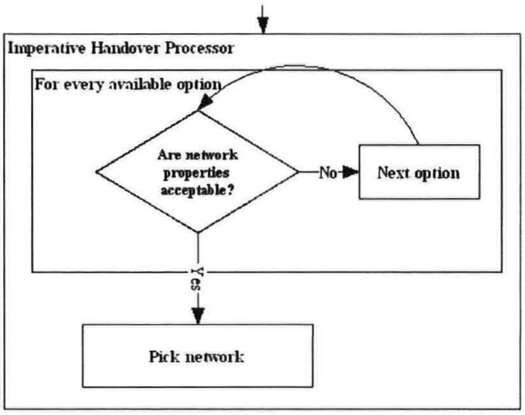
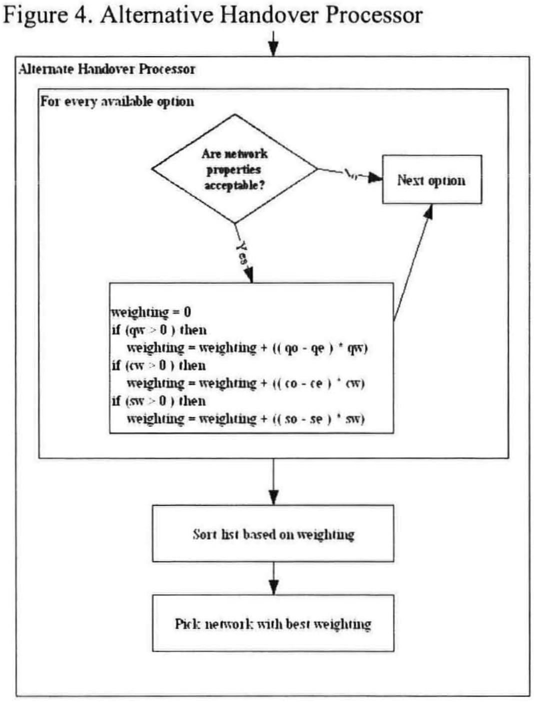


Figure 4. Alternative Handover Processor



3. Handover

Providing users with freedom to choose has to be backed with the freedom of mobility. If the users have to restrict their movements or reinitiate their channels upon reattachment, then it would be pointless to provide the freedom of selection. In some scenarios however, freedom of choice has to be sacrificed for the sake of maintaining a connection. For instance, a user is downloading the latest HP drivers of his printers while sitting in a coffee shop. After finishing his coffee he leaves the shop without realising the download still has not completed. Due to the short range of the WLAN hotspots the user is out of reach almost immediately while the download is still in progress. Handover now has to be performed very quickly in order to maintain connection with the web server. Since many web servers don't support route optimisation the task becomes enormous.

3.1. WLAN binding information

Hotspot binding tables would have to be changed to provide seamless handovers between WLAN and HMIP networks. Upon registration in a hotspot the MN would be provided with virtual and private bindings, if the hotspot supports an extended service set. To provide seamless handovers, hotspots would have to store a mobile node's home address. Since hotspots operate in very small areas, much smaller than a cell, the MN could use its HMIP VCoA as the HA for the hotspot.

The binding table looks similar to the HMIP binding table. RHA (Real Home Agent) is the real home agent of the terminal whereas HA is the first alternative choice (based on service selection). By using HA as the first alternative, MN would reduce the handover latency and would be very useful in imperative handover situations.

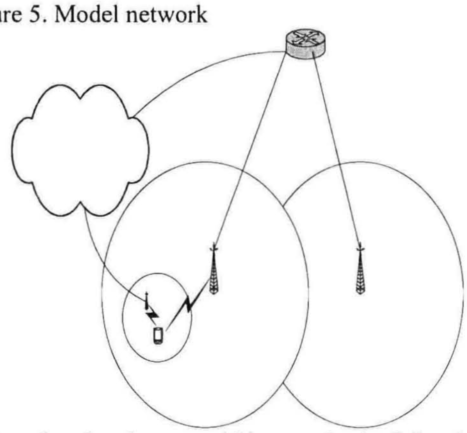
Table 2. WLAN binding table

MAC	PCoA	VCoA	Credit	Seamless	RHA	HA
-	-	-	-	-		-

3.2. WLAN – HMIP handover

Scenario – A MN moves into the hotspot and wants to enjoy high speed wireless access provided by the hotspot whilst still being connected to the HMIP network for voice and other services. MN initiated a data channel but moves out of the hotspot coverage. If the hotspot does not support cross-network seamless handover, MN's call would be dropped or the MN would have to move back within the hotspot coverage area, hence restricting the movements of the MN.

Figure 5. Model network



Seamless handover could happen in the following ways –

1. CN supports route optimization – If the CN supports route optimization, MN could send a binding update so that the CN could start transmitting the packets to its new link attachment. Packets already in transit would be rerouted by the hotspot access router to the MN's home address.

2. CN does not support route optimization – If the CN does not support route optimization then all the future packets would be rerouted to the MN's HA where they would be tunnelled to MN's current link attachment.

3.3. HMIP-WLAN handover

These handovers would only be useful if the CN supports route optimisation. Since many web servers don't support route optimisation MN's packets would have to be re-routed by the MAP to the hotspot whereby the packets would be travelling through the global Internet domain at the best effort service.

Packet latency HMIP to WLAN

$$= HMIP_{(Wired\ +wireless\ )delay} + BE + Hotspot_{(wired\ +wireless\ )delay}$$

Packet Latency HMIP $= HMIP_{(Wired+wireless)delay}$

Hence, this scenario is only useful if the CN supports route optimization or both the cells are a part of the same network.

3.4. Cross-network handovers

Inter-network handovers are not useful for short duration channels, as the handover would be subject to accounting and security. However, to provide the user with freedom of choice, seamless inter-network handovers are important. Handovers between two networks could be made seamless by introducing coverage area overlaps [11]; this would however only work if the handover is decided by movement. For instance, a user is on his way to attend a very important meeting but is not sure about some minor details. He calls up his boss by dialling his number and selecting the radio button 'Very Urgent'. The mobile terminal initiates the call with the user's carrier network but whilst the user is on the call the terminal starts looking for a network with lower load and better QoS. The terminal transfers the user's call onto the new network seamlessly. Security and billing issues in the above scenario are beyond the scope of this paper.

Call setup and Handover procedure:

1. Call initiated through carrier network.
2. List of candidate networks acquired.
3. User preferences algorithm executed.
4. Network selected.
5. Soft registration acquired.
6. Authenticated on the new network.
7. User profile downloaded by the new network from Home Agent.
8. New network accepts user call handover.
9. Handover processed.

4. Conclusion

This paper outlines a handover strategy for 4G wireless networks. Handovers are classified as imperative and alternative, and the decision criteria can be classified as static or dynamic. Though it would ideal to have a central database for service advertisements, a single frequency band cannot be ruled out altogether because of the flexibility it provides. Cross-network handover latency could be reduced by soft registrations. Soft registrations enable terminals to acquire candidate network capability information as well. The algorithm for selecting candidate networks would enable terminals to select networks based on user preferences. In the end, a new binding update table for WLAN's is suggested along with the procedure to establish cross-network handovers and WLAN-HMIP handovers.

5. References

[1]"4G - Beyond 2.5G and 3G Wireless Networks," vol. 2004.www.mobileinfo.com/3G/4GVision&Technologies.com

[2] Axel Neumann, Andreas Festag, "Hierarchical Mobile IP Implementation Experiences." Berlin: Siemens IPv6 Workshop, 2002

[3]C. Castelluccia, "A Hierarchical Mobile IPv6 Proposal," *INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE*, vol. 226, 1998

[4]Clapton.A, *Future Mobile Networks: 3G and beyond*: Institution of Electrical Engineers, 2001.

[5]D Wisely, E Matjina, "Evolving systems beyond 3G - the IST BRAIN and MIND projects," *BT Technology Journal*, vol. 21, pp. 102-120, 2003

[6]Feyrer, "Introduction to IPv6," 2001.http://www.onlamp.com/lpt/a/onlamp/2001/05/24/ipv6_tutorial.html

[7]Hesham Soliman, Claude Castelluccia, Karim El-Malki, Ludovic Bellier, "Hierarchical Mobile IPv6 mobility management (HMIPv6)," *IETF*, 2002.http://www.ietf.org/internet-drafts/draft-ietf-mobileip-hmipv6-07.txt

[8]Hsieh Robert, Seneviratne Aruna, Soliman Hesham, El-Malki Karim, "Performance analysis on Hierarchical Mobile IPv6 with Fast-handoff over End-to-End TCP," presented at GLOBECOM, Taipei, Taiwan, 2002

[9]P. Kankaanpaa, "Candidate Access Router Discovery," 2003

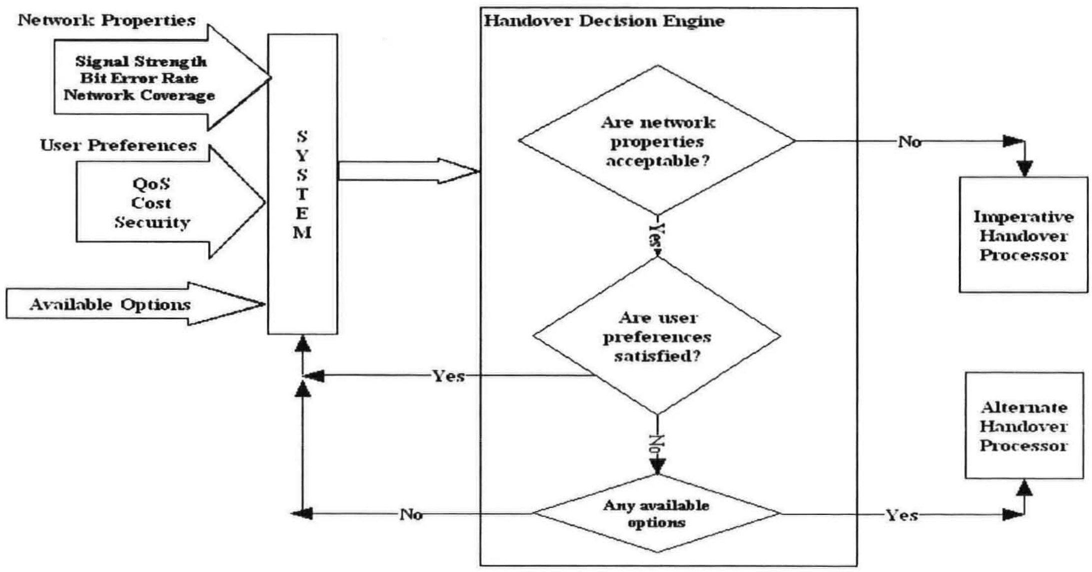
[10] L Burness, David Higgins, A Sago, P Thorpe, "Wireless LANs - present and future," *BT Technology Journal*, vol. 21, pp. 32-47, 2003

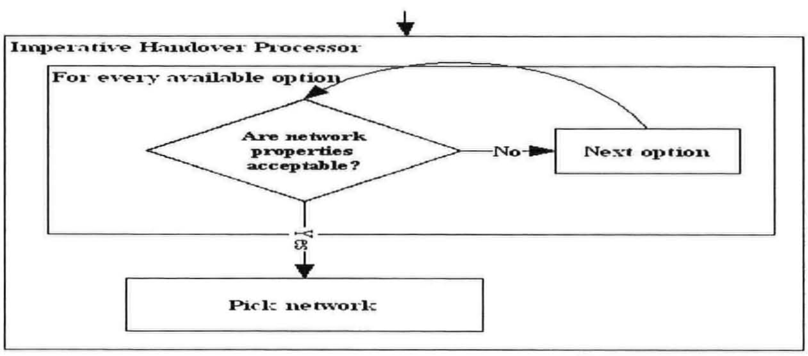
[11] Saurabh Bansal, Aboubaker Lasebae, Richard Comley,

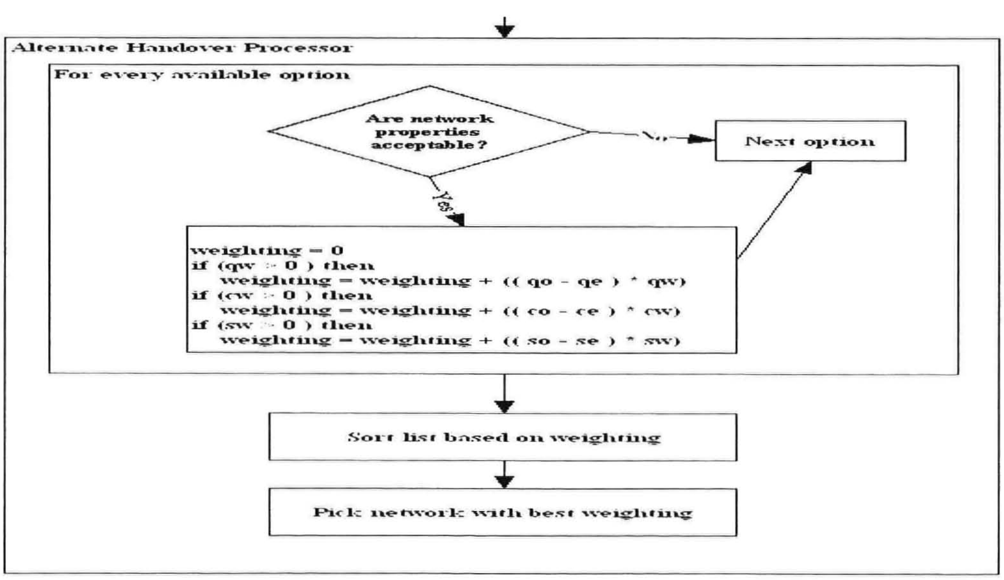
"Global Handover Latency Minimisation in HMIPv6," *WSEAS Transactions on circuits and signals*, vol. 2, 2003

[12] Wenhui Zhang, Juergen Jaehnert, Klaus Dolzer, "Design and Evaluation of a Handover Decision Strategy for 4th Generation Wireless Networks,"

Appendix A – Service Selection Algorithm







Appendix B – Service Advertisement Information

Provider	Frequency	Service	Price	QoS	Seamless	Carrier
A		Voice	Low	BE (Best Effort)	Yes	0
B		Voice	Medium	PriQue	Yes	1
C		Data	Free	BE	Yes	1
A		Data	Medium	BE	Yes	0

Service advertisements for future mobile communication networks

Dr. Aboubaker Lasebae
a.lasebae@mdx.ac.uk

Mr. Saurabh Bansal
s.bansal@mdx.ac.uk

Prof. Richard Comley
r.comley@mdx.ac.uk

ABSTRACT

Fourth generation mobile communication networks are characterized by heterogeneous access networks and IP based transport technologies [1]. Realization of handovers to the best network section while considering QoS and Authentication, Authorization, Accounting and Charging (AAAC) calls not only for seamless handover protocols, but also intelligent service advertisement mechanisms. Support for multi-operator and multi-access type scenarios would introduce further handover and call setup delays due to the overhead of scanning [7]. To minimise the scanning overhead yet providing users with the freedom of choice calls for efficient service advertisement mechanisms [8]. Service advertisements could be delivered to the terminal on a reserved frequency band or via a centralized database. Though the frequency band mechanism is faster and more flexible than the database access method, no knowledge of the operating band would make it unusable. The two advertisement mechanisms complement each other to provide a perfect advertisement delivery mechanism.

1. INTRODUCTION

Fourth Generation is a collection of concepts which would enable seamless access across different types of networks [2]. Different access technologies give users greater choice in selecting services that can reflect differences in Quality of Service (QoS) support, business models and service providers. Wireless Local Area Networks (WLAN) could be a cost effective complement to 3G systems in urban blanket deployments for multimedia applications and provide users high speed access to the internet in hotspot environments [3]. All the different wireless standards are expected to inter-work to provide users with a wide range of services at a higher grade in the fourth generation systems. Real-time services with good quality should be generally accessible regardless of the technology and the access network, uninterrupted during a handover. Such a heterogeneous design poses challenges in the handover design, service advertisement and call setup. Integration of seamless handovers with QoS and AAAC (Authentication, Authorization, Accounting and Charging) coupled with the possibility of multi-operator scenarios make it even more complicated.

Support for multi-operator scenarios calls for efficient service advertisement mechanisms. Since the future mobility terminals would support more than one access mechanism; operators would vary in terms of size. For instance, there could be a coffee shop offering high speed data access in it's hotspot to a country-wide operator offering voice, video, data and other services. To enable users to be notified of services in the surrounding areas, there has to be an efficient service advertisement mechanism [8].

In 2G and 3G networks, handover measurements are controlled by the network. In the UMTS system, prior to handover decisions, neighbouring cell parameters, such as frequencies and scrambling codes are sent to the mobile node by the UTRAN/RNC [8]. In WLAN 802.11b, wireless nodes have to scan all the 13 channels to find the corresponding access point, which can take up to 400ms[7]. Large scale scanning adds to the layer 2 handovers. Though scanning a wide band of frequencies adds to the handover latency, without doing so the mobile node cannot make a pre-selection handover decision from all the available options. Mobile nodes could have a service contract with some network operators but there could be other operators available who may offer services on a pay-per-use basis.

To support multi-operator service scenarios terminals would have to do large scale scanning to find out the available candidate networks. However, large scale scanning would add to the call setup and handover latency, also it would be a waste of terminal battery power.

2. HANDOVER CLASSIFICATION

Mobile terminals in 2G and 3G networks send measurement reports to the network to enable the network to make handover decisions for the mobile node. Mobile assisted handovers work well in non-heterogeneous networks [8] within the administrative domain, whereas in heterogeneous networks they have some disadvantages. Handovers within a domain don't require any security considerations whilst if the handover is cross-domain the node would have to be authenticated in the new domain. The service profile of mobile nodes would vary from domain to domain whereas it stays the same within a single domain. Accounting and authorization information would have to be refreshed every time the node is handed over to a new administrative domain, which adds to the handover latency. Network controlled handovers would prevent a user from exercising their right of choice by looking elsewhere other than to the subscribed network. However, mobile controlled handovers require networks to disclose more capability information. Also, networks would need to make provision for service advertisements to provide nodes with a full list of services available.

To enable the mobile node to exercise its right of freedom it requires a list of service providers available in the region, besides it's main service provider.

3. SERVICE ADVERTISEMENT DELIVERY

We have outlined three means for delivering service advertisements. The three ways in which channel specification could be delivered to the mobile node without the overhead of scanning are:

- Neighbouring cell specification [7] could be delivered by the MN's current point of attachment upon authentication. MN's current AP could send it a list of neighbouring cells. After authentication, the network could provide the node with the list of candidate networks available in the surrounding area. MN's current network might not be willing to provide a list of candidate networks as it is a competition issue. Also, if the MN's subscribed network does not offer services in a particular area, then MN would be unable to find out about the candidate networks. UMTS networks could provide this information as a value-added location based service. Neighbouring cell specifications would not be pursued due to the lack of transparency.
- All the service advertisements for WLAN and mobile telephony could be made on a single **frequency band**, so the mobile node only has to scan for one frequency band to find out all the available options. This raises the issue of regulating the frequency band, though this would enable mobile nodes to connect to competitor networks in areas where its subscribed network does not provide any service. A single frequency band would complicate terminal design. It might be difficult for the terminal to find the frequency band when travelling abroad. However, even with these limitations it provides the most operator independent platform.
- A **central database** enabling the network operators to maintain their zone pricing policies. Network operators could allocate a small band of frequency from their allocated band to enable the mobile nodes to query the database during call initiation and handovers. This would enable network operators to have a zone pricing policy and reduce the amount of cell specification data transferred to the mobile nodes. Querying the database would however increase the call setup time. Users should be able to use the database whenever they require but could also wish to ignore it. Database access would increase the handover latency. When travelling abroad it is more practical for terminals to use the database access method because they would not know the single frequency band.

4. SERVICE CHARACTERISTICS

The service advertisements would deliver the following service characteristics to the terminal:

- Provider – Service Provider ID
- Frequency – Service operation frequency
- Bandwidth – Available bandwidth on the channel
- Service type – Type of service (e.g. voice)
- Price – Cost of using the service
- QoS Class – Type of quality of service offered
- Security class – Type of security offered
- Support for multi-operator handovers

Upon collecting the service characteristics, the terminal can collect the more dynamic service characteristics –

- Received signal strength
- Bit error rate
- Network coverage

The mobile node could store the service characteristic information in a table. The stored information could be utilized by the service selection algorithm to make handover and call setup decisions. The dynamic properties of the service would require a periodical refresh.

5. REGISTRATION

HMIP allows MN's to hold multiple registrations but different billing policies might limit the registration capability of the MN. MN might not be able to register itself in a network because of a credit issue. To aid pre-handover selection and lower handoff latency it would help if the MN is able to register with all the possible options. Until the MN initiates a service, MN's registration would be in a soft state. Holding soft registration would enable mobiles to acquire candidate network capability information. Candidate network capability information could be acquired using the CARD protocol. Soft registrations would also be useful in acquiring dynamic service characteristics. Registration could be renewed or dropped after the validity period expires.

6. ADVERTISEMENT MECHANISMS COMPARISON

- Transparency – Both the mechanisms are transparent. The frequency band option allows the service providers to advertise their services themselves and the database access method allows operators access to edit their records.

- Regulation – The frequency band method raises an issue of regulating the frequency band whereas the database would have to be managed by someone allowing relevant access to users and service providers.
- Cost of Ownership – Frequency band option would increase the cost for service providers who would have to and maintain install their own advertisement equipment. However, the cost of ownership could be reduced by sharing of equipment. The database access mechanism adds the cost of editing and maintaining the records by service providers.
- Delays introduced – Frequency band operation introduces the delay of scanning the whole band of frequency and then downloading the relevant advertisements. The database call setup time and download of records would introduce delays in the database access mechanisms.
- Complement - The two methods complement each other and the terminal can decide which method to use when collecting service advertisement information or when one is not available.

7. ANALYSIS

7.1 Frequency Band Mechanism

Scanning the single frequency band to listen to advertisements would be the delay introduced by the frequency band mechanism. The frequency band would work like a radio reception service where the terminal would tune into different radio frequencies for service advertisements.

Frequency band Handover latency =

$$(\sum_{Carrier=0}^{N} Downlink_{Delay}) + (SelectionAlgorithm_{Delay}) + (Registration_{Delay}) + (Handover_{Latency}) \text{ ---- } (1)$$

Downlink delay represents the time taken by the terminal to tune into a frequency and download the advertisement. Selection algorithm delay represents the delay introduced by the service selection algorithm [9]. Upon selecting the service the latency introduced the registering in the new network is represented by the Registration delay.

Test Scenario – Due to the limitations imposed by our equipment the maximum number of service providers was limited to 13. A 3Com 802.11b access point and a LAN PC card were used for the tests. A laptop with an 802.11b card was used to measure the time taken to scan the 13 channels.

7.2 Database Mechanism

Database access Handover Latency =

$$(DBCall_{Setup}) + \text{[Call setup delay]}$$
$$DBExecuteQuerry + \text{[Querry execution delay]}$$
$$(\sum_{Carrier=0}^{N} DBRecord_{Download}) +$$
$$(SelectionAlgorithm_{Delay}) +$$
$$(Registration_{Delay}) +$$
$$(Handover_{Latency}) \text{ ---- } (2)$$

Test scenario

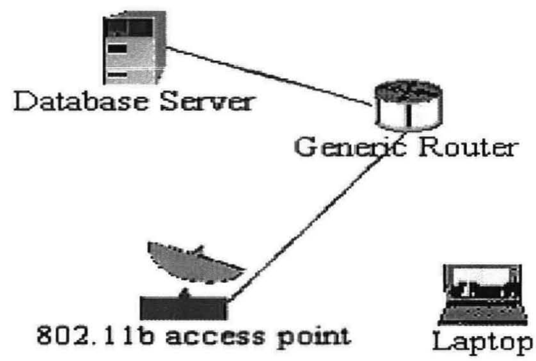


Figure 1 -Test Network

A database populated with 1000 records divided in 50 regions was used alongside an 802.11B access point, laptop and Linux machine acting as a router. The database was searched 30 times per region with different numbers of service providers randomly.

8. Results

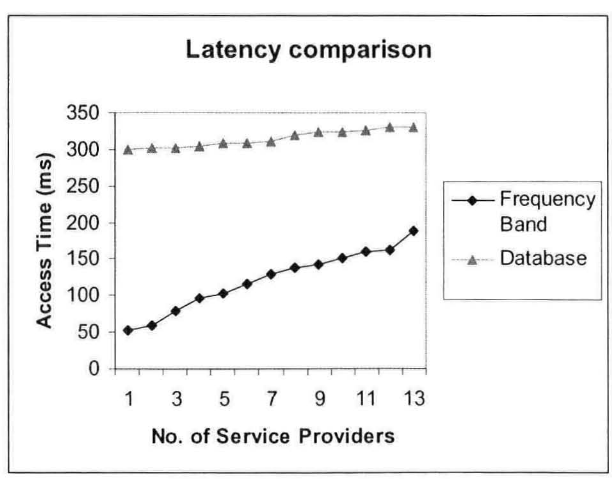


Figure 2 - Latency comparison

Results reflect that frequency hopping is much quicker than database access method. Access time does not increase significantly in the database access method but they do increase significantly with the frequency access method. Access time in database access method does not increase significantly because call setup and query execution only has to be done once and the time difference is only relative to the number of records to be downloaded which can be grouped together in a single packet. Whereas in the Frequency access method terminal access time is relative to the number of channels with nothing common in between them.

9. Conclusion

Service advertisement mechanism would form the basis of multi-operator multiple access type networks. Frequency access method and database access method working independently or in tandem provide a perfect balance for delivering service advertisements to the terminals. Frequency access method is quicker than database access method but if the operating frequency is not know it would be unusable.

10. REFERENCES

[1] "4G - Beyond 2.5G and 3G Wireless Networks," vol.2004.www.mobileinfo.com/3G/4GVision&Technologies.com

[2] Clapton.A, *Future Mobile Networks: 3G and beyond*: Institution of Electrical Engineers, 2001.

[3] D Wisely, E Matjina, "Evolving systems beyond 3G - the IST BRAIN and MIND projects," *BT Technology Journal*, vol. 21, pp. 102-120, 2003

[4] P. Kankaanpaa, "Candidate Access Router Discovery," 2003

[5] L Burness, David Higgins, A Sago, P Thorpe, "Wireless LANs - present and future," *BT Technology Journal*, vol. 21, pp. 32-47, 2003

[6] Saurabh Bansal, Aboubaker Lasebae, Richard Comley, "Global Handover Latency Minimisation in HMIPv6," *WSEAS Transactions on circuits and signals*, vol. 2, 2003

[7] Wenhui Zhang, Juergen Jaehnert, Klaus Dolzer, "Design and Evaluation of a Handover Decision Strategy for 4th Generation Wireless Networks,"

[8] Saurabh Bansal, Aboubaker Lasebae, Richard Comley, "Freedom to choose along with the Freedom of Mobility", ICTTA'04 IEEE Section France Damascus, 2004

Service Advertisements for Fourth Generation Mobile Networks

Aboubaker Lasebae, Saurabh Bansal, Richard Comley
School of Computing Science, Middlesex University, London UK
{s.bansal, a.lasebae, r.comley}@mdx.ac.uk

Fourth generation mobile communication networks are characterized by heterogeneous access networks and IP based transport technologies [1]. Heterogeneous access across networks provides users with great flexibility in selecting services based on their requirements, but such heterogeneous access poses great challenges. Heterogeneous access would enable mobility users to control the handover process and decide which network they get handed over to. Users will get greater flexibility in exercising choice when making handover decisions. Future mobility networks would cater for various access mechanisms and operators would vary in terms of coverage and access technology. Support of multiple-operators requires efficient service advertisement mechanisms [8]. We propose a hierarchical operator independent ***Service discovery data repository*** which is not dependent on the ***Service discovery access mechanism***, with the capability of enabling location based, specialized, discrete QoS and cost based searches. The repository is the first of it's kind, enabling the users to explore services based on the requirements.

1. INTRODUCTION

Fourth Generation is a collection of concepts which would enable seamless access across different types of networks [2]. Different access technologies give users greater choice in selecting services that can reflect differences in Quality of Service (QoS) support, business models and service providers. Wireless Local Area Networks (WLAN) could be a cost effective complement to 3G systems in urban blanket deployments for multimedia applications and provide users high speed access to the internet in hotspot environments [3]. All the different wireless standards are expected to inter-work to provide users with a wide range of services at a higher grade in the fourth generation systems. Real-time services with good quality should be generally accessible regardless of the technology and the access network, uninterrupted during a handover. Such a heterogeneous design poses challenges in the handover design, service advertisement and call setup. Integration of seamless handovers with QoS and AAAC (Authentication, Authorization, Accounting and Charging) coupled with the possibility of multi-operator scenarios make it even more complicated.

Support for multi-operator scenarios calls for efficient service advertisement mechanisms. Since the future mobility terminals would support more than one access mechanism; operators would vary in terms of size. For instance, there could be a coffee shop offering high speed data access in it's hotspot to a country-wide operator offering voice, video, data and other services. To enable users to be notified of services in the surrounding areas, there has to be an efficient service advertisement mechanism [8].

Service advertisements can be categorized into:

1) **Service discovery access mechanism** – The access mechanism defines how the users can deploy the service discovery mechanism on their terminals. Wu's

common core network and Bansal's frequency band mechanism are the two existing access mechanisms. Gang Wu proposes an IPv6 Common Core Network (CCN) acting as a central hub for all mobile terminals. All access points are principally connected to this network [9]. The network would guarantee QoS guaranteed routing and seamless handovers among radio access networks. Whereas Bansal proposes broadcasting advertisements as part of router advertisements on a reserved frequency band [10]. The service discovery access mechanism is beyond the scope of the paper.

2) **Service discovery data repository** – The data repository outlines how the service discovery data is organized in a repository.

2. Requirements for service advertisement data repository

- **Support for Multiple Operators** - Multiple operators offer great flexibility to the users wishing to seek alternate networks for better services. Since, one of the design goals for future mobility networks is to provide heterogeneous access across networks, it is very important for the service advertisement mechanisms to support multiple operators.
- **Operator transparency and dependence** - The operator dependent service advertisement mechanisms would raise competition issues and the terminal would be limited in its ability to explore services. Operator independent service discovery mechanisms allow greater flexibility to the users in exercising their right of choice.
- **Reduce the overhead of scanning** - In 2G and 3G networks, handover measurements are controlled by the network. In UMTS system, prior to handover decisions, neighbouring cell parameters, such as frequencies and scrambling codes are sent to the mobile node by the UTRAN/RNC. In WLAN 802.11b wireless nodes have to scan all the 13 channels to find the corresponding access point, which can take up to 400ms [7]. Large scale scanning adds to the L2 handovers. Since, future wireless terminals would be able to support more than one interfaces, the device would have to scan for all the usable frequencies available to the device. Large scale scanning would add to delays in call setup and handovers. Large scale scanning would also drain resources of the terminals. Service advertisement mechanisms should not increase the scanning overhead of terminals.
- **Inter-network recourse discovery** - Since, the future mobility networks may offer various access mechanisms; the users should be able to explore services within the network [11].
- **Service discovery by coverage area** - The access technologies of mobile node would be limited in their reach, so the service advertisement mechanisms should enable the user based on the coverage area.
- **Provisions for specialized searches**

  The usage of mobile users can be broadly classified into –

  - **Standard services** – Standard services do not vary in terms of price, e.g. browsing the internet, etc.
  - **Specialized services** – Specialized services vary in terms of price, e.g. a local call would not be priced the same as an international call even though they both belong to the real-time voice telephony services. Specialized

services include any non-standard services which the operator may offer, e.g. live broadcast of Madona concert, etc.

Storing the pricing for every service by every operator would be a waste of the terminal's resources. So the advertisement mechanism should enable mobile node to establish the cost of using the services as and when required.

3. **Existing approaches**

Melazzi et. al. have outlined a procedure for resource discovery on mobile node's current network to outline various access mechanisms offered by the operator. Melazzi approach employs CARD (Candidate Access Router Discovery) protocol [4] to identify access routers available [11]. The approach discourages use of more than one interface at a time and can only be used for resource discovery within the network. Melazzi does not define any structure for the repositrory.

Neighbouring cell specification [7] could be delivered by the MN's current point of attachment upon authentication. MN's current AP could send it a list of neighboring cells. After authentication, the network could provide the node with the list of candidate networks available in the surrounding area. MN's current network might not be willing to provide a list of candidate networks as it is a competition issue. Also, if the MN's subscribed network does not offer services in a particular area, then MN would be unable to find out about the candidate networks. UMTS networks could provide this information as a value-added location based service. The approach does not allow query based specialized searches and fails to define a structure for the repository.

4. Our Approach – Data Repository

4.1 Hierarchical layout with zonal divisions

Nationwide wireless coverage area would be split up in zones based on the volume of wireless traffic. Zonal division of London may look like the London borough map in the figure below. Zonal division would enable us to focus on more localized services and differentiate between large and small operators. Traditionally any operator providing services citywide or nationwide was termed as a big operator even though that operator may have a very limited presence in a particular area. This paper would refer to operators providing services throughout the zone as large operator whereas operators having a small presence in the borough as small operators.

Figure 1 – London Borough Map Source http://www.ldan.org.uk/cms/view/findservice/ViewLondon.asp

Instead of a having a central database we propose a distributed hierarchical database,. The database would have several hierarchies based on country, city and zone divisions. The hierarchical distribution would enable terminals to locate repositories and search their local repository. It will also enable zonal repositories to look after a localized user. The zonal databases would cater for standardized services would work in conjunction with the operator run databases providing for their own standard and specialized services. Operator databases would enable the user explore other services offered by its operator. Distributed repositories would enable much less information being stored on the central database. Different levels of hierarchies would hold a varied base of information. The table below outlines data to be held on the databases –

Database	**Contents**
World	Links to all the continent databases
Europe	Links to all the European country databases
UK	Operational details of all the standardized services of UK wide operators
	Links to all the County databases
London	Operational details of all the standardized services of London wide operators
	Links to all the Zonal databases
City	Operational details of all the standardized services of Zone 'City' wide operators
	Standard and Specialized details of small operators
	Links to the databases of large operators
Operator B	Standard and specialized services of operator B

The varied levels of hierarchies would enable the user locating service providers in a foreign territory. The top level hierarchies and user' subscribed to providers can be pre-configured on the subscriber identity module (SIM) of the terminal.

4.1.2 Location based updates

The current 3G specification [12] enables the retrieval of location co-ordinates of the terminal by the network or the terminal. Terminal can extract the Global Positioning System (GPS) information from the Radio Network Controller (RNC) or the Servicing Mobile Location Controller (SMLC). The terminal can gather the information periodically or on demand. MN would be able to retrieve the zonal information and it's co-ordinates in the zone.

Location based information retrieval would enable the MN to greatly reduce the information overload by only retrieving what it requires. MN would specify its location and coverage area it seeks when retrieving information.

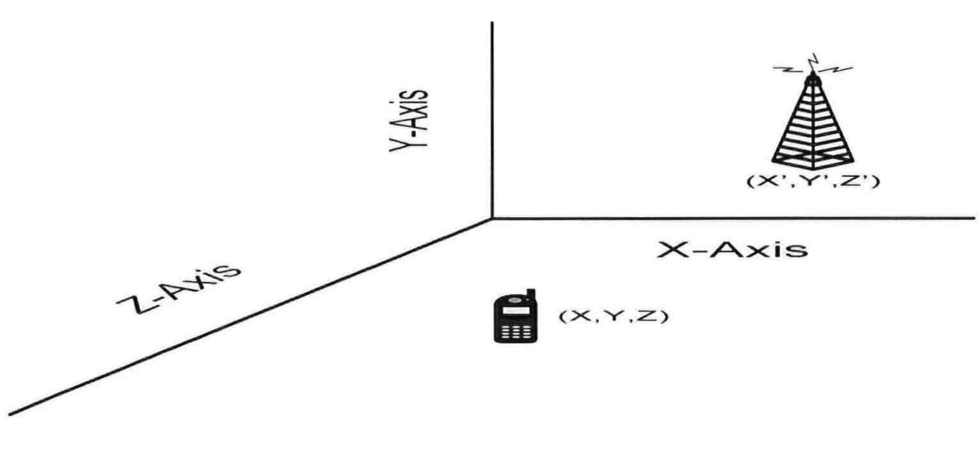
Location information though very valuable but puts extra strain on the database server by having to calculate terminal's position with reference to the operator's access point. To reduce this burden we would only allow location tracking for smaller operators rather than all the operators. Larger operators would have their location based services listed under smaller operators and would be available for location based search. E.g. Operator A provides 3G voice, video and data services but also provides high speed Wi-Fi internet service on Heathrow Airport Terminal 4. In this scenario operator A would have two entries in the central database but only Wi-Fi service would be available for location based search.

Terminal's location from the Access Point can be worked out using the triangulation algorithm [12].

Distance between terminal and access point =

$$\sqrt{(X'-X)^2 + (Y'-Y)^2 + (Z'-Z)^2}$$

where X, Y, Z are the co-ordinates of the terminal and X', Y', Z' are the co-ordinates of the terminal.



The distance between the access point and terminal enables the repository to analyse whether the access point is within the coverage area requested by the MN. To reduce the burden on the database distance calculation would only be performed when the operator fulfills all the other requirements of the user, which will be discussed in the subsequent sections.

4.1.3 QoS

QoS is the "collective effect of service" performance, which determines the degree of satisfaction of a user of a service [13]. QoS in itself is a collection of traffic engineering, policy management, QoS middleware, session management, network QoS mechanisms and signalling mechanisms.

QoS can be categorized into bandwidth, packet drop rate, jitter, latency and bite error rate. Different applications have different requirements for the above mentioned metrics. E.g. ftp applications are affected more by packet drop rate, bit error rate and bandwidth than latency and jitter. To cater for different types of applications we propose to document QOS in the repository as –

1) Minimum bandwidth offered by the network
2) Average packets/million dropped
3) Average latency – The average delay experienced by the users in the network
4) Average Jitter – The average jitter experienced by the people in the network.
5) Average BER – The average bit error rate experienced by the users.

The approach relies on the honesty of the operators to advertise true information to the users. The user could calculate various QoS metrics itself, but this approach has the following disadvatages –

1) Mobile terminals have limited resources and would end up spending considerable amounts on these calculations. Besides bandwidth all the other values are dynamic and therefore would have to be refreshed by the user at regular intervals.
2) Mobile IP's one of the big drawbacks is the amount of control data on the networks [6]. If all the users started maintaining their own performance metrics, the networks would have even more control information flowing through the network.
3) The terminal would have to maintain it for every operator available which could be several in certain areas.

To deal with the issue of honesty the user can maintain a feedback (or the past experience) of the operators it has used services from [10]. The user can then discard operators with adverse rating. Maintaining a feedback would be much more convenient and manageable than maintaining local QoS performances.

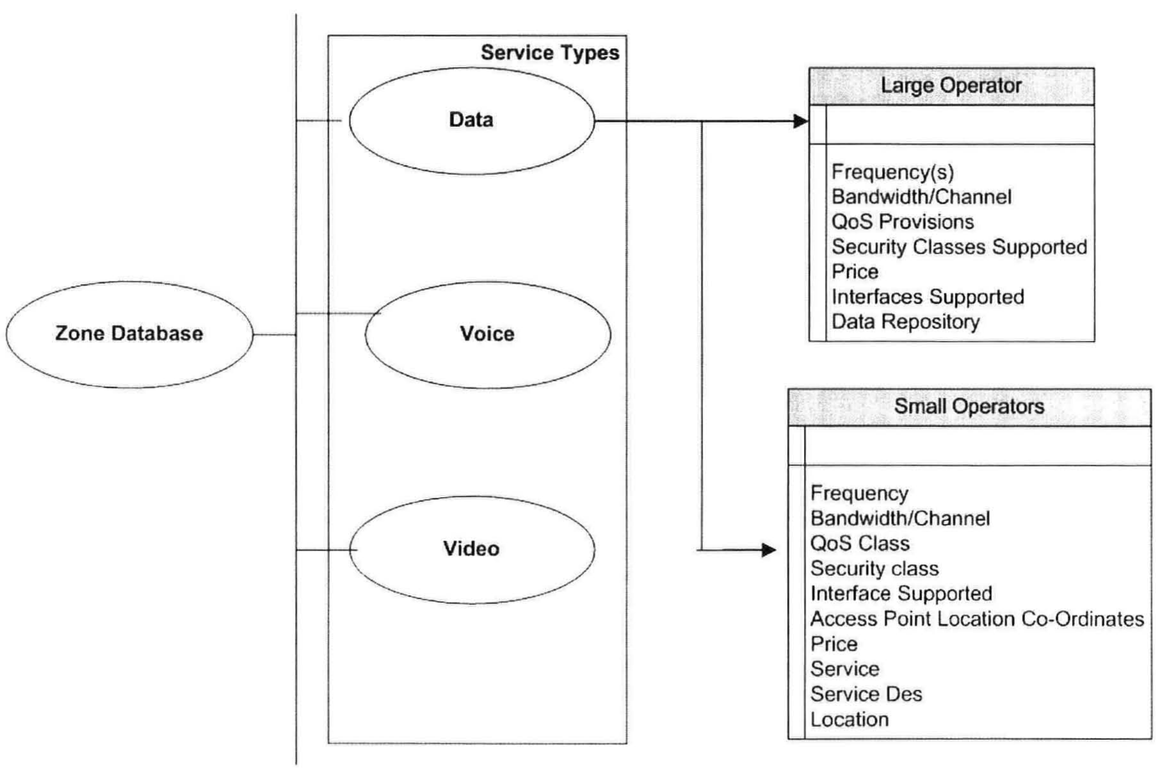
4.1.4 Cost

Cost of using a service is often ambiguous because of the varying units. Some operators charge per min whereas some charge per byte downloaded for browsing services. To enable users to compare service provider's charges, we propose to break this information as-

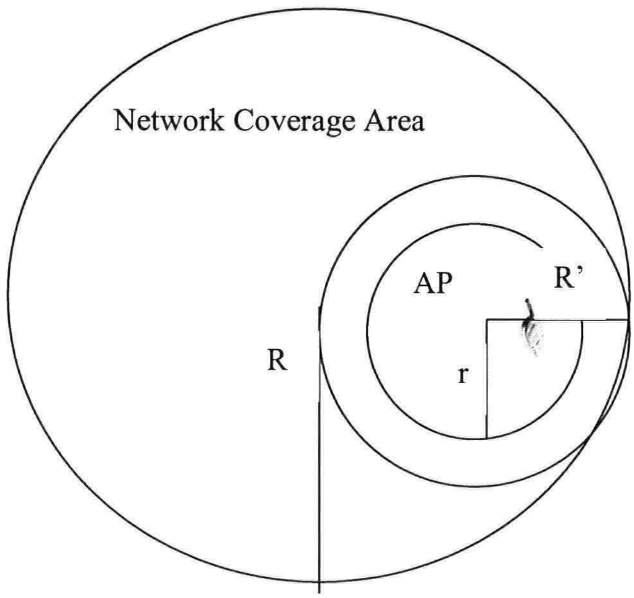
1) Charge Model – Charge model would hold information regarding the cost policy of the provider. E.g. per min, per sec, per byte downloaded, etc.
2) Cost – Cost would hold cost value per unit of the charge model.
3) Sliding scale protection - Sliding scale prices vary depending on the time the services are used for [14]. The user might get an attractive price to start with but if the price increases with usage the network would loose its cost benefit and may not satisfy user's requirements. To minimize the affect of marketing policies we propose to document cost as usable period prices in the repository. Cost would be documented in chunks of 1min, 5min, 10min and over 10mins.

4.1.5 Conceptual design of the database

The databases would be categorized by the main service type's data, voice and video. The data would also be subdivided based on operator type: large or small. The main difference between large and small operator would be the location tracking. Only zonal databases would allow location tracking and accommodate smaller operators. Larger operators would list out details relevant to the zone, e.g. operators frequency band is 1.8 GHz to 2 2GHz but only uses frequencies 1.90 to 1.92 GHz in particular zone. It would reduce MN's scanning overhead if unused frequencies in the zone are removed from the information passed to it. Zone databases will only allow location tracking for smaller operators but operators can choose to provide location tracking on their databases. Location based tracking can also be a complement to predictive handover schemes, if any such handover management policy is implemented by the operator.



Effective network coverage area would also be documented in the records of small operators. The coverage area of the network is not the absolute coverage area of the network because radius of the coverage area of the network might provide misleading information to the terminal. For instance in the figure below R is the radius of the network coverage and r is the radius of access point coverage area. Providing R would not provide any useful information to the MN because it would not be able to relate it to the centre of the network whereas comparing r and R' would give MN a rough idea of guaranteed coverage in any direction if it moves out of the access point coverage area by the same network. R' is the biggest circle in the network that can be drawn with AP's centre as its centre.



4.1.6 Queries based on MN's requirements

The user would be able to specify exactly what they are looking for and search the database more effectively. Users would be able to query the database based on any combination of fields. Queries can be based on any combination of the fields. Some of the important fields are

1) Interface – The terminal can specify the interface type in query.
2) QoS – The user/application can specify what QoS class it is looking for.
3) Security - The user/application can specify what security class it is looking for.
4) Price – The users can specify the price band for the service they desire to use.
5) Service description – User can search for special service like the U2 concert live broadcast, etc.
6) Location – Smaller networks are searchable based on their location and network coverage area.

The users can search the database in any combination of fields but are discouraged to base their searches on QoS, security and price because if; what user requires is not exactly matched by any network, the user's service selection program [10] might have to make some tradeoff's and hence would require the details of other operators providing that service.

5. Performance analysis

The database advertisement mechanism would be affected by the delays introduced by the access mechanism employed by the terminal and the database query execution and results retrieval.

Database access Latency =

AccessMechanismDelay	+	[Delays introduced by the access mechanism]
DBExecuteQuerry	+	[Querry execution delay]
DBresults Re*trieval*		[Delay in retrieving the records from the database]

-------------------------------- (1)

Equation 1 gives the total delay introduced by the service discovery. The data repository does not have any control on the delays introduced by the access mechanism which includes transport latency. The database delays include the time taken by it to execute the query and do all the necessary calculations like location based information retrieval.

5.1 Test Scenarios

To test the efficiency of the database we have devised the following test scenarios –

1) Number of parameters against time – Since the user is allowed to narrow down his/her search based on their exact requirements; the queries would vary in terms of parameters. The total number of parameters in the database permitted for search is 14. The first test would identify delays introduced due to the change in the number of search parameters.
2) Location based search – The test would analyse the delays introduced by the location tracking feature when in use and when not in use.

To limit the influence access mechanism factors in the evaluation, the test would be carried out on one machine acting as both client and server. The configuration of the test machine is –

1) Pentium IV 1.6 GHz

2) 256 MB RAM

3) Microsoft Access XP

4) Visual Studio .Net

6. Results

The affect of the number of search parameters is reflected in the graph below. The latency does not increase significantly with the increase in the number of search queries as shown in the graph below. Hence, the users should not be discouraged by having to limit the number of search parameters.

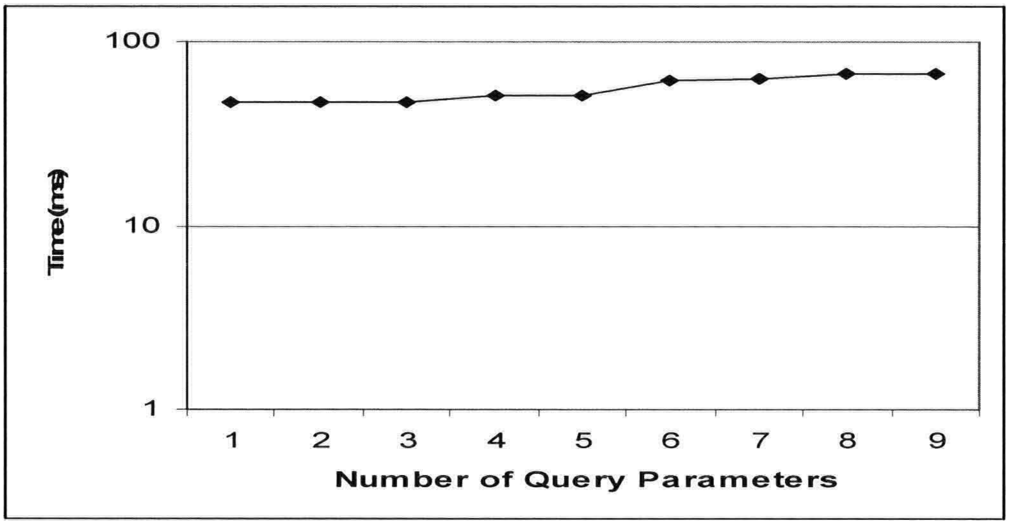
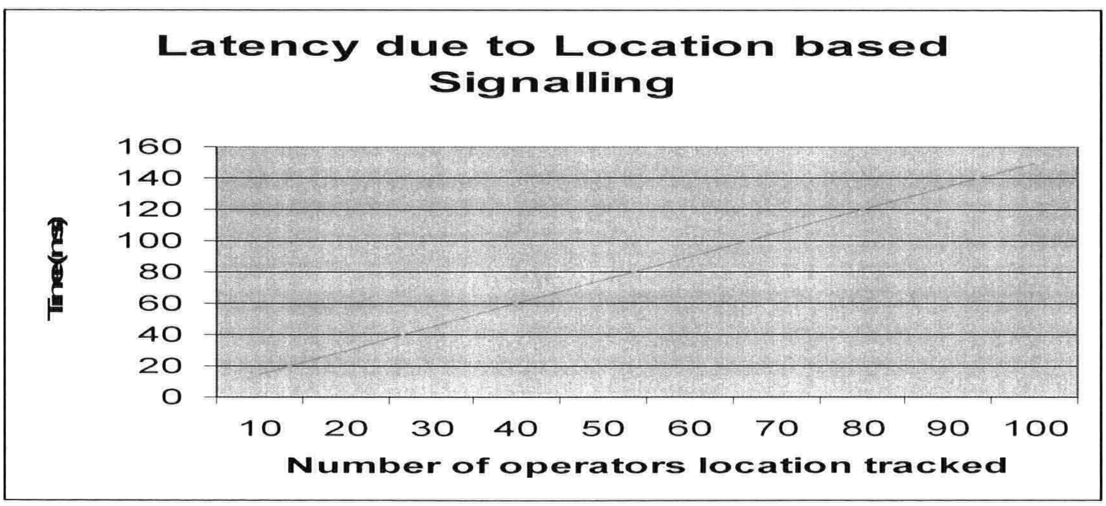


Figure 2 - Latency against the number of search parameters

The affect of location based search latency are illustrated in the graph below. The results suggest that for every 10 operators it takes on an average 15 ms for the distance to be worked out between the terminal and the access point. The load of location based searches would increase with the number of users as well. The load of location calculation could be shared between the terminal and the data repository server. The graph is very linear and the latency does not increase as more and more distance calculations are performed.



7. Drawbacks

1) The users should use the database reasonably to make sure everyone gets fast and efficient access. The users should generalise their searches to enable their selection algorithms [8] enough information to find suitable alternatives if users requirements are not completely met by any operator. The database use might have to be charged to enforce users to generalise their search rather than having to make repeated searches.
2) The repository raises the issue of management. Operators should have editable access to their own records and the users should have a non-editable access to their own records. The database would have to be managed by a neutral body to enable fair access to everyone.

8. Conclusion

Service advertisement mechanism would form the basis of multi-operator multiple access type networks. The data repository is not dependent on the access mechanisms whilst still enabling the user to express their requirements and retrieving results based on that. The repository reduces the effective scanning time, enables location based searches and specialized searches. Location tracking has its advantages but requires considerably more processing time than other queries. Location tracking load could be reduced by sharing the load between the terminal and the repository server. The other parameters do not put a strain on the server but careless use would limit the benefits of the service. The repository service might have to be charged to force users to use with consideration for others. Data repository is the first hierarchical advertisement retrieval system which is independent of the access technology. The advertisement repository enables inter and intra network resource discovery.

9. REFERENCES

[1] "4G - Beyond 2.5G and 3G Wireless Networks," vol.2004.www.mobileinfo.com/3G/4GVision&Technologies.com

[2] Clapton.A, Future Mobile Networks: 3G and beyond: Institution of Electrical Engineers, 2001.

[3] D Wisely, E Matjina, "Evolving systems beyond 3G - the IST BRAIN and MIND projects," BT Technology Journal, vol. 21, pp. 102-120, 2003

[4] P. Kankaanpaa, "Candidate Access Router Discovery," 2003

[5] L Burness, David Higgins, A Sago, P Thorpe, "Wireless LANs - present and future," BT Technology Journal, vol. 21, pp. 32-47, 2003

[6] Saurabh Bansal, Aboubaker Lasebae, Richard Comley, "Global Handover Latency Minimisation in HMIPv6," WSEAS Transactions on circuits and signals, vol. 2, 2003

[7] Wenhui Zhang, Juergen Jaehnert, Klaus Dolzer, "Design and Evaluation of a Handover Decision Strategy for 4th Generation Wireless Networks,"

[8] Saurabh Bansal, Aboubaker Lasebae, Richard Comley, "Freedom to choose along with the Freedom of Mobility", ICTTA'04 IEEE Section France Damascus, 2004

[9] Gang Wu, Paul J. M. Havinga, Mitsuhiko Mizuno, "Wireless internet over heterogeneous wireless networks", IEEE, 2001

[10] Aboubaker Lasebae, Saurabh Bansal, Richard Comley, "Service advertisements for future mobile communication networks", IEEE GCC'04 Bahrain, 2004

[11] N Blefari-Melazzi, D. Di Sorte, M. Femminella,, L. Piacentini, G. Reali, "Resources Discovery and Selection in Wireless Networks", IEEE Communications Society, 2004

[12] Saimak Naghian, "Hybrid predictive handover in Mobile Networks", IEEE, 2003

[13] R. Wisely David, IP for 3G : Networking technologies for mobile communications. New York ; Chichester: Wiley 2001, 2001

[14] http://www.virginmobile.com/mobile/services/tariff/tariff.html