Intelligent Proactive Handover and QoS Management using TBVH in Heterogeneous Networks

Fatema Sabeen Shaikh

January 2010

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

School of Engineering and Information Sciences

Hendon. NW4 4BT

United Kingdom

ABSTRACT

In a fourth generation (4G) communication system consisting of a diverse variety of wireless access standards, network availability, coverage and Quality of Service (QoS) may rapidly change relative to the speed and motion of the mobile node. Deliverance of seamless connectivity involves the development of intelligent proactive mechanisms for efficiently predicting vertical handovers. The complexity of the challenge increases further with the consideration of random device mobility patterns. Geographical topologies such as indoor and outdoor environments also exert additional constraints on network coverage and device mobility. The ability of a device to acquire refined knowledge about surrounding network coverage can significantly affect the performance of vertical handover prediction and QoS management mechanisms.

This research study adopts a proactive, mathematical model-based vertical handover prediction approach to propose a unique set of solutions that equip the multi-interfaced 4G client with new intelligence which allows it to accurately predict network coverage boundaries at a given location relative to the motion and speed of the mobile node. Specialised client-based mathematical models targeting indoor and outdoor environments dynamically apply knowledge on device location and direction, network coverage and geographical topology to derive a new metric called Time Before Vertical Handover (TBVH). This metric provides a clear, quantitative measure of the time the device has available to spend in the current network before it performs a vertical handover. The proposed TBVH models are simulated in OPNET Modeler and an evaluation of their performance reveals superiority over Received Signal Strength-based techniques in their accuracy to correctly predict vertical handovers.

The second part of the research study proposes a client-based Stream Bundle Management (SBM) Layer for downward QoS management which addresses issues including network selection, QoS negotiation and resource management. The increased awareness of a network's availability through TBVH allows the SBM layer to not only select the best network which meets the requesting application QoS requirements, but to also negotiate resources dynamically and efficiently in a manner which results in the avoidance of unnecessary vertical handovers. The SBM layer demonstrates improved performance and intelligence over existing network selection approaches, particularly in its ability to intelligently manage network resources and avoid unnecessary vertical handovers.

He is like the sun: to the eye at a distance

it seems small, but when near,

it dazzles the sight.

- AL BUSIRI

To the beloved one Muhammad who is praised in the heavens and earth, may abundant peace and blessings of the Creator shower upon him, his family and companions forever.

To my parents, for all their love, kindness and sacrifice,

and

To my dear sister Kalthoum for all her support.

CONTENTS

LIST OF FIGURES

Figure.   **Page**

LIST OF TABLES

Table.   **Page**

ACKNOWLEDGEMENTS

LIST OF PUBLICATIONS

- G. Mapp, **F. Shaikh,** M. Aiash, R. Vanni, M. Augusto, E. Moreira, *"Exploring Efficient Imperative Handover Mechanisms for Heterogeneous Networks"* Proceedings of EUPS, August 2009.

- **F.Shaikh**, A. Lasebae, G. Mapp, "*Proactive Policy Management for Heterogeneous Networks." P*roceedings of ICTTA, April 2008.

- G. Mapp and **F. Shaikh,** J. Crowcroft, D. Cottingham. "*Y-Comm: A global architechture for heterogeneous networking.*" Proceedings of WICON'07, October 2007.

- **F. Shaikh**, G.Mapp, A.Lasebae, "*Proactive policy management using TBVH mechanism in heterogeneous networks.*" Proceedings of NGMAST'07, September 2007.

- G. Mapp, D. Cottingham, **F. Shaikh**, P. Vidales, L. Patanapongpibul, J. Baliosian and J.Crowcroft, " *An Architectural Framework for Heterogeneous Networking*", Proceedings of WINSYS 2006.

- **F.Shaikh**, A. Lasebae and G. Mapp, "*Client-based SBM Layer for Predictive Management of Traffic Flows in Heterogeneous Networks*", Proceedings of IEEE ICTTA 2006.

- **F.Shaikh,** A. Lasebae and G. Mapp, "*Stream Bundle Management Layer for Optimum Management of Co-existing Telemedicine Traffic Streams under Varying Channel Conditions in Heterogeneous Networks*", Proceedings of MIT 2005.

- **F.Shaikh**, A. Lasebae and G. Whitney, "*Critical Analysis of High-Speed Wireless Standards for Transferring Telementoring Information*", Proceedings of IEEE ICTTA 2004.

CHAPTER 1

4G Handover and QoS Management Issues: An Overview

1.1   Introduction

In the field of wireless communications, ongoing developments in wireless technologies have been coupled with a remarkable increase in nomadic computing activities, with users demanding improved network connectivity and services for supporting high-speed multimedia services like interactive video telephony and voice over IP (VoIP). While the demand for wireless connectivity has increased, unlike wired networks where it can be fulfilled by upgrading communication media, in wireless networks, resources like bandwidth remain limited. This requires the development of sophisticated and cost-effective solutions that can efficiently accommodate the increasing number of users on to these limited resources, while ensuring the delivery of adequate Quality of Service (QoS). Wireless connections also display an additional constraint on the guaranteed delivery of network resources as channel resource availability fluctuates continuously due to user mobility and the inherently unstable characteristics of wireless links like co-channel interference, multi-path fading and bursty errors due to noise. These issues along with others have resulted in the rapid proliferation of wireless technologies with the development of a wide range of broadband wireless access standards targeting various domains, offering different levels of QoS and coverage at more or less affordable costs.

This chapter introduces the reader to some fundamental concepts and unravels the hidden problems in the rapidly evolving field of 4G heterogeneous networking in the area of vertical handovers. Section 1.2 introduces the wireless 4G heterogeneity paradigm and section 1.3 discusses the handover concept. Section 1.4 uncovers important problems which remain unresolved in the area of seamless connectivity during vertical handovers and which serve as a motivation for this research study. Section 1.5 outlines the research questions. The limitations of scope for this study are listed in section 1.6, main contributions are listed in section 1.7, and finally the chapter concludes by describing the structure of the thesis in section 1.8.

1.2   Fourth Generation Heterogeneous Networking

Table 1.1 summarises the defining characteristics of popular and upcoming wireless access standards. Each network provides a different level of QoS in a particular domain and no standard can be considered capable of providing ubiquitous broadband coverage at all times by itself yet. While cellular technologies like WCDMA and GPRS promise the provision of wide coverage and wireless data services, problems like high round trip delays, packet loss, traffic burstiness and low bandwidths [Rodriguez, et al. 2004] hamper their ability to deliver high QoS to end users. Wireless LANs (WLAN) like IEEE 802.11 succeed in the delivery of high data rates and lower delays, but data transmission is restricted to small coverage areas called hotspots. WIMAX on the other hand, needs to undergo further development in order to support good service in indoor environments. These different capabilities offered by individual networks coupled with the surging interest in "anytime anywhere" high-speed connectivity have led to the conceptualisation of Fourth Generation (4G) heterogeneous networking.

The vision of 4G heterogeneous networking is the provisioning of universal connectivity and mobility

Table 1.1 Wireless Access Standards

Coverage Type	Network	Coverage	Data Rate	Reference
Personal Area Network (PAN)	Bluetooth	10 m	up to 700 kbps	[Ganz, et al. 2004]
	Ultra Wide Band	10 m	252 Mbps – 1 Gbps	[Arslan, et al. 2006]
Local Area Network (LAN)	IEEE 802.11 (a-z)	35 – 5000 m	11 – 600 Mbps	[Bing. 2008]
	HIPERLAN 2	70 – 300 m	25 Mbps	[Zahariadis. 2003]
Wide Area Network (WAN)	UMTS/WCDMA	20 km	up to 2 Mbps	[Smith, et al. 2007]
	WIMAX	up to 50 km	up to 70 Mbps	[Pareek. 2006]
	LTE	up to 100 km	up to 326 Mbps	[Technologies. 2009]
	GSM	up to 35 km	9.6 – 144 kbps	[Smith, et al. 2007]
	Satellite	global coverage	Up to 2 Mbps (UL) 38 Mbps (DL)	[Ganz, et al. 2004]

through the seamless integration of different network access technologies offering diverse levels of

QoS. The popular description states that these different networks will all converge down to a common core IP-backbone [Tafazolli. 2005] and facilitate interoperability at all levels to provide a ubiquitous networking environment. This will enable multi-interfaced mobile devices to roam freely among networks without experiencing disruptions like connection loss during handovers, while giving them the choice of the best available location-based network services.

It is widely accepted that different networks in the heterogeneous environment will integrate in a loosely-coupled manner, with each network domain being independently deployed by different service providers [Song, et al. 2005]. The provision of continued connectivity and an acceptable level of QoS to users while they roam across different networks continue to be the dominant topic of interest with intense research activity being carried out in both academia and industry [Shenoy, et al. 2008]. On the other hand, location-awareness and multi-interfaced capabilities have already become indispensible features in mobile devices and network service providers are busy promoting Internet usage to smart internet phone users through affordable mobile internet tariff plans. Therefore, 4G heterogeneous networking can no longer be regarded as an exciting nascent concept confined to research laboratories; it has grown tremendously to become the dominant paradigm and can be viewed as the culmination of developments taking place in different areas in the field of communications.

1.3   Handovers in heterogeneous networking

Mobility is the key defining feature of all wireless networks which differentiates them from their wired counterparts. Ever since their inception, the smooth execution of handovers has been an important factor in gauging the performance of wireless networks. In a wireless network, coverage on the wireless peripheral side of the network is provided by an entity called the Base Station (BS) or Access Point (AP) (terms used interchangeably). The coverage area of a BS is called a cell. A mobile node (MN) residing in a cell attaches itself to the BS for communication purposes. The MN can roam among different BS cells during the lifetime of a connection. This means the migration of a connection from one BS to another, which is also called a handover.
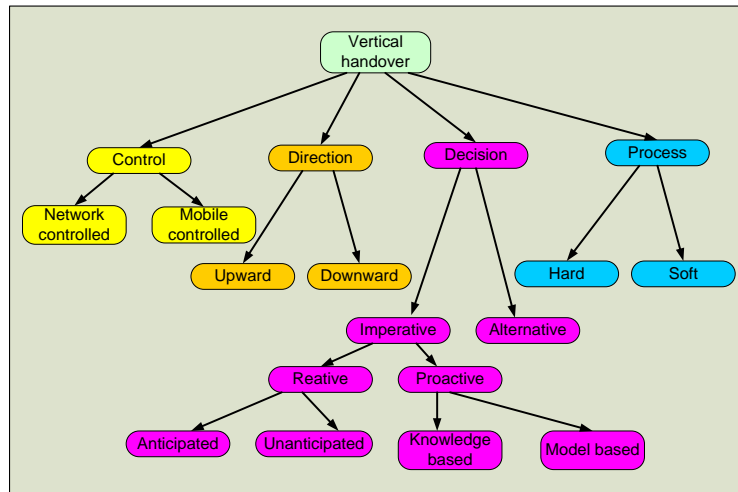
Figure 1.1 Classification of vertical handovers
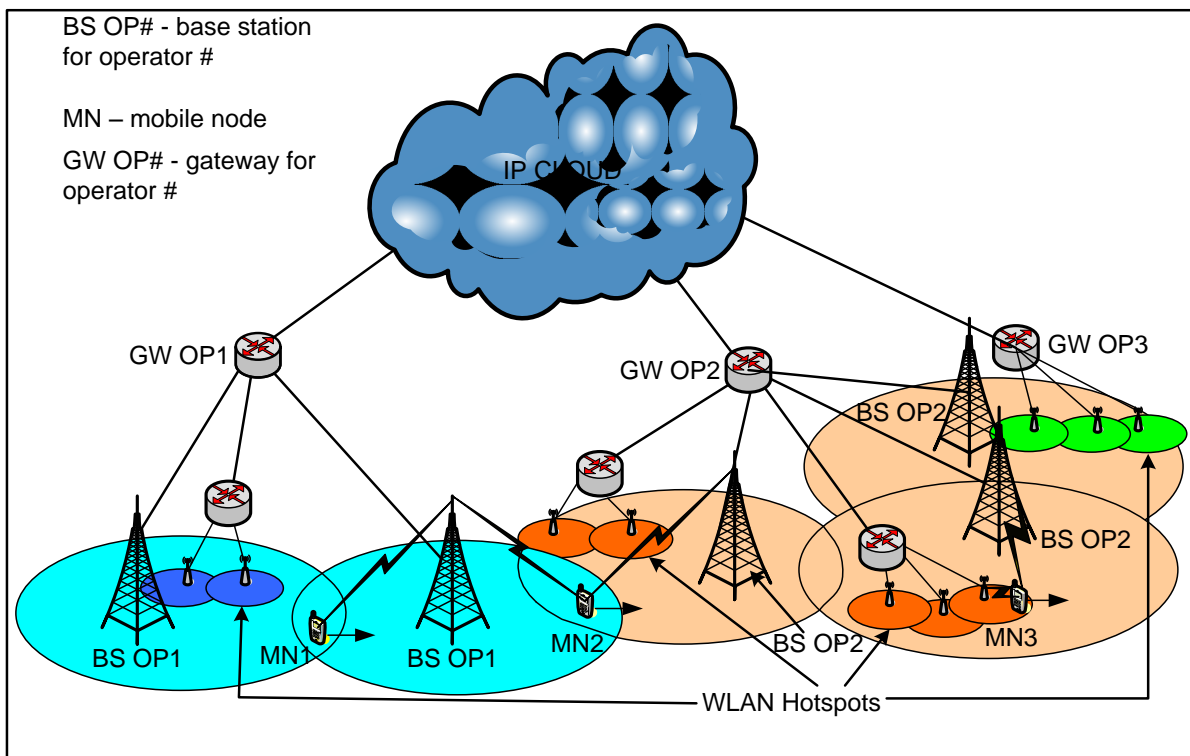
Figure 1.2 Handover scenarios in 4G networks

A **handover** is the process where the MN changes radio transmitter or access media used to provide the bearer services, while maintaining a defined bearer QoS [3GPP-Release 5]. A handover operation that can minimise or eliminate the delay for establishing a new connection with a BS is called a **fast handover**, and when the data loss is minimised it is named a **smooth handover** [Persson, et al. 2005]. A combination of both these qualities results in a **seamless handover** [Perkins. 2002]. Some of the key desirable features of a seamless handover are negligible delay, maximum throughput and low packet loss ratio along with the elimination of the ping pong effect [Ylianttila, et al. 2001]. The ping

pong effect is an undesirable effect caused when a MN performs frequent handovers between two BSs due to fluctuations in channel resources, resulting in additional signalling overhead and latencies.

Handovers can be classified into a wide range of categories based on the decision factors that result in their execution. This sub-section limits itself only to the classification of vertical handovers relevant to this study as shown in figure 1.1. Figure 1.2 demonstrates the co-existence of two different access technologies, WLAN and UMTS deployed by two different operators represented by blue and orange coverage, together with the independent deployment of WLAN pico-cells represented by green coverage. A **horizontal handover** takes place when the MN changes its point of attachment from one BS to another BS belonging to the same technology, and at the same level of network hierarchy. A **vertical handover** happens when the MN switches connection to a new BS belonging to a different access technology higher or lower than the current network in the hierarchy. During a vertical handover if the connection with the old network's BS is broken before being established with a new network's BS, the MN is said to perform a **hard handover**. In a **soft handover** the connection with a new network is established before the connection with the old network is lost.

In an **upward** vertical handover, the MN switches from a smaller-coverage, high-bandwidth network to a larger-coverage, low-bandwidth network e.g. WLAN to UMTS. The **downward** vertical handover mechanism is opposite to the upward vertical handover. In figure 1.2, the movement of MN1 from the coverage of one BS to another represents a horizontal handover. MN3 undergoes an upward vertical handover as it is moves out of WLAN coverage in to UMTS coverage.

Categorisation based on the entity that decides to perform a handover results in **client-controlled** and **network controlled** handovers. In a network-controlled approach, the network maintains an up to date knowledge of context information at the MN and decides when and how it should perform a vertical handover. This task can be very complex in the case of multi-interfaced devices where the MN is connected simultaneously to several BSs of different networks and experiences largely varying QoS at each interface. In a loosely coupled scenario it becomes increasingly difficult to assign complete control to any single BS. Even if this takes place, a continuous flow of the latest context information from the MN's multiple interfaces to the controlling BS involves the transfer of a huge amount of information, resulting in large network overhead. The network-controlled approach requires a high level of interactivity among BSs belonging to different network domains, controlled by

independent service providers. This will result in increased complexity due to the resolution of a large number of technical and administrative issues arising from the sharing of confidential network and customer information, something service providers may not be willing to do [Shaikh, et al. 2006].

In the client-controlled approach, as the multi-interfaced client is directly connected to different networks it possesses up-to-date context knowledge of the medium access, network and transport conditions for each active network interface. Hence it is in a more superior position to take decisions on important issues such as handovers, QoS management, and network selection. The MN in this case must possess the ability to negotiate QoS and switch to an appropriate network at the right time to get the best utilization of network resources [Shaikh, et al. 2006, Chen, et al. 2005]. In 4G heterogeneous clients, the crucial role of handover related decision-making has to a large extent shifted from the network side towards the client side and an increasing number of studies have adopted the client-controlled approach for vertical handovers [Ahmed, et al. 2006, Kassar, et al. 2008, Song, et al. 2005, Vidales. 2005, Xiaohuan Yan, et al. 2008].

Based on the category which defines the reason behind handover, vertical handovers are classified as imperative and alternative. An **imperative handover** takes places because a MN has determined through technical analysis that it is good to do so. A failure to perform an imperative handover can result in a severe loss of performance. **Alternative handovers** take place when the preference is given to a network based on other non-technical reasons such as price or incentives. This research study focuses on both imperative and alternative handovers and demonstrates how an increased intelligence about network availability can result in a more accurate prediction of handovers.

Imperative handovers can be **reactive or proactive** in nature. In reactive handovers decisions are mainly reactions to available network information and no attempts are made to predict likely future conditions. Clients in reactive schemes explicitly request context information from lower layers [Kupper. 2005]. Feedback obtained from lower layers then alerts mechanisms about change in network conditions [Mapp, et al. 2006]. **Anticipated** reactive handovers are soft handovers where the MN is able to handover to other BSs. **Unanticipated** reactive handovers are hard handovers where there is no new BS to which the MN can attach itself.

In a proactive handover the MN attempts to predict future conditions through the evaluation of measurable network parameters like coverage and channel conditions. Context information is automatically sent to the client without explicit requests for it [Kupper. 2005]. Proactive mechanisms possess the ability to reserve resources in advance based on the knowledge of network parameters such as topology, coverage and positioning information. They can be further classified into **knowledge-based** and **modelling-based** proactive handovers. The knowledge-based approach makes use of pre-recorded network coverage information together with a MN's location to predict the availability of different networks at a particular location. The modelling-based approach predicts future conditions with the help of mathematical models. This approach is more flexible in accommodating random MN movements and offers the advantage of being easily adoptable by both simulated and real-time systems. This research study adopts the proactive, model-based vertical handover prediction approach.

1.4   Seamless vertical handover – the challenges

In the wireless heterogeneous environment, the majority of problems in the case of handovers stem from the issue of uncertainty in the prediction of network coverage and the duration of availability of network services. The situation becomes even more complicated with the consideration of random device mobility patterns [Liu, et al. 2008]. Although the importance of this issue has been acknowledged by the wireless community, the implications to overall system performance have not been explored. This section unravels the true extent to which the problem of increased network unpredictability can directly affect the performance at the network, device and application levels, particularly during vertical handovers.

In wireless heterogeneous environments the MN at any instant may fall under the coverage of several different networks that vary in physical, medium access and link layer characteristics [Song, et al. 2005] and offer a variety of different services. Handovers are no longer executed simply for the sake of maintaining connectivity during mobility but instead the whole process adapts a more opportunistic approach for deriving maximum benefit from all available networks. The handover decision phase now becomes an even more complex and crucial step towards achieving seamless mobility in fulfilment of the vision of being 'Always Best Connected' [Gustafsson, et al. 2003].

The most wide adopted concept for the client-based vertical handover approach consists of three well-defined phases – system discovery, handover decision and handover execution [Shenoy, et al. 2008, McNair, et al. 2004, Wen-Tsuen Chen, et al. 2005]. This research study focuses mainly on the first two stages of a vertical handover i.e. system discovery and handover decision as correct decisions taken in these two stages will play a crucial role in the successful implementation of a seamless vertical handover. The main goal of vertical handover optimisation techniques is to devise solutions that can minimise and even eliminate disruptions caused due to handover. An effective way of achieving this is to equip the network and MN with the ability to proactively detect vertical handovers before they actually take place so that the devices can start procedures to prepare and adjust to impending changes in network conditions.

1.4.1   Uncertainty in system discovery phase

The system discovery phase is when the MN detects new wireless networks and learns about their offered services [Stevens-Navarro, et al. 2008]. In a multi-interfaced client this phase consists of monitoring all interfaces for channel conditions. In the case of active interfaces, a MN periodically checks for service advertisements specifying signal quality, available resources and offered QoS.

In heterogeneous networking, different wireless networks of diverse coverage and services are independently deployed, forming an overlapped composite coverage area consisting of PANs, LANs and WANs. Due to uncertainty in network availability and wide variations in coverage characteristics, it becomes difficult to decide when and for how long a device should scan channels to sense for a network on a particular wireless interface. A multi-interfaced MN must activate all its interfaces to receive service advertisements from different BSs. In order to ensure that a new wireless network was detected as soon as it became available, a common practice in the past was to keep all interfaces activated all the time. However the simultaneous activation of multiple interfaces means a huge increase in power consumption.

Previous results have shown that among the three handover phases, the channel scanning period in the system discovery phase is the most time consuming, contributing to almost 90% of the total handover latency [Mishra, et al. 2003]. The most frequent vertical handovers in a typical 4G environment are expected to take place at the LAN-WAN level. Therefore it is essential to minimise

the number of instances a MN will need to go through the system discovery phase through the improved prediction of strong BSs for attachment. It is also important to minimise the unnecessary attachment to temporary BS coverage which can lead to unnecessary vertical handovers.

1.4.2 Uncertainty in handover decision phase

The handover decision phase in heterogeneous networking is the most important step that directly affects a MN's communication [Persson, et al. 2005]. It is an increasingly complex process which aims to answer three fundamental questions about vertical handovers:

- *'When?'* – The quest for the answer to this question has resulted in the area of handover prediction.
- *'Which?'* – The answer to this question forms the area of network selection.
- *'How much?'* – The answer to this question is sought through resource allocation and QoS management techniques.

Each of these fundamental questions forms a unique area of research attracting an immense amount of activity in both academia and industry.

1.4.2.1 The importance of correct handover prediction

In the handover decision phase, correct handover prediction is of significant importance. An incorrect answer to the question WHEN? which relies on the knowledge of a network's availability can lead to an overall degradation in performance due to instability in other phases of the vertical handover, even resulting in connection loss. In fact the correctness and accuracy of decisions made in all the three handover stages is largely dependent on the accuracy of prediction of the availability of a network's coverage. The performance of all mechanisms linked directly or indirectly to vertical handovers can be greatly enhanced by obtaining an answer to the complex yet increasingly important question

"*When is the device expected to perform a vertical handover?*"

Therefore it is not an exaggeration to state that an accurate knowledge of the duration of availability of a network in relation to a MN's motion within that network is crucial to the successful management of all aspects of functioning in 4G heterogeneous networks.

1.4.2.2 The need to avoid unnecessary vertical handovers

For vertical handovers in heterogeneous wireless networks to be successful one important issue that needs to be eliminated is that of unnecessary vertical handovers. This means that the MN should remain connected to the new network for duration equal to the handover recovery period. This is the time in which the data received on the new interface is equivalent to at least the amount that would have been received on the old interface in the duration equal to the total handover procedure. Otherwise the handover will be considered as unnecessary if the MN is forced to perform a vertical handover once again before the recovery duration period expires.

The main causes for unnecessary handovers are the failure to recognise temporary coverage, unavailability of required resources and congestion in the new network. Among these, the problem of predicting temporary coverage still remains largely unsolved. For instance, A MN roaming into the strong but temporary coverage of a WLAN may have access to the most optimal resources and the most favourable channel conditions. Despite all this, the fact that it will have to perform an upward vertical handover before successfully utilising these resources means that their availability is virtually useless unless it is harnessed in the correct manner. An unnecessary vertical handover actually results in an increased signalling overhead and delay.

1.4.2.3 The impact of geographical topology on handover prediction

In the prediction of vertical handovers, a crucial piece of information that has so far been ignored by studies is the effect of geographical topologies and physical boundaries. This knowledge is of significant importance as it can affect the validity of a MN's decision to perform a vertical handover. Take the example of a MN which is located inside a closed indoor environment and which moves towards the boundary of WLAN coverage. According to a pure Received Signal Strength (RSS) based handover prediction approach rapidly decreasing RSS indicates that this MN is moving towards the coverage boundary, prompting the device to begin preparation for vertical handover. However if the network's coverage boundary falls close to but beyond the environment's physical boundary like a wall, the reality is that the device cannot experience a vertical handover as it will be prevented from exiting from the current coverage by the physical boundary. Scenarios like these are becoming increasingly common due to the widespread deployment of WLAN hotspots. Therefore a key

requirement for 4G heterogeneous networking is a more detailed and refined knowledge of the geographical topology surrounding the MN.

1.4.2.4 The importance of intelligent, context-aware QoS management

The second and third stages in the handover decision phase which attempt to answer the questions WHICH? and HOW MUCH? are closely inter-related and are responsible for network selection and downward QoS management respectively. Downward QoS management in 4G wireless clients consists of a set of sophisticated mechanisms that enable applications to specify their QoS requirements and then execute solutions that manage these requirements intelligently over available network channels.

The main goal during a vertical handover is to make the most efficient use of network resources and provide an improved level of QoS to applications by reducing packet loss and latencies. A major challenge faced by a multi-interfaced client is the effective delivery of multimedia traffic across diverse network channels, by minimising forced termination of ongoing connections during handovers, while accommodating new connections. In the presence of multiple networks, a roaming MN is required to continuously analyse application demands, device behaviour and network conditions at each active physical interface before mapping applications on to the most appropriate network. One of the key objectives of this research study is to demonstrate the positive influence that the knowledge of time before vertical handover exerts on refining the MN's choice of network and resources resulting in a significant improvement in performance.

Network selection consists of attaining the most favourable tradeoff among user preferences, service applications and network conditions [Song, et al. 2005]. The device needs to possess the ability to reconfigure itself dynamically. It needs to select and adapt to the most appropriate wireless access standard for handling conditions encountered in specific service area regions and times zones of the days [Glisic. 2006]. The presence of multiple network channels at the MN in heterogeneous networks, offering different levels of QoS increases the complexity of multi-class traffic management issues such as resource management, traffic scheduling and flow control. Downward QoS management at the MN requires answers to several key issues including:

- The QoS requirements of application streams.

- Most suitable networks among currently available ones for allocating a particular call.

- The current and likely future conditions of these networks.

- How long are these networks likely to remain available (time before vertical handover).

Once again the ability to predict surrounding network coverage emerges as an increasingly indispensible requirement for the 4G client. In fact the availability of network coverage can become a precondition for the collection of other network QoS parameters [Qingyang Song, et al. 2005]. This new level of awareness will empower it to refine its performance which in turn will have a deep impact on the accuracy of calculated results.

The issue of the avoidance of unnecessary vertical handovers can be a tricky one in the presence of multi-class traffic and requires a refined knowledge of an application's resource demands, the actual availability of these resources and the duration for which they are likely to remain available. The possibility of a handover being unnecessary is dependent mainly on whether the new network can satisfy the requesting traffic stream's resource demands in the limited period of connectivity. For instance, a downward vertical handover will not be considered futile if the enqueued data in the MN is a set of emails which can easily be sent in the limited period of time that a MN connects to the temporary but free hotspot. However, the new connection may not be suitable for starting a VOIP connection and the handover in this case will be considered unnecessary.

1.5   Research Questions

Targeting the unaddressed challenges in vertical handover prediction and traffic management in heterogeneous clients that arise due to unpredictable network coverage, this research study aimed to find out answers to a set of important research questions which include:

- **How can a MN roaming within a network predict future network availability relative to its motion within the network, and determine how long it has before it performs a vertical handover?**

  This research question required the development of a new set of mechanisms which proactively and dynamically empowered the MN with increased awareness about its surrounding coverage environment. The main goal here was to propose a new dynamically derived parameter that provided different device layers with a clear, quantitative measure of the time available in a

particular network until a vertical handover or Time Before Vertical Handover (TBVH). A key advantage of this unique parameter was that it could be applied to different domains and assisted the device layers to proactively prepare for vertical handovers in advance.

- **How can the knowledge of network coverage availability be utilised to**
  - **assist in the minimisation of unnecessary vertical handovers?**
  - **optimise the performance of network selection and QoS management mechanisms in a heterogeneous client?**

The answer to this research question was the development of a new QoS management layer called the Stream Bundle Management (SBM) layer which inculcated new information on coverage availability (TBVH) into its improved network selection, call admission control and resource management strategies. It combined the context knowledge on user objectives, application requirements, network conditions and coverage availability to take informed decisions on multi-class traffic management in a heterogeneous client. An important task was the correct application of the knowledge of context to recognise and avoid unnecessary vertical handovers.

1.6   Limitations of Scope

- As this research study mainly delves into optimisation of the handover decision phase which takes place before the actual execution of the vertical handover, improvements to the handover execution phase fall outside the scope of this thesis. It instead adapts the current popular solutions that exist in areas like mobility management and network security.
- The solutions proposed in this thesis are based on a set of achievements in a variety of related areas including hardware optimisations, localisation techniques, context-awareness solutions, and even human-computer interaction. The aim is to concentrate on applying attractive solutions available in these areas and not to work on improving their performance.
- The most frequent occurrence of vertical handovers in the 4G heterogeneous environment is likely to take place when a MN moves in and out of a smaller WLAN coverage. Therefore an important step towards seamless roaming involves mitigating the effects of handovers near WLAN boundaries. The research study mainly targets this issue in WLAN during the development

of TBVH prediction models and assumes a uniform WAN coverage for other networks higher up in the hierarchy.

1.7   Main Contributions

By answering the two research questions introduced in section 1.5, this thesis made an important set of contributions which include:

- An exhaustive critical review of existing solutions in the areas of handover optimisation, network selection and QoS management in 4G heterogeneous networks which unravels important deficiencies that hamper the successful realisation of seamless connectivity and roaming.

- Simulated models for open, enclosed and closed environments that proactively derive new context information on geographical topology and network coverage.

- A new metric called Time Before Vertical Handover which provides a clear, quantitative measure of the duration of a network's availability. It displays the ability to be applied easily to a wide variety of handover prediction and network selection algorithms to improve their performance.

- The client-based Stream Bundle Management layer for downward QoS that possesses the unique ability to process a large variety of complex context information in an efficient manner to minimise disruptions due to vertical handovers that can disrupt a smooth user experience.

- A network selection mechanism that takes into account a large variety of information including user objectives, application requirements, network conditions, device position and mobility patterns, coverage availability and geographical context before bundling traffic streams on to available wireless channels.

- An intelligent resource management and call admission control mechanism that proactively negotiates resources with the network in order to minimise the occurrence of unnecessary vertical handovers.

1.8   Structure of dissertation

The outline of the thesis is as follows:

- **Chapter 2** provides an overview of background concepts. It introduces the reader to latest developments in the field of heterogeneous wireless networks and related areas which the study aims to harness in order to fulfil the research goals.

- **Chapter 3** is a comprehensive literature review of related work in the field of handover prediction. It acknowledges important research work conducted in this area which the study builds upon and identifies key deficiencies in existing approaches due to the inability of current devices to predict network coverage availability.

- **Chapter 4** proposes a unique set of solutions to solve the problem of network coverage prediction in WLAN through the development of a set of proactive mathematical models that dynamically derive the new parameter TBVH.

- **Chapter 5** discusses the modelling and simulation of the solutions proposed in chapter 4 in OPNET Modeler. Simulation results are discussed in detail and the performance of TBVH based handover prediction mechanisms is evaluated against the performance of RSS based handover prediction.

- **Chapter 6** is the beginning of the second part of the research study which deals with performance enhancement through intelligent network selection and QoS management. This chapter is a literature survey of existing solutions in network selection and resource management techniques. It highlights important achievements and recognises key gaps in knowledge which need to be addressed.

- **Chapter 7** introduces the concept of the Stream Bundle Management layer for downward QoS management. It explains in detail the development of novel multi-class traffic stream management strategies which include network selection, call admission control and resource management.

- **Chapter 8** demonstrates the successful functioning of the SBM layer in its refined ability to intelligently select the best network interface for a requesting traffic stream and to proactively detect and avoid unnecessary vertical handovers.

- **Chapter 9** concludes the thesis with a summary of the main contributions of this research study, along with a discussion on future work.

CHAPTER 2

Research Context – Networks Technologies and Architecture

2.1 Introduction

This chapter introduces the reader to different relevant background concepts in 4G heterogeneous networking needed to clearly understand the purpose of this research study. It also provides an overview of some recent important technological achievements in the field of wireless communications which have been adapted by this research study in order to fulfil the research goals outlined earlier in chapter 1.

The chapter begins with section 2.2 which is a discussion on individual wireless access technologies - IEEE 802.11 Wireless LAN (WLAN), IEEE 802.16 (WiMAX) and UMTS and elaborates on the main features relevant to this study. After a brief discussion on mobility management in section 2.3 it covers the main features of the IEEE 802.21 standard for 4G heterogeneous clients in section 2.4. In section 2.5 the chapter explores available location positioning solutions for both indoor and outdoor environments and finally section 2.6 culminates in to a discussion on one of the latest ambitious developments in the field – the YCOMM (Y-Communication) framework developed specifically to encapsulate the key challenges in 4G heterogeneous networking. This section outlines how the various technological developments discussed in this chapter and the solutions proposed by this research study will fit into the YCOMM architecture.

2.2 Overview of Select Wireless Access Standards

While the 4G heterogeneous network consists of a wide variety of wired and wireless network access standards exhibiting diverse features, this research study focuses only on the challenges arising from the integration of wireless networks. Among the commercially available wireless access standards this section limits itself to discussing the main features of three of the most popular ones – WLAN,

UMTS and WiMAX. Due to the immense vastness of the topics, the section further limits itself to a concise overview of technical features most relevant to this research study.

2.2.1 IEEE 802.11 Wireless LAN

The IEEE 802.11 Wireless Local Area Network (WLAN) is a set of wireless standards operating in the 2.4, 3.6 and 5GHz frequency bands. Ever since their deployment in 1999, WLANs have played a critical role in addressing the rising popularity and demand for wireless access at low costs. They can no longer be regarded as a luxury but rather as an essential means of communication in different domains like homes, businesses and educational institutions. Depending on its various flavours, the wireless access standard is capable of providing a wide range of data rates between 1 to 600MBps within a coverage radius of 50-300m and forms one of the key candidates of 4G heterogeneous networking. As WLAN is the key candidate technology for which the handover optimisation solutions proposed in this study are built, this sub-section describes in detail some of the important features based on the IEEE LAN/MAN Standards Committee's specification document [LAN/MAN Standards Committee. 1999].

2.2.1.1 The reference network architecture

The IEEE LAN/MAN Standards Committee has classified WLANs into two categories based on the manner in which mobile nodes communicate with their neighbours. In the first type called Ad hoc topology the Independent Basic Service Set (IBSS) consists of independent nodes coming together in an ad hoc manner where each station can form an independent connection with each neighbour. Ad hoc networking forms a vast research area in itself and falls beyond the scope of this study.

The Infrastructure topology consists of an access point (AP) configured to operate on a single channel and which forms the interface between the wireless edge and wired core of the network. All mobile nodes attach themselves to the AP in order to connect and communicate with other network elements. A Basic Service Set (BSS) consists of the AP and all the mobile nodes associated with it. When multiple APs connect together to the same wired network to offer a larger wireless coverage, the arrangement is known as the Extended Service Set (ESS) (Figure 2.1). Every WLAN is assigned
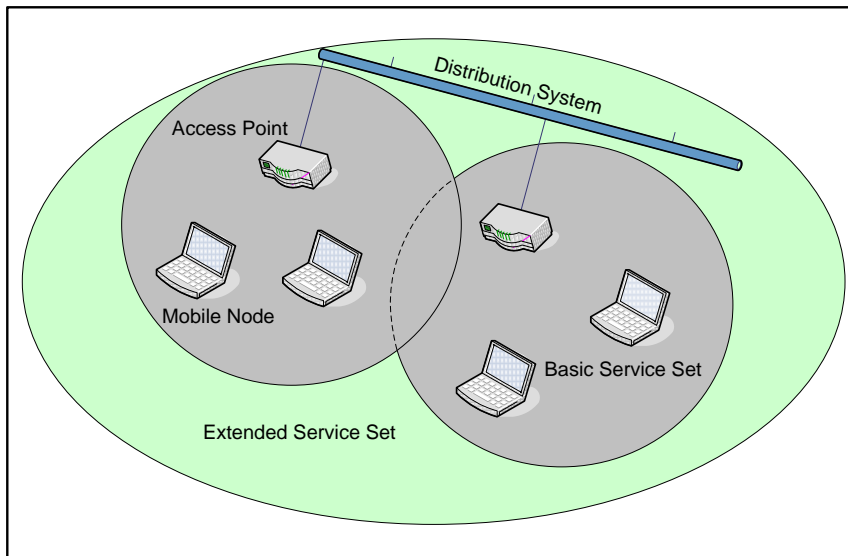
Figure 2.1 The Extended Service Set (ESS) in an infrastructure topology

a unique identifier called the Service Set Identifier (SSID). All APs and MNs trying to connect to a specific WLAN must use the same SSID otherwise they will not be permitted to join the BSS/ESS. A combination of the APs and wired network is known as the Distribution System (DS). A MN moving between different APs senses their signals and attempts to connect to the one which offers the strongest signal and best QoS. It thus experiences a handover which if not executed smoothly, can cause disruption to ongoing connections and even result in connection loss.

2.2.1.2 Connection establishment in WLANs

In WLANs, the absence of a stable physical network connection like a wired connection means that a MN needs to spend additional time and resources in continuously scanning its surroundings for optimum wireless coverage. Even while connected to an AP a MN is not guaranteed connectivity throughout the required period due to mobility and fluctuating channel conditions and thus may have to migrate its data connection to another AP in order to prevent connection loss. As a result wireless devices are constantly on the lookout for APs offering optimum connectivity. In WLANs, connection establishment is carried out in four steps through the transmission of a set of management frames between the MN and AP [Bing. 2008] as shown in figure 2.2:
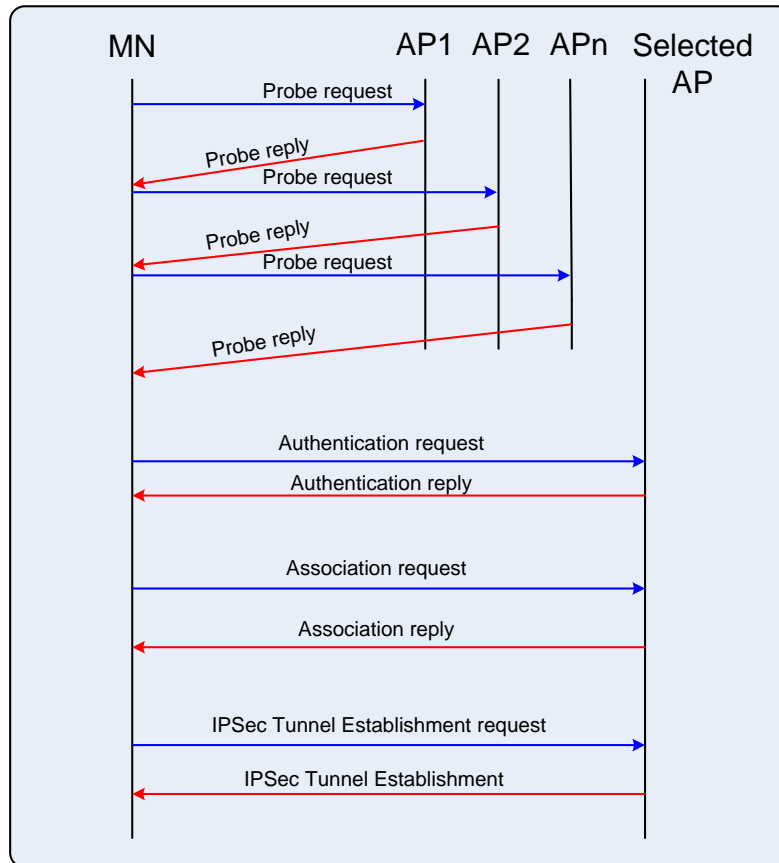
Figure 2.2 WLAN Connection Establishment [Bing. 2008]

- **Discovery**: In this step a wireless device scans radio channels for detecting the presence of access points and other stations. The scans may be either active or passive. In an active scan a MN sends an explicit probe request to each AP which then transmits beacons in response. In the passive scan a MN simply listens to radio channels for beacons transmitted by APs and then decides the final choice.

- **Authentication**: This step deals with the authentication of two devices at the start of an association. It is achieved through the exchange of special authentication packets. An association request progresses only after successful authentication, otherwise the recipient of the authentication request sends a de-authentication notification to the requesting device.

- **Association, Disassociation and Re-association**: This step involves the creation of an association between two stations or between a station and AP. A MN intending to connect to an AP in a BSS sends an association request. When handing over from an old AP to a new AP the MN sends a disassociation request to the old AP and a re-association request to the new AP.

37

Each MN can be connected to only one AP at any given time. It also consists of mechanisms for QoS management and call admission control.

- **Confidentiality**: This step is needed in order to ensure an encrypted form of communication between two devices sharing a public wireless interface. In a BSS an AP is responsible for enforcing security policies and advertises them in beacon and probe response packets.

At this point, it is important to refocus the reader's attention on the problem of uncertainty in system discovery which was highlighted in chapter 1, section 1.4.1. In order to avoid unnecessary vertical handovers to WLAN and the increased overhead associated with going through the connection establishment phase repeatedly, it is important that the correct choice for an AP is made right at the first step of connection establishment, i.e. during the system discovery phase. Otherwise efforts made in the remaining three phases can be rendered futile. This is the reason why this research study focuses mainly on improving the intelligence of the device during the system discovery and association phases. These phases are explained in more detail in the next sub-section 2.2.1.3. As for authentication and encryption techniques, they are beyond the scope of this study and the solutions proposed here rely on existing security standards to achieve the desired level of security.

2.2.1.3 AP discovery and association

In a WLAN MN, the AP discovery phase mainly consists of scanning for the availability of different APs before finalising the choice of a particular AP to associate with. Scan procedures can be initiated in two ways – manually by a user, or automatically through the deployment of system selection algorithms in response to certain triggers [Bing. 2008]. As this study aims to provide proactive solutions for improved handover prediction, it focuses mainly on the enhancement of automatic scanning procedures. WLAN scanning is further classified as active or passive based on how the MN acquires the Beacon frames from neighbouring APs.

Passive scanning

In order to join a particular extended service set (ESS) through passive scanning, a MN tunes in to different channels consecutively and listens for Beacon frames that match the Service Set ID (SSID) of the ESS. The MN may sometimes listen to each channel for up to a maximum duration (e.g. 100ms

[Bing. 2008]) in order to ensure the receipt of a Beacon. Passive scanning can thus be more time consuming as a MN may end up spending the maximum duration in almost each channel.

Active scanning

In active scanning a MN explicitly request a Beacon frame on each channel through the transmission of probe request packets. It then listens to the channel for a Beacon from the AP (e.g. 20ms [Bing. 2008]) and moves on to scan another channel as soon as it receives a reply. The MN then creates a list of all available APs at the end of the scan procedure.

Next scan event delay

In order to ensure stability in the choice of an AP the current practice is to perform a number of iterations of scan and then select the AP with the strongest overall RSS which occupies the highest position in the candidate list for the longest duration. It is thus important to establish a certain delay between two iterations. The MN waits for this time period and then resumes scanning again.

Active Set Selection

Some of the popular criteria currently used for selecting an AP for attachment are as follows:

- An AP with the strongest filtered RSS value above -75 dBm. This AP may not have been the best candidate for all iterations but nevertheless will be selected if it has a rapidly increasing RSS value.
- The AP that remains as a valid choice for the longest duration.
- A weighted combination of the above two criteria.

In order to prevent unnecessary vertical handovers this research study refines the procedure of AP selection by applying the new metric TBVH. This is combined with the knowledge of RSS and a requesting application's required resource duration to select the best AP.

2.2.1.4 QoS management

The IEEE 802.11 Medium Access Control (MAC) layer has two defined modes of operation for the provisioning of basic QoS – the Distributed Coordination Function (DCF) for contention-based

Table 2.1 Traffic categories in IEEE 802.11e [Ganz, et al. 2004]

User Priority	Traffic Category	Traffic Type
1	TC1	Background
2	TC2	Spare
0	TC0	Best Effort
3	TC3	Excellent Effort
4	TC4	Controlled Load
5	TC5	Video < 100ms delay + jitter
6	TC6	Voice < 10ms delay + jitter
7	TC7	Network control

medium access and the Point Coordination Function (PCF) for contention-free medium access. Initially, these two QoS [Ganz, et al. 2004] mechanisms employed simple procedures for QoS provisioning. There was no classification of traffic which meant that all packets in the same MN were treated equally. Packet scheduling for transmission at the MN was also carried out in a simple First-In-First-Out (FIFO) manner. In order to resolve these problems, the 802.11 standards committee proposed the IEEE 802.11e extension designed specifically to provide MAC enhancements for improved QoS provisioning.

2.2.1.5 Traffic classification

Traffic in IEEE 80.11e is classified based on traffic category and traffic stream [Ganz, et al. 2004].

- **Traffic category**: this mechanism assigns a distinct priority to packets belonging a certain application type. Categories range in an increasing priority order from 0 to 7 as shown in table 2.1.
- **Traffic streams**: after being classified into traffic categories based on the type of application, each category can be further classified into 8 traffic streams. This facilitates the refined specification of quantitative measures for each stream such as packet size, arrival rates, maximum delay, etc.

2.2.1.6 Polled channel access

Call admission control is an inherent feature of the polled channel access mechanism in the 802.11 wireless standard [Bing. 2008]. A MN wanting to set up a traffic stream connection sends the request to the AP which then performs the resource availability check. If the request can be fulfilled, the AP

responds back with a service period schedule based on submitted parameters like data rate, frame size, delay bound and maximum service interval.

2.2.1.7 Contention-based channel access

In contention based channel access [Bing. 2008], admission control is an optional feature where the AP decides when MNs should transmit based on their access category parameters. Admission control is not imposed on all access categories. The AP advertises the access categories that fall under admission control. A MN that needs to use that access category must first send a request to the AP with a list of parameters similar to the ones in polled channel access. If the request is accepted, the AP allocates a channel time for the MN for uplink transmission based on the submitted parameters. If the request is rejected, a MN may still transmit but with a lower priority access category. It is thus required that there be at least one access category that does not employ admission control. A key enhancement to the WLAN based call admission control (CAC) scheme is proposed by this study which actually allows a MN to negotiate for increased resources with the AP, enabling a stream transfer to complete while the device remains connected to WLAN, hence avoiding an upward vertical handover.

2.2.2   UMTS

The Universal Mobile Telecommunications Service (UMTS) standard [Smith, et al. 2007, Wisely. 2009] was developed by the Third-Generation Partnership Project (3GPP). It evolved from the Global System for Mobile (GSM) communications to provide support for enhanced features such as high data rates and packet data transfer capabilities. The standard was designed for world-wide use and supports compatibility for both packet-switched and circuit-switched data transmission.

2.2.2.1 Salient Features

- **Physical Layer:** The radio access consists of Wide band Code Division Multiple Access (WCDMA). WCDMA supports both Frequency Divisoin Duplex (FDD) and Time Division Duplex (TDD) modes of operation.

- **Data rates:** UMTS supports different data rates, 144 kbps for satellite and rural outdoor environments, 384 kbps for urban outdoor and 2048 kbps for indoor and low range outdoor environments.

- **Radio Link Control:** The radio link control (RLC) part of the data link layer takes care of issues such as acknowledged and unacknowledged data transfer, transparency, QoS settings, error notification, and the establishment of RLC connections.

- **Security:** UMTS incorporates improved security features in the form of five security groups:

  - Network access security provides secure access to users to 3G services. It is designed to protect attacks on the radio access link.

  - Network domain security features take care of security in the core network and protect the network against attacked from the wired interface.

  - User domain security features consist of mechanisms that enable secure access to MNs.

  - Application domain security features enable the secure exchange of messages between the user and provider domains.

  - Visibility and configurability security allows for the configuration of security features by a user on the device.

2.2.2.2 The UMTS architecture

The UMTS network architecture [Ganz, et al. 2004] consists of three entities as shown in figure 2.3.
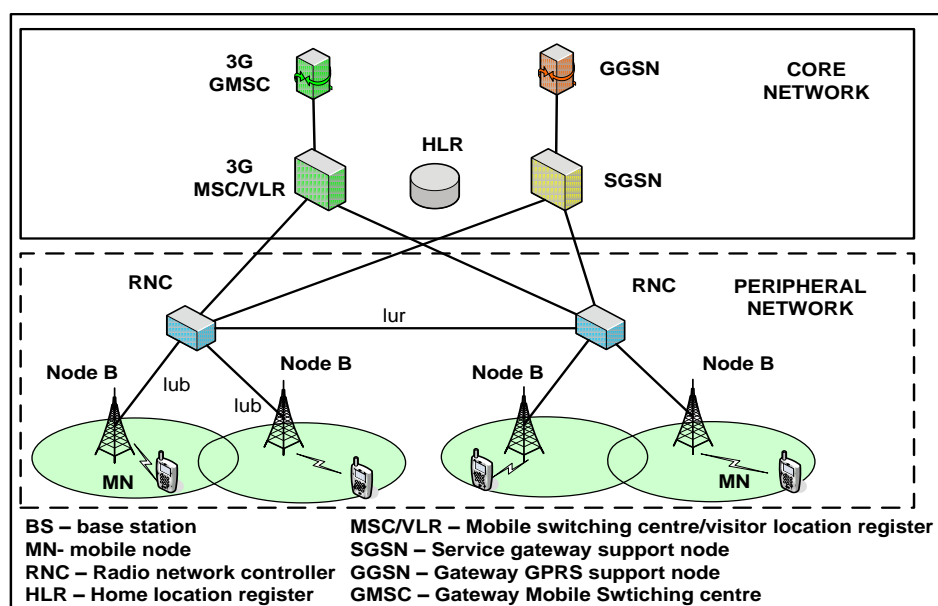
Figure 2.3 The UMTS network infrastructure [Ganz, et al. 2004]

42

They are the mobile node (MN), UMTS Terrestrial Radio Access Network (UTRAN), and the Core Network. The MN and Node B (base station).communicate using the wireless radio interface. A group of Node Bs is controlled by the Radio Network Controller (RNC). RNCs communicate with each other through the lur interface which transfers signalling and user data. Node Bs and RNCs together form the UTRAN. An RNC is connected to the core network through the lu interface which supports both data and voice services. An internal connection between the RNC and Node B for the transfer of information is called the lub interface.

2.2.2.3 Radio Resource Control (RRC)

The radio resource control layer [Ganz, et al. 2004] exists in both the MN and Node B and executes network layer signalling. The main functions of the RRC are as follows:

- Call admission control where the RRC selects the parameters describing the radio channel based on information from the higher layers and allocates the required resources between the MN and Node B with the desired QoS.
- Execution of both handover decision and execution functions including preparation for handovers, cell reselection and update procedures.
- Controlling measurement decisions for parameters both on the device and network sides.
- Execution of functions such as encryption, power control and integrity protection.

2.2.2.4 QoS Management

QoS management in UMTS consists of mapping application requirements on to offered UMTS services on both uplink and downlink channels. The standard defines four QoS classes:

- **Conversational**: This class is meant for applications like telephony, Voice Over IP (VOIP), and video conferencing. As this consists of real-time traffic, delay limits are very strict in order to avoid low call quality.
- **Streaming**: This class contains real-time applications such as real-time video and audio streams.
- **Interactive**: This class is used by applications such as web browsing that request data from remote servers. It is characterised by the request-response time delay pattern of the end user.

- **Background**: This traffic consists of data transfer applications such as file transfer and email. These applications are insensitive to delay.

End-to-end QoS delivered to the user can be specified by Service Level Agreement (SLA) between domains, parts of the network and operators.

2.2.3   IEEE 802.16e WiMAX

The Worldwide Interoperability for Microwave Access (WiMAX) standard is an emerging broadband wireless standard that has gained immense popularity due to its promise of mega-bit speeds. Developed by the IEEE 802.16 group, several types of WiMAX solutions have been certified by the WiMAX Forum in the past few years. This sub-section focuses on the IEEE 802.16e-2005 version called mobile WiMAX due to its support for mobile applications and provides a summary of the important features as specified in the standard document release 2005.

2.2.3.1 Salient Features

Mobile WiMAX offers the following set of features designed to provide high-speed broadband wireless access, flexibility and multi-user diversity:

- **Orthogonal Frequency Division Multiple Access (OFDMA):** The Physical Layer is based on Orthogonal Frequency Division Multiplexing (OFDM) which facilitates operation in non line of sight conditions by offering resistance to multipath effects.
- **High data rate support:** WiMAX employs the 64-QAM modulation techniques with rate 5/6 error-correction coding to achieve peak data rates as high as 74Mbps when operating in the 20MHz spectrum. In the 10MHz spectrum, using a Time Division Duplex (TDD) scheme with a 3:1 downlink-to-uplink ratio, peak data rates of 25Mbps downlink and 6.7Mbps uplink can be achieved. Service providers can achieve even higher data rates of up to 350Mbps by combining several channels for a single transmission.
- **Support for advanced antenna techniques:** WiMAX provides support for the attachment of multiple antennas which allows the implementation of advanced schemes such as beamforming, space-time coding and spatial multiplexing to improve system capacity and spectral efficiency.
- **Robust security support:** The standard uses the Advanced Encryption Standard (AES) which support strong encryption. The authentication architecture is based on the flexible Extensible
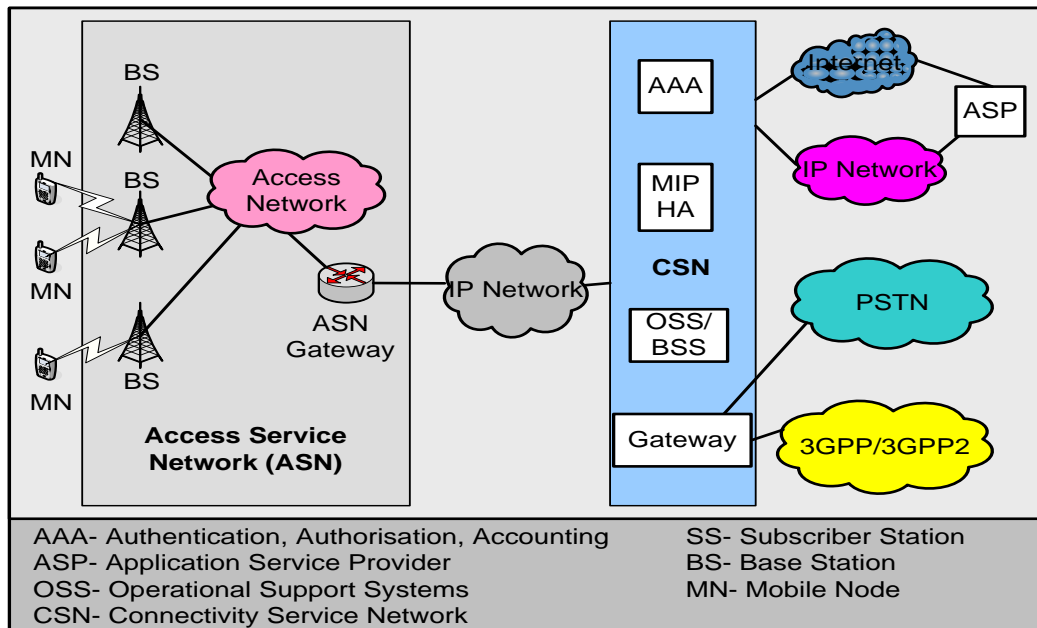
Figure 2.4 The Wimax Reference Architecture [Ganz, et al. 2004]

- Authentication Protocol (EAP) which offers a variety of user credentials including username/password, digital certificates and smart cards.

2.2.3.2 The WiMAX Architecture

Figure 2.4 illustrates the WiMAX reference model proposed by the WiMAX Forum's Network Working Group (NWG). It is based on an all-IP platform with all end-to-end services employing IP-based protocols for mobility management, QoS provisioning, end-to-end transport and session management. The architecture consists of three logical components:

1. The Mobile Node (MN) which form the edge of the wireless access network.

2. The Access Service Network (ASN) which consists of base stations and ASN gateways.

3. The IP-based core Connectivity Service Network (CSN) that provides all IP-based core functions.

This logical split in the network architecture facilitates different entities to be governed by different business systems [Ganz, et al. 2004]. Some functional entities and interfaces defined by the WiMAX Forum NWG and which are important for this study are as follows:

- **Mobile Nodes (MN)**: These are stationary or nomadic devices equipped with WiMAX interface cards through which they can connect to the ASN network.

- **Base Stations (BS)**: They provide air interfaces to the MNs and are directly involved in micro mobility management functions involving handovers such as handover triggering and tunnel establishment, radio resource management, QoS enforcement, security and session management.

- **Access Service Network Gateway (ASN Gateway)**: This entity is a layer 2 traffic aggregation point in the ASN. It also acts as a higher level point in the network management hierarchy for key decisions such as radio resource management, admission control, mobility management, authentication and QoS and policy enforcement and connectivity with the CSN.

- **Connectivity Service Network (CSN)**: This is the core network that provides connectivity to the Internet, ASP, public networks and corporate networks. This entity is owned by the network service provider and includes the AAA servers needed for the authentication of devices and connections. It acts as a highest logical entity in the reference architecture and provides services for IP address management, roaming between different ASNs, location management, and interworking with other access networks like PSTN.

2.2.3.3 Quality of Service Management

In the WiMAX network the BS is responsible for controlling all uplink and downlink connections. A unidirectional MAC layer connection with a unique connection identifier (CID) is established between the BS and MN before data transmission. This CID is used as a temporary address for data transmission over the link. WiMAX QoS also consists of service flows which are unidirectional flows of packets with a particular set of QoS parameters. Each flow is identified by a service flow identifier (SFID). These service flows can also be mapped on to DiffServ to enable IP-based QoS. Applications in the WiMAX network are scheduled by the BS MAC scheduler. These services in the order of decreasing priority include Unsolicited grant service (UGS), Real-time Polling service (rtPS), Non-real-time Polling service (nrtPS), Best-effort service (BE) and Extended real-time Polling service (ErtPS).

2.3   Mobility management for seamless roaming

Mobile IP [Johnson, et al. 2004] has emerged as one of the strong candidates for the provisioning of seamless macro-mobility support to a mobile node (MN) roaming among different network domains in a 4G heterogeneous environment. It targets one of the main challenges in mobility – the change of IP
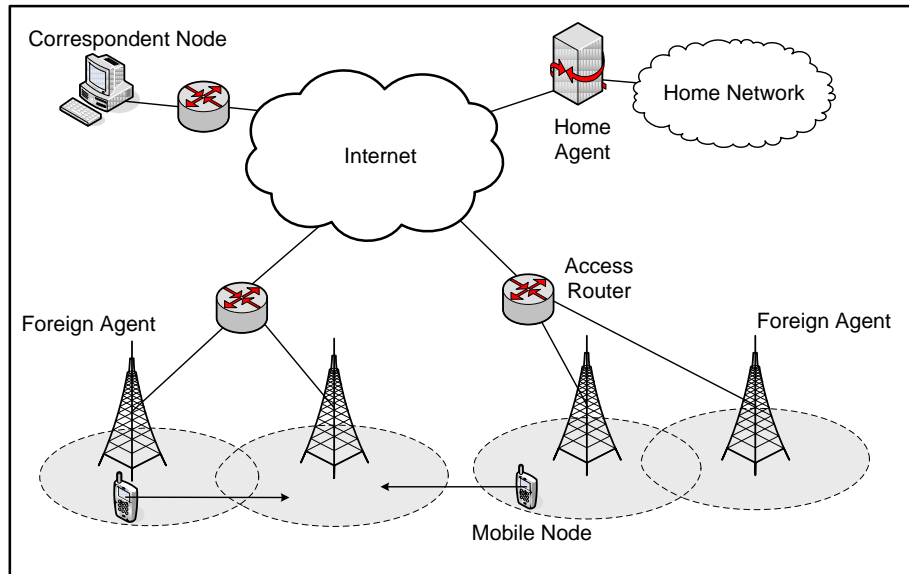
Figure 2.5 IP based mobility management in wireless networks

address when a MN changes its point of attachment to the Internet which can cause a disruption to ongoing transport connections.

Figure 2.5 shows an IP-based mobility management scenario. In this protocol the MN is able to retain its address and existing connection while communicating with a correspondent node (CN) on the Internet even if it gets attached to a new network. In basic Mobile IP implementation, every MN belongs to a home network (HN) which assigns it an IP address. When the device moves outside its HN and gets connected to a foreign network (FN), it is assigned a temporary IP address called the Care of Address (CoA) and the home agent (HA) in the HN is updated accordingly through a binding update. Any packets arriving from the CN for the MN first go to the HA which then implements IP tunnelling and encapsulates the existing packet in a new header with the CoA as the destination field and then forwards it to the foreign agent (FA). Enhancements in Mobile IPv6 [Shenoy. 2005] include the ability of a MN to configure its own CoA eliminating the need for FAs, and the direct transmission of binding updates to the CN by the MN without going through the HA.

Micro-mobility solutions were developed to reduce the overhead associated with the frequent transfer of address information [Banerjee, et al. 2004]. In this case the HA was not alerted of the change in the MN's location as long as it remained within the same domain, no matter how many times it changed its point of attachment. Hierarchical mobile IPv6 [Campbell, et al. 2000] is an example of micro-mobility solution where each MN was assigned two CoA: a regional CoA and a link CoA
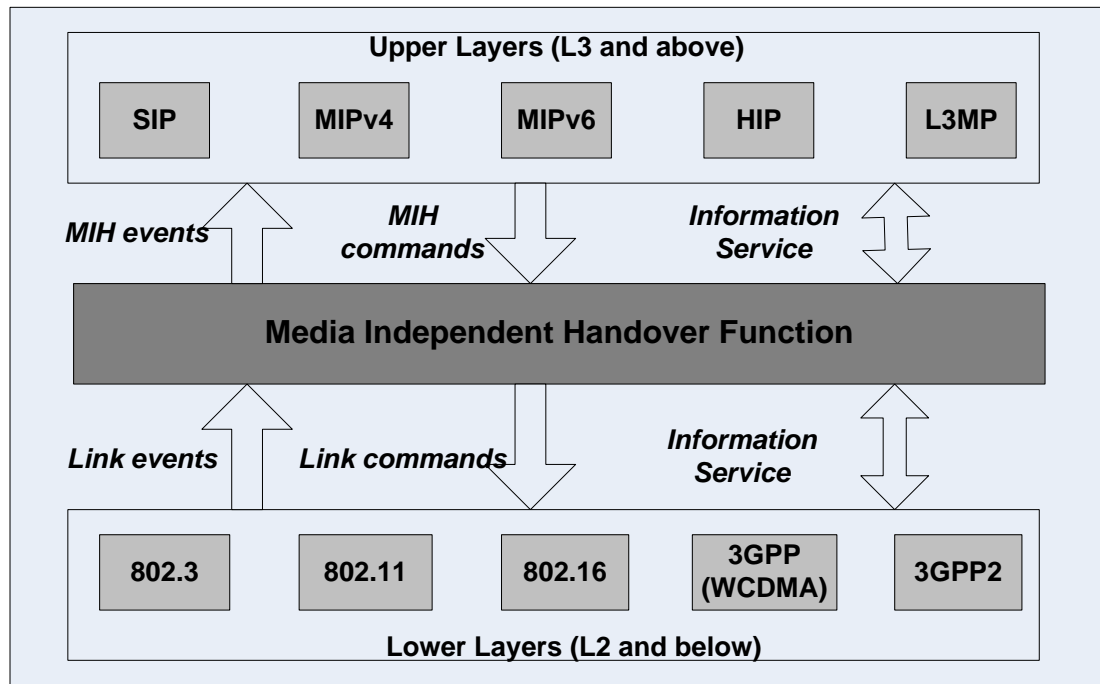
Figure 2.6 Information flow between MIH components [Dutta et al, 2005]

[Schiller. 2003]. Only the link CoA changed with handovers. The regional CoA did not change as long as the MN remained connected to the same domain.

This research study does not delve further into the issue of mobility management and concerns itself mainly to the phases before and after a vertical handover.

2.4   The IEEE 802.21 Framework

The IEEE 802.21 WG [Dutta et al, 2005] has proposed a client-based framework that aims to offer different services for the optimisation of vertical handovers between different networks such as WLAN, WiMAX, UMTS and GPRS while interfacing with policy and higher layer mobility protocols. The various factors affecting handovers decisions are service continuity, application class, quality of service, network discovery and selection, security, power management and handover policy. The 802.21 framework employs information gathered from both the MN and network to choose the most appropriate network. The main component of the 802.21 framework is the Media Independent Handover Function (MIHF) shown in figure 2.6 which acts as a unified interface that exposes network independent service primitives to higher layers. It communicates with higher and lower layers through Service Access Points (SAPs) and offers three different services:

- **Media Independent Event Service (MIES):** This event service reports local events (events taking place in a client) and remote events (events taking place in network elements) to upper layers. The MIES service provides a reactive service in which a client registers for the notification of certain events which the MIES reports in the form of triggers. Some examples of events are "Link Up", "Link Down", "Link Parameters Change", "Link Going Down", and "L2 Handover Imminent".

- **Media Independent Command Service (MICS):** The MICS commands are employed by the higher layers to find out the status of connected links and for the execution of higher layer mobility and connectivity decisions to lower layers. Some examples of commands are MIH Poll, MIH Scan, MIH Configure, and MIH Switch.

- **Media Independent Information Service (MIIS):** This service provides a set of mechanisms that enable the MIHF to obtain both static and dynamic network information within a geographical area. Static information can be the names and providers of neighbouring networks, and dynamic information can be parameters such as channel information, MAC addresses, security information and other higher layer service information.

As network detection policies provided by the MIHF function are mainly reactive in nature, they do not possess the ability to deduct further information on the future conditions of available networks. Triggers offered are also primitive in nature and do not accommodate refined knowledge about device and network contexts such as device location, network coverage and geographical topologies. This introduces the need for the development of new proactive mechanisms that consider a wider variety of context information in decision making.

With seamless roaming a distinguished feature of 4G networking, devices must possess the key capability to dynamically adjust their behavioural characteristics to the environment in which they are located. Hence location prediction techniques play a key role in improving device context awareness leading to the successful exploitation of location-based network services. With the advent of pervasive computing, the last decade witnessed an exponential increase in the demand for location-aware applications and devices for both indoor and outdoor environments. The field of location prediction has been the subject of intense research activity and a wide range of innovative tracking solutions

have been introduced in the commercial market in the past five years [Bing. 2008]. Let us then move on to explore available solutions in the area of indoor and outdoor location positioning.

2.5   Location Tracking in Wireless Networks

Location awareness or localisation forms one of the main dimensions of context [Muthukrishnan, et al. 2005]. It plays an indispensible role in the ubiquitous delivery of uniform services to MNs roaming among different networks by acting as a bridge between the virtual and physical world [Estrin, et al. 2002]. This section provides an overview of the different location tracking solutions currently available in the market. It adopts the popular taxonomy based on the target environment, i.e. indoor and outdoor localisation solutions, and discusses noteworthy solutions developed in both areas.

A number of studies in the recent past have demonstrated how the knowledge of location helps to improve performance in wireless networks [Cottingham, et al. 2005, Bahl, et al. 2000, Fulop, et al. 2008, Akyildiz, et al. 2002]. Localisation plays a pivotal role in this research study which employs both indoor and outdoor positioning techniques for determining the positions of devices in a given network coverage.

2.5.1   Outdoor Localisation Techniques

Over the years, satellite-based Global Positioning System (GPS) [GPS.gov website] has emerged as the dominant location prediction technology for outdoor environments. The US Defence Department estimates the accuracy of civilian GPS units to be better than 15 m with a precision of 95%. GALILEO, its European counterpart aims to provide an accuracy of less than 1 m [Hinch. 2004]. Increasing user demands for location awareness have resulted in continued enhancements to the GPS standard, with some solutions even reaching an accuracy of up to 1 cm [VBOX website]. However, as GPS accuracy can suffer due to issues such as reduced line-of-sight with satellites due bad weather conditions or the unavailability of enough base stations, alternative network-based location prediction techniques have been proposed such as Observed Time Difference of Arrival and Direction or Angle of Arrival. Researchers continue to propose other sophisticated mobile-based and network-based solutions such as track-before-detect schemes [Algeier, et al. 2008] and Bayesian Filtering Formulations [Khalaf-Allah. 2008] with positioning accuracy reaching up to 1m.

This study adopts GPS as the outdoor localisation technique due to the wide availability of GPS functionality on mobile phones. The location estimation accuracy of standard GPS which ranges between 5-15 m [Muthukrishnan, et al. 2005, Hinch. 2004] is acceptable for the prediction of handovers in large coverage cells such as UMTS as the large handover threshold circle radius results in less sharply defined coverage boundaries.

2.5.2 Indoor Localisation Techniques

Indoor positioning of physical objects is an important issue in the development of context-aware applications in the ubiquitous computing environment [Fukuju, et al. 2003]. Much research work has focused on the development of indoor positioning models through the use of different wireless access technologies.

The Active Bat location system [Harter, et al. 2002] used a combination of RF and ultrasound time-of-flight to estimate distance based on which the position is determined. It achieved an accuracy of up to 9 cm. Zhou et al. [Junyang Zhou, et al. 2008] designed an indoor location scheme that utilized information on both GSM and WLAN signals to estimate the location of a device. The system was stable and reached centimetre-level accuracy. The low-cost indoor positioning system proposed by Randell et al. [Randell, et al. 2001] utilised a combination of radio frequency and ultrasonics to detect the position of an indoor device. The observed accuracy was 25 cm for worst case scenarios. The SmartLOCUS positioning system [Brignone, et al. 2003] developed by HP Labs synchronised radio frequency and ultrasound differential time-of-flight measurements to create a self-organising coordinate system offering an accuracy of 2-15 cm. Ekahau developed a software-based probabilistic positioning framework [Bing. 2008] based on Bayes theorem which considered a realistic scenario where the measured signals are inherently noisy. The system achieved an accuracy of 1-3 m.

This research study proposes adopting the Ekahau positioning system as the localisation solution for the indoor environment as it is a WiFi-based system which can be deployed without making any changes to the existing network infrastructure [Bing. 2008]. The setup involves calibration and access point configuration phases which coincide with the study's own network configuration phase. Thus it is both feasible and easy to setup a location environment that complements the proposed solutions.

2.6 The Y-COMM Framework for Heterogeneous Networking

This chapter so far discussed technological developments in different areas in the field of communications. Heterogeneous networking can be viewed as the culmination of these technological developments and in order to successfully derive maximum benefit from these new features, there is a growing need for the development of a structural framework that establishes clear links among these different developments. The successful implementation of seamless roaming requires a close co-existence of different network standards like UMTS, WIMAX and WLAN at both the network and device levels. In order to achieve seamless interoperability, components of the 4G protocol stack will exhibit more complex functionality than components of the normal OSI protocol stack due to the additional tasks they will need to support [Shaikh, et al. 2007]. Some of the key requirements include reconfigurability, QoS management, and policy management for vertical handovers [Mapp, et al. 2006]. The development of multi-interfaced devices has resulted in an expansion at the Medium Access Control and Physical layers, along with a convergence at the Network layer through the adaptation of IP [Mapp, et al. 2006]. Heterogeneous networking also requires mobile clients to liaise closely with base stations and adopt more superior decision-making roles in order to ensure smooth vertical handovers.

Technological developments in both wired and wireless networking have resulted in core networks evolving into high speed, low latency wired networks, and peripheral networks increasingly adopting the wireless paradigm. Both sides thus display largely different characteristics in terms of bandwidth, latency, availability and error distribution properties [Mapp, et al. 2006]. It is thus necessary to address the split between the peripheral and core sides, and between users and services [Crowcroft, et al. 2007]. These factors, including others have given rise to the demand for the development of a new architectural framework that encapsulates the key challenges of heterogeneous networking and which acts as a guide in the development of its products.

The Y-Communication (Y-COMM) framework [Crowcroft, et al. 2007, Mapp, et al. 2009] was proposed jointly by researchers at the Networking Research Group at Middlesex University, Computer Laboratory at Cambridge University, Samsung Research and Deutsche Telekom. This architectural framework aims to address the new challenges faced at the network, device and application levels as
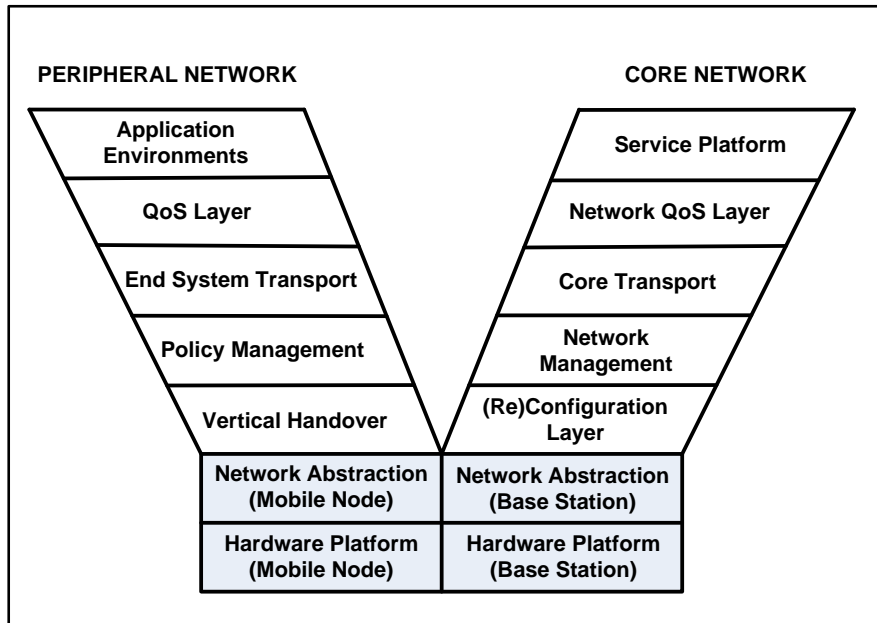
Figure 2.7 The YCOMM architectural framework [Crowcroft, et al. 2007]

a result of heterogeneous networking. It adopts a layered approach and acts as a reference model similar to the OSI reference model [Mapp, et al. 2009]. It consists of two frameworks – peripheral and core, as shown in figure 2.7. The Peripheral framework targets the specification in heterogeneous clients, and the Core framework targets services and infrastructure design in core networks. This section summarises the main features of the YCOMM framework. Taking the UMTS network architecture shown in figure 2.3 as an example, it also provides an explanation on how the framework can map on to existing network components.
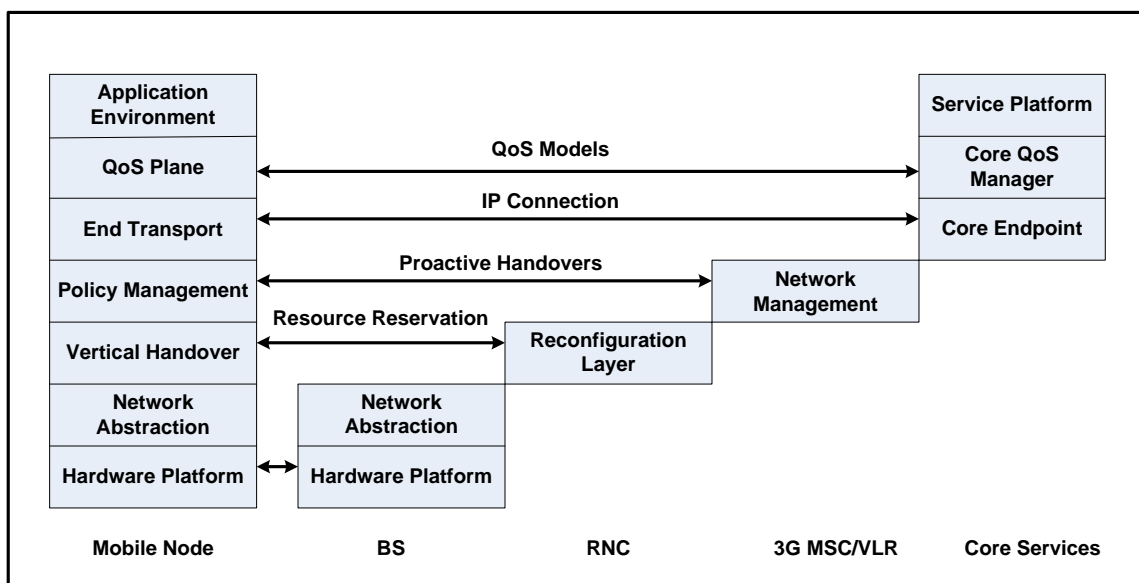
Figure 2.8 Mapping YCOMM on to UMTS network components [Mapp, et al. 2009]

2.6.1 The Peripheral Framework

The peripheral side framework stack shown in figure 2.8 resides in the MN. The MN may be connected simultaneously to a set of different networks.

2.6.1.1 Hardware Platform Layer (HPL)

This layer defines the hardware components and technologies for supporting different wireless networks. It is responsible for defining different characteristics such as electromagnetic spectrum, modulation techniques and MAC algorithms for acquiring and reserving channels. It is further divided into vertical sub-layers, one for each wireless standard supported by the device.

2.6.1.2 Network Abstraction Layer (NAL)

The main function of this layer is to provide a common interface for different network technologies. It is also responsible for maintaining and controlling the network on a MN. The IEEE 802.21 standard discussed in section 2.4 fits in to the network abstraction layer.

2.6.1.3 Vertical Handover Layer (VHL)

This layer is responsible for the execution of vertical handovers and is able to accommodate both network-controlled and client-controlled handovers. Its responsibilities include acquiring resources for handover, managing signalling and context transfer during vertical handovers.

2.6.1.4 Policy Management Layer (PML)

The Policy Management Layer evaluates all circumstances when a handover should occur. It takes into consideration various factors such as signal strength, channel network and transport conditions and coverage for decision-making. It supports both reactive and proactive mechanisms. Location information supplied by positioning techniques discussed in section 2.5 forms an input to this layer from lower layers and is utilised for the derivation of new context parameters.

2.6.1.5 End Transport Layer (ETL)

This layer combines the function of both network and transport layers of the OSI model. It thus examines the addressing, routing and transport issues in peripheral networks. It is also responsible for making end-to-end connections across the core network.

2.6.1.6 Quality of Service Plane (QoS-P)

The QoS Plane layer is responsible for the management of upward and downward QoS in heterogeneous clients. Downward QoS involves an application specifying its QoS requirements to the system and the system running mechanisms to maintain this QoS at all times over different varying network channels in a transparent manner. Upward QoS involves the development of intelligent applications which try to adapt to changing QoS. The QoS Plane also acts as an information repository, monitoring active channels for changes in QoS to ensure stable operation.

2.6.1.7 Applications Environment (APP-Env)

The intention here is to adopt a toolkit approach which aids in the development of application environments where there is a direct vertical interaction between applications and lower layers. The layer specifies a set of objects, functions and routines to build applications which are capable of interacting directly with the YCOMM layers and benefiting from their new features.

2.6.2 The Core Framework

The core framework's protocol stack is distributed across the different core components in the network as shown in figure 2.8.

2.6.2.1 Hardware Platform Layer (HPL)

This layer in the core framework runs hardware for a specific wireless access technology designed for the base station side. In UMTS, this layer runs in the BS.

2.6.2.2 Network Abstraction Layer (NAL)

This layer focuses on controlling the functions of the base station belonging to a specific wireless standard. This layer also runs in the BS in UMTS.

2.6.2.3 Reconfiguration Layer (REL)

This layer controls key infrastructure such as routers, switches and other components using programmable networking techniques. It runs inside the RNC in the UMTS network and employs programmable techniques to configure individual RNCs.

2.6.2.4 Network Management Layer (NML)

The function of this layer is to act as a management plane for controlling networking operations in the core network. It manages a number of networks in an integrated fashion and can also gather information on peripheral networks. This can then be forwarded to the Policy Management layer on the client side. This layer runs inside the MSC and manages all the RNCs in a local coverage area. These RNCs may belong to different access networks. It knows the status of each wireless network and its topology and shares the information with the Policy Management Layer at the client side.

2.6.2.5 Core Transport Layer (CTL)

This layer is responsible for the transport of data through the core network with a given QoS and specified level of security. When IP packets are tunnelled through the core network, this layer negotiates the required resources between two core end points.

2.6.2.6 Network QoS Layer (NQL)

This layer takes care of the QoS issues within the core network. A special concern is to take care of QoS issues at the interface between the core and peripheral networks by preventing overloading. The negotiation of the required level of QoS for a new connection by the MN takes places between the QoS Plane in the MN and the Network QoS Layer at the core side. The Network QoS Layer then returns two core end points to establish the new connection.

2.6.2.7 Service Platform Layer (SPL)

This layer allows for the installation of special services on various infrastructure components in the core network. Services can be installed simultaneously on different networks.

2.6.3 Role of proposed solutions in YCOMM

This research study makes two important contributions to the peripheral side of the YCOMM framework at the Policy Management Layer and the QoS Management Plane.

2.6.3.1 Contributions at the Policy Management Layer [Mapp, et al. 2009]

As the policy management layer mainly evaluates the circumstances when a handover should occur, the proposed TBVH models for determining the time before vertical handover reside in this layer. The MN actively communicates with the Network Management layer in the core framework to determine context parameters such as network topologies and QoS for each network. This information is combined with the MN's context information to determine if and when a vertical handover will take place. Once the estimate for TBVH is derived, it is passed on to other layers such as the QoS Plane and Vertical Handover layer so that they start preparing for the vertical handover by proactively negotiating for resources with the core network. This facilitates seamless vertical handovers. The generic message exchange sequence between the MN and core network is shown in figure 2.9. A more detailed explanation on this topic is covered in Chapter 4.
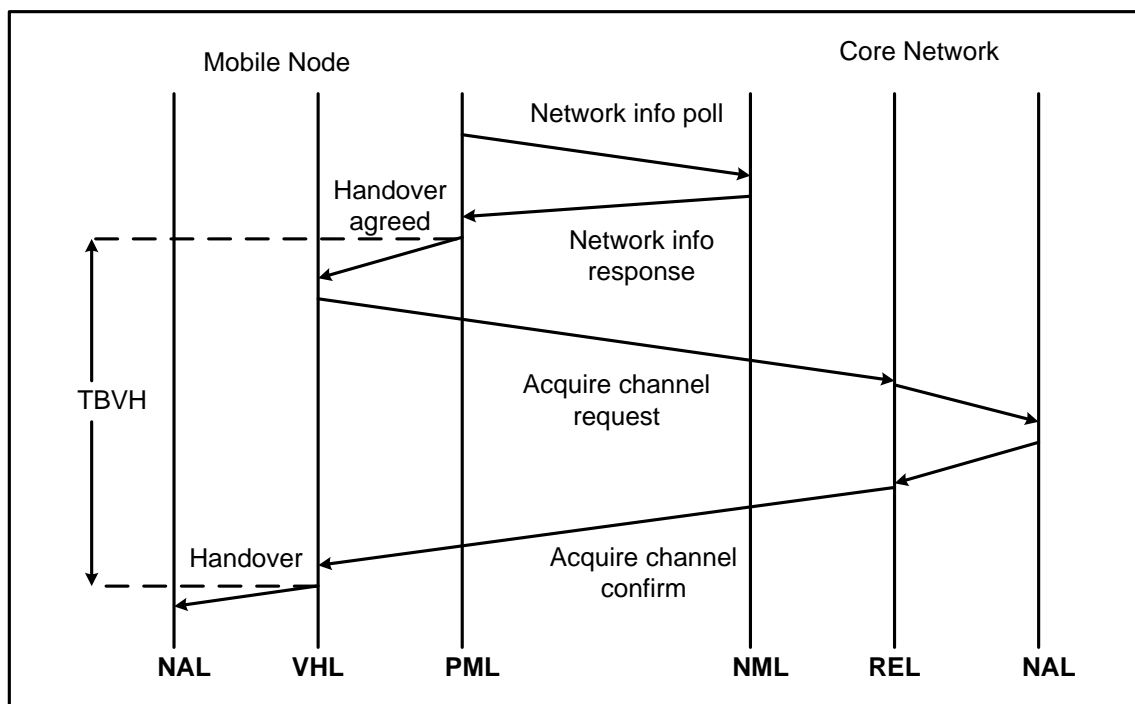
Figure 2.9 MN-CN Message Exchange in preparation for a vertical handover [Mapp, et al. 2009]

2.6.3.2 Contributions at the QoS Plane

The proposed Stream Bundle Management layer for handling downward QoS resides in the QoS Plane of the MN. This layer collects context information from the network, client and application domains to make intelligent choices in network selection and QoS management. This is one layer that actively applies information on TBVH to choose a stable network interface for a requesting application by avoiding unnecessary vertical handovers. A detailed explanation of this topic is covered in Chapter 7 of this thesis.

2.7   Chapter Summary

This chapter provided an overview of technical concepts in the field of wireless communications which will facilitate in understanding the design and functioning of new solutions proposed in subsequent chapters. It also introduced the reader to some of the latest technological developments which have directly influenced this research work. This completes the first introductory part of the thesis. The next three chapters discuss existing challenges in vertical handover prediction, the proposed TBVH-based vertical handover optimisation concepts, theory and results.

CHAPTER 3

Related Work – Handover Prediction Techniques

3.1   Introduction

The main purpose of this chapter is to provide a comprehensive literature review of related work conducted in the area of handover optimisation through improved prediction of vertical handovers. It highlights noteworthy achievements of some previous research works and identifies unresolved issues and gaps in knowledge which this dissertation aims to fulfil.

Due to the ever existent demand for seamless mobility, handover optimisation has always been an area of immense research activity, and is likely to remain so in the future. While conducting the literature review, the following three popular categories emerged in handover prediction techniques:

1. History based prediction
2. Coverage based prediction
3. Mathematical modelling based prediction.

3.2.   History-based handover prediction

In literature, a number of studies adopted the history-based approach for improved context awareness in handover prediction mechanisms. The main assumption of this approach was that a MN's movement patterns in the future were mostly likely to be similar to past patterns and thus could be predicted from stored sequences of data. The main drawback of this approach was that it relied heavily on large amounts of stored historical sequences. The large availability of information was often insufficient to fully capture the random behaviour of MNs, e.g. pedestrian behaviour. Path prediction failed as soon as the MN strayed away from the predetermined route. This section analyses some interesting history-based handover prediction schemes which proposed some novel ideas aimed to overcome these problems and improve prediction during handovers.

Liu and colleagues [Liu, et al. 1996] proposed a set of mobile motion prediction algorithms designed to predict the future location of a mobile user according to the user's movement history. The

assumption was that users had a degree of regularity in their movements which could be recorded to predict their future movement patterns. The cellular system was divided into service areas, each area forming a set being represented by a state variable. Based on the MN's resident time, a state was classified as stationary, transitional or a boundary state if located at the boundary of the service area. MN movements were modelled by a discrete-parameter and discrete-state stochastic process consisting of states. This study recognised the importance of coverage boundaries which contributes successfully towards improved context awareness. In an extension to this study, Stathes and Merakos [Hadjiefthymiades, et al. 2003] applied the concept of states to propose a path prediction algorithm based on learning automaton. The shortcoming of both these approaches was that they were limited to predicting the state of a MN within a service area and did not capture or predict its actual trajectory or refined movements within that service area.

[Navidi, et al. 2004] and [Liang, et al. 2003] proposed predictive distance-based mobility management schemes that attempted to predict the future location of a MN. While Liang based this on the Probability Density Function (PDF) of the MN's location given by a Gauss-Markov model, Navidi did not assume any specific mobility model and suggested a history-based approach for location prediction based on previous reported locations. Both these approaches utilised context information mainly for reducing the paging overhead between a MN and BS and did not address traffic management issues. While the TBVH based handover prediction model proposed in this thesis is not history-based, it is slightly similar to Navidi's model in the sense that it does not make any assumptions about mobility patterns like Liang's approach and dynamically calculates the TBVH components based on real-time measurements of different parameters.

A recent study, [Michaelis, et al. 2008] proposed improved handover prediction for high priority users like rescue teams in very densely populated areas such as stadiums. The mechanism mainly relied on large amounts of stored historical sequences for the development of a variety of mobility models which aimed to realistically model the behaviour of the MN in different situations like vehicular, pedestrian and group-dependent scenarios. Acknowledging the computational and memory demands incurred due to training and storing data patterns for each user, mobility prediction was limited to high priority users only. However all this information was simply used to predict the next cell to which the MN was expected to perform a handover. Realising the importance of knowing the time until the next

handover for smooth mobility management, the authors proposed calculating it as the exponentially smoothed mean based on all stored MN residence times. The problem with this technique was even without considering random MN movement patterns, mean residence times could vary largely based on MN speeds.

Another study [Fulop, et al. 2008] applied the Markovian modelling approach to improve the predictability of the random walk model with the user moving between different states. Cell dwell time was history based and the approach did not capture the anomalies in handover prediction arising due to topological factors.

3.3. Coverage based handover prediction

Schemes from this handover prediction category employed previously stored knowledge about network coverage to predict the duration and quality of future coverage for a MN. The drawback of this approach is that it involves the storage of large amounts of coverage information which automatically introduces the increased overhead associated with the data gathering phase during the construction of coverage databases. In order to remain functional the databases require the frequent updating of coverage data which by no means is a simple task when considering large coverage areas like cities.

Soh et al. [Soh, et al. 2004] proposed a proactive mobility prediction technique that applied both MN positioning information and road topology knowledge to predict the time a MN had before performing a horizontal handover. The study demonstrated how the knowledge of time before handover (TBH) helped in improving the resource reservation efficiency and network performance. However, as this approach relied mainly on large volumes of data on road maps stored in prediction databases inside every BS, it was unable to predict the path and TBH of a MN when it strayed away from the road topology stored in the database. Thus the accuracy of TBH prediction decreased considerably when the MN exhibited random mobility patterns.

Balasubramanium and Indulska [Balasubramaniam, et al. 2004] evaluated a rich set of context information during the vertical handover decision process and acknowledged the important role device location played in the choice of the correct interface for handover and QoS mapping. However

the approach relied heavily on coverage maps for predicting network coverage and device positioning and therefore suffered the problems common to this category.

Predicting the important role location and context/situation knowledge would play in the efficient integration of different technologies E. Cianca and colleagues [Cianca, et al. 2006] proposed a network-based middleware solution were intelligent agents in the access network provided support to users in tasks like network selection, resource and handover management, service discovery and QoS parameter adaptation. This approach too relied on coverage maps to predict the MN residence time. Being network controlled, it was not active enough to react to sudden changes in network conditions at MN interfaces.

A recent study conducted in the area of handover management [Cottingham. 2009] employed coverage maps for improved network coverage prediction exclusively in vehicular environments. A noteworthy achievement of this study was the presence of real-time coverage data gathered by diligently driving a vehicle around a city and measuring the received signal strengths of different detected networks. However the study itself honestly acknowledged the lengthy duration of the year and half long data gathering phase together with the problem of outdated records. It also accepted the unsuitability of RSS-based handover prediction methods for smaller coverage networks like WLAN which suffer from rapid fluctuations in RSS, an important point which this research study advocates as well.

3.4   Mathematical calculation and modelling based handover prediction

The mathematical calculations/modelling based handover prediction category consists of theories that aim to predict future handover conditions by dynamically applying mathematically derived formulae and models to available network information. The main drawback of this approach is that information processing for results can be computationally intensive so it is important to develop efficient solutions which do not exert great demands on the mobile node's limited computational resources.

Ebersman and Tonguz in [Ebersman, et al. 1999] acknowledged the importance of knowing the time a MN had before handover and proposed the Signal Prediction Priority Queuing (SPPQ) method. This calculated the expected time a MN had before it reaches the cell boundary and performed a horizontal handover based on the change in received signal strength (RSS). This study succeeded in

deriving an estimate of time before handover (TBH). It also proposed the interesting concept of employing the values of RSS and change in RSS to assign different priorities to MNs. However, as the TBH solutions did not consider the position and direction of a moving MN, they failed to capture the accuracy of the MN's movement. According to the SPPQ method, a user that experienced degradation in RSS was assigned a higher priority. However in reality a mere decrease in RSS did not mean that the user was near the boundary, and so the mechanism that predicted the time before a call was lost sometimes failed to predict the correct value. This approach also assumed only the outward movement of calls and was not flexible in considering random MN movements in all directions.

The authors in [Shen, et al. 2000] proposed an adaptive fuzzy inference system which aimed to predict the probability that a MN would enter a particular cell in time (t). This approach utilised current and previous power signal measurements to predict future mobility information. All BS power signals measurements received at a MN were sent to the mobile switching centre (MSC) through the home BS which then performed the prediction calculations. This scheme took into account realistic measurements to predict the next cell to which the MN was expected to handover, however it was limited to that function only and did not predict the time a MN was likely to remain in the coverage of the current cell. Also, being computationally intensive, the scheme demonstrated acceptable performance in homogeneous networks until the BS count was less than 6, after which computational overhead increased significantly. This will clearly cause problems in heterogeneous networks where a MN may sense many different BSs from different networks at any given instant.

Makela et al [Makela, et al. 2000] proposed a predictive vertical handover approach which employed MN movement pattern classification as a means to predict a MN's position and distance to cell border. Unfortunately this technique lacked accuracy and was not equipped to handle sudden changes in MN movement patterns, particularly in indoor environments. It mainly focussed on deriving the best time for handover in terms of actual handover execution and did not focus on predicting how long the MN had before handover.

Ylianttila et al [Ylianttila, et al. 2001] developed an extension to the dwell timer [Hatimi, et al. 1999] scheme which was the predefined time for which a MN remained in the old network taking samples of the RSS from the AP and comparing them with a predefined threshold. If these samples taken in the

dwell time were below the threshold then the MN initiated the handoff to the other network. The Ylianttila study further proposed using a predefined dwell-timer for different data rates. While the study proposed intelligent solutions to eliminate the ping-pong effect, it did not provide a quantitative measure of the time the MN was expected to dwell in the old network. It also suffered from the inaccuracy resulting from employing only RSS as the threshold parameter. Bing H. et al in their study [Bing, H. et al. 2003] demonstrated how when distance criterion was taken into account, the handover probabilities are smaller than those only based on RSS criterion.

H. Wang and colleagues [Wang, et al. 1999] defined the useful concept of stability period which was the minimum duration for which the new network had to consistently display the lowest value of f_n in order to make it better choice for handover. This period was

$$T_{makeup} + T_{handoff}$$

T_{makeup} was the amount of time required to make up for loss of data or money due to handover latency,

$T_{handoff}$ - the handover latency itself.

Values of these two parameters were based on recent measurements in the past. The study also assumed the availability of context information such as network maps which could aid in the deduction of their values.

Chen et al. [Chen, et al. 2004] further refined the stability period through the utility function which was the sum of the product of network parameters and their assigned weights. The utility ratio was the ratio of target network utility by current network utility. The duration of stability period increased or decreased dynamically based on the decrease or increase in utility ratio. Both Chen and Wang focused on deriving a quantitative measure of the time needed to overcome the effect of a vertical handover which in turn would decide whether it was suitable to switch to the new network. This can be very useful during the prediction of unnecessary handovers. However the studies did not predict how long the MN was expected to reside in the current network coverage. In order to encourage the MN to handover only when it needed a better network, W-T Chen and Y-Y Shu [Chen, et al. 2005] proposed modifying the sequence of steps for vertical handover to handoff decision, system discovery

and handoff execution. This prevented a MN from unnecessarily repeating the first two steps in handover management when its QoS demands were met. In the event of choosing a new network, the dwell timer was applied to prevent the ping-pong effect.

Another study [Guo, et al. 2005] proposed the combination of Fuzzy logic and modified Elman neural network (MENN) for making handoff decision. The study identified parameters influencing the decision such as bandwidth, velocity and number of users. However the approach did not consider detailed QoS profiles of networks, topological information or user preferences.

Duong et al [Duong, et al. 2004] proposed a forced handover prediction scheme which proactively estimated the best moment in time for transferring user related context information to assure the shortest waiting time of the transferred context at the new access network. The MN then made a forced handover after the scanning-to-handover time expired, regardless of whether the handover condition was satisfied or not. The problem with this scheme was that it could actually aggravate the problem of unnecessary vertical handovers. Also, the time before handover could only be predicted very close to the handover during the scanning period and thus could not be applied at an earlier stage for traffic allocation and network selection.

Zhu and McNair [McNair, et al. 2004] proposed a policy-based vertical handoff decision algorithm where the calculated cost provided a measurement of the benefit of handing off to a certain network. However elimination parameters considered were RSS and channel availability only which do not provide sufficient information if the MN wants to avoid unnecessary vertical handovers.

Areej Saleh [Saleh. 2004] in her thesis proposed a location-aided handover decision algorithm for the prediction of unnecessary vertical handovers. This study was limited to only predicting handovers once the time needed for the actual handover was known. Issues such as the drastic changes in a MN's direction were acknowledged but not investigated.

In order to solve incertitude in vertical handovers due to the Line of Sight vulnerability in 60GHz LOS interfaces, Wang and colleagues [Wang, et al. 2007] proposed an algorithm based on Decision Theory which aimed to predict the duration of disruption in LOS communication which in turn helped to decide whether or not the device should switch to WLAN. The scope of this approach was limited to

resolving incertitude in LOS communication in indoor environments and it did not consider issues arising in WLAN based vertical handovers due to topological effects.

A recent study [Liu, et al. 2008] extended the traditional hysteresis-based and dwelling-timer based algorithms and proposed a new general handover decision algorithm for triggering handovers in heterogeneous wireless networks at the appropriate time. It was mainly built on the RSS-based solutions and assumed a stable signal coverage area for both WLAN and 3G networks. A limitation of this study was the assumption that there were no geographical topology related restrictions on the MN's movements and allowed the device to roam freely in any direction with any speed. These conditions however do not exist in realistic scenarios as will be demonstrated in later chapters in this thesis. In pedestrian indoor environments for instance, a MN frequently encounters obstacles such as walls and furniture which make a significant impact on its movements and its ability to exit network coverage. Rapid RSS fluctuations are also a common problem in WLANs, especially when a MN moves at higher speeds.

X. Yan et al [Xiaohuan Yan, et al. 2008] proposed a mathematically derived model based on the prediction of travelling distance which aimed to avoid unnecessary vertical handovers from cellular networks to WLANs. The proposed approach relied mainly on RSS for MN-AP distance calculation and was based on the assumption of a circular WLAN coverage. While this approach did succeed in predicting unnecessary vertical handovers to WLAN, it had several shortcomings. First the solution could predict an unnecessary vertical handover only when the MN's trajectory actually cut the WLAN cell coverage boundary and could not capture the random movements of the MN within the WLAN cell. For example the technique would not work if a MN entered the WLAN cell, stopped, changed direction and moved inwards towards the centre. Secondly, the accuracy of the proposed solution in predicting unnecessary vertical handovers increased only for MNs travelling with speeds above 15 m/s. As it is a widely accepted practice not to assign MNs travelling at higher speeds to WLAN cells, the technique simply reinforced the validity of this practice. In fact results demonstrated that the accuracy of the system decreased up to 70% when MN speeds fell below 10 m/s. A reason for high accuracy in higher speeds was that the faster the MN, the lesser was the RSS sampling rate.

3.5   Critical review outcomes

The main purpose of this chapter was to provide a thorough evaluation of related solutions proposed during the last decade in the field of vertical handover prediction. One of the main outcomes of this exhaustive survey was the revelation that despite the availability of a rich variety of context information, wireless heterogeneous devices still lack the intelligence to recognise their surrounding environment, particularly the precise knowledge of network coverage availability. The review evaluated a number of studies that proposed solutions to tackle the issue of predicting network coverage. However none of them succeeded in providing a simple and effective solution that dynamically predicted the duration of network coverage availability for a MN from available context information. In other words they failed to provide a dynamic and flexible answer to the research question

How can a MN roaming within a network predict future network availability relative to its motion within the network, and determine how long it has before it performs a vertical handover?

An important reason for this failure is the lack of new type of context information that specifically recognises coverage boundaries. While it may be argued that coverage based prediction techniques do provide a means to obtain this information, the approach does not facilitate the dynamic detection of boundaries by a MN which is a crucial requirement of 4G clients and requires the continuous supply of coverage data.

The second important deficiency that emerged common in almost all approaches was their inability to accommodate the truly random movement behaviour of MNs, particularly in pedestrian environments. The third deficiency unravelled was the failure to empower the MN with proactive mechanisms that enabled it to calculate network coverage availability dynamically as per its needs. The fourth deficiency was the lack of a hybrid mechanism that functioned correctly while considering both vehicular and pedestrian speeds and behaviour. This was largely because most studies chose RSS as the key decision parameter for handovers. All these deficiencies form the key gaps in knowledge which this research study aims to fill with a successful contribution of new ideas in the field of handover prediction.

3.6   Chapter Summary

This chapter reviewed important research work related to this research study in the area of vertical handover prediction. It not only acknowledged key contributions made by these studies, but also highlighted important issues including the absence of proactive knowledge of network availability which they failed to tackle. It explained how these unresolved issues contributed towards degrading the performance of handover optimisation techniques. The next chapter proposes new solutions that aim to increase the context awareness of a MN in a manner which grants it the ability to proactively predict vertical handovers.

CHAPTER 4

TBVH-assisted Vertical Handover Prediction

4.1   Introduction

The accurate prediction of network coverage allows the MN to proactively prepare for anticipated changes in QoS and channel conditions during vertical handovers. The two factors that can play a fundamental role in correct decision-making during the handover phase are:

1) Knowledge of network availability and coverage boundaries,

2) Knowledge of a MN's location and movement direction.

This chapter focuses on the combined implementation of new and existing concepts in a very unique manner to develop a set of mathematical models for they dynamic derivation of a novel parameter - Time Before Vertical Handover (TBVH). TBVH is a composite parameter calculated by the MN. The solutions proposed in this chapter are aimed at successfully answering the crucial research question

How can a MN roaming within a network predict future network availability relative to its motion within the network, and determine how long it has before it performs a vertical handover?

The chapter proposes a set of mechanisms that target the more challenging random mobility patterns of pedestrians in both enclosed and outdoor environments. However they can be applied to vehicular environments as well.

The size of WLAN hotspots can range between a single room to many square metres of overlapping coverage located in public places like restaurants, libraries, universities and airports [Bing. 2008]. As devices moving in or out of WLAN coverage are more likely to experience frequent vertical handovers due to its smaller coverage area, it is crucial to focus on the development of new techniques for enhancing WLAN coverage predictability and system discovery. As a result WLAN is taken as the principle technology for determining TBVH. However the operational principles of these solutions are not limited to WLAN and can be applied to other wireless technologies as well. This chapter consists mainly of two sections. Section 4.3 introduces a novel set of solutions for the intelligent detection of

surrounding coverage context and section 4.4 covers an extensive explanation on the derivation of TBVH models.

4.2   Problems with RSS as a reliable handover trigger

Recognising the significance and importance of improved handover prediction mechanisms the Wireless World Research Forum (WWRF) has emphasised the fact that future wireless devices must possess the ability to efficiently adapt their behaviour to context information like location, situation, personal role, task and environment [Tafazolli. 2005]. Reacting to the need of context-awareness, an increasing number of studies proposed innovative solutions on how to derive different types of context information including user location, customer status, network statistics, device mobility patterns, speed and direction, user activity history, network coverage maps and application characteristics. However, the majority of studies adopted Received Signal Strength (RSS) as a key indicator of network availability. In wireless networks, although a rapidly deteriorating value of RSS can be a good indicator that the MN is approaching the coverage boundary and may soon perform an imminent handover (horizontal or vertical), in heterogeneous networking the metric alone cannot be considered a reliable trigger due to several reasons:

- The RSS from different networks varies significantly because of differences in coverage and differences in techniques employed at the physical layers due to which they cannot be easily compared [Liu, et al. 2008]. RSS fading patterns can also be very different due to large differences in BS-MN distances for different networks.

- RSS measurements alone cannot provide answers to complex questions such as the precise knowledge of how long the MN is expected to remain in the AP's coverage. This knowledge is important for decision-making during both horizontal and vertical handovers as it can at a very early stage allow the MN to take important decisions on matters of resource allocation and QoS management.

- Rapid variations in signal levels due to phenomena such as multipath fading and shadowing and sudden changes in MN speeds and directions make it difficult to predict future RSS and signal quality [Makela, et al. 2000].

- A MN employing RSS as a handover trigger is programmed to scan available channels once the RSS from the current AP falls below a threshold. However if a MN is powered up at the border of

a set of WLAN cells with measured RSS approaching the handover threshold, it will keep scanning for new APs [Liao, et al. 2006] despite the fact that the APs currently available may be able to fulfill its resource demands, resulting in unnecessary vertical handovers.

Therefore what is needed is a more robust and proactive metric that not only gives the current status of network coverage availability but which can also predict for how long the coverage and network services are likely to remain available.

4.3   Intelligent coverage context detection for WLAN

In the heterogeneous 4G wireless environment one of the main challenges for a multi-interfaced client is the activation of interfaces for the detection of a network's coverage. In order to ensure that a new wireless network was detected as soon as it became available, a common practice in the past was to keep all interfaces activated all the time. This led to a huge increase in power consumption due to simultaneous activation of multiple interfaces. To tackle this issue, some early studies like [Katz, et al. 1999] proposed the periodic activation of interfaces. But the frequency of interface activation directly affected system discovery time [Chen, et al. 2004].

Another proposed solution [Wu, et al. 2002] was the development of a Location-Service (LS) Server which stored the precise information on coverage areas of wireless networks. Upon receiving the location update from the MN, the server supplied it with information on reachable networks. This technique however was both time and resource intensive and successful detection of new networks was directly dependent on how frequently the location server was updated.

To address the issue of coverage information complexity in the LS server, Chen et al [Chen, et al. 2004] proposed the ideal coverage concept for downward vertical handovers where the LS server stored only the position information of the AP of a reachable network, its ideal coverage radius and minimum coverage radius. A MN roaming towards this new AP became aware of its coverage and activated the corresponding interface once it reached the ideal coverage area of the new AP. This effort succeeded in reducing the unnecessary 'on' time for WLAN interfaces and reduced the complexity of coverage data stored in the LS server. It however did not eliminate the problem of updating information in the server.

The above mentioned solutions proved useful in introducing a level of predictability to the problem of uncertainty in network coverage, but they did not consider an important point. In a loosely-coupled, multi-operator scenario it will be difficult to decide who controls the centralised server. Moreover, small independent hotspot providers would be under no obligation to share their coverage information with others by uploading it on the LS Server. Thus it is not possible for the MN to always have correct information about the availability of other networks in its vicinity by relying solely on the LS server and it may have to eventually switch on all its interfaces to discover new networks in its vicinity.

In order to eliminate the need to rely on an external entity like the LS server for coverage information, this study proposed a more feasible approach which is to empower each network with the ability to identify its own coverage boundaries and to dynamically supply the MN with this information. This can be achieved by adopting a more distributed approach where the BS is equipped with additional intelligence and functionality which allows it to store information about its own coverage and the coverage of other reachable networks in its vicinity. This information is then supplied to an incoming MN.

This section introduces the reader to a novel approach where a MN attached to a WLAN AP is automatically supplied with new context information about the coverage characteristics in its vicinity. One of the key deficiencies identified in earlier approaches in chapter 3 was the inability of devices to detect coverage boundaries. The next sub-section 4.3.1 discusses the concept of forced vertical handovers and how it is applied to create a crisp coverage boundary.

4.3.1   Forced vertical handover for coverage boundary prediction

A forced vertical handover means that a MN is forced to switch communication between network interfaces e.g. WLAN to UMTS when resources on the current interface start diminishing. The concept of forced handovers was adopted by studies in the past as a means for improving performance when a MN's connection was migrated from one interface to another. Forced handovers are triggered by physical events related to network interface operability issues. As service disruption is proportional to the delay after which the MN becomes aware of disconnectivity, early detection and correct triggering are two important requirements in a forced handover [Bernaschi, et al. 2005]. Let us take a look at how different studies have applied this concept.

4.3.1.1 Related Work

The POLIMAND project [Aust, et al. 2003] proposed a policy-based Mobile IP handover decision solution where handovers were forced based on parameters measured by the Generic Link Layer (GLL) – a simple layer similar in functionality to the current 802.21 standard. Vertical handovers decisions were based on information received from the GLL. Bernaschi et al [Bernaschi, et al. 2005] also proposed forced vertical handovers triggered after the occurrence of events which indicated deterioration in link quality. Another study that adopted this concept was the Dailalos project [Aguiar, et al. 2006] where signal strength and QoS measures from the network were the main deciding factors which forced a MN to perform a vertical handover. During their experiments on vertical handovers, Charavorty et al [Chakravorty, et al. 2004] proposed forced handovers between networks by keeping the router advertisement frequency constant on one link and varying it on the other. All these studies were reactive in nature and focussed on forced handovers as a response to event triggers which signalled a drop in the current link's QoS or signal strength levels.

Proactive forced handover schemes focus on utilising improved context awareness for a MN to force the vertical handover to occur at a planned moment. [Duong, et al. 2004] was one such study that proposed proactively estimating the best moment in time for transferring user related context information to ensure the shortest waiting time of the transferred context at the new access network.

All these studies adopted the concept of forced handovers during the actual switching of network interfaces but none of them focused on applying the concept to achieve improved vertical handover prediction. RSS was used as an indicator of deteriorating network conditions but it did not succeed in reducing uncertainty associated with the detection of coverage boundaries. As explained in earlier chapters, the precise prediction of available coverage forms crucial information which can be employed successfully by the layers at both the device and network side in various ways, resulting in enhanced performance.

As the first step towards developing an effective solution, this research study proposes applying the concept of forced vertical handovers to WLAN in a unique way to reduce fuzziness at the WLAN coverage boundary, replacing it with a clear and crisp cell structure. The fundamental principle is to not wait until the quality of the WLAN link deteriorates completely but to perform a vertical handover at
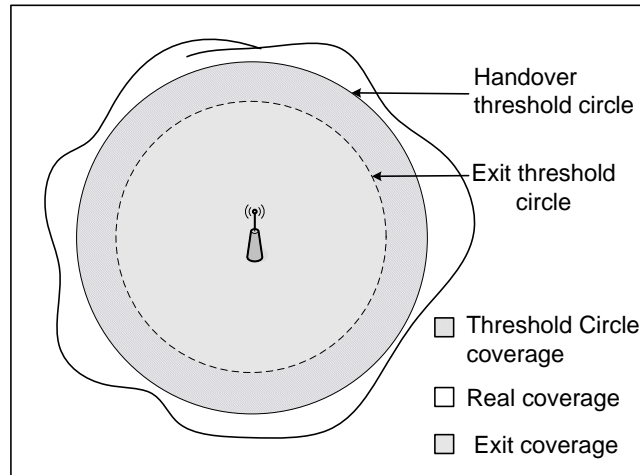
Figure 4.1 Threshold circles

a specific, predetermined point, close to the time of the actual handover, even slightly before it. This introduces a level of certainty in the whole handover process due to the introduction of precise knowledge of coverage boundaries. Shorter hops in the WLAN network led to lower variations in throughput and channel fading [Bing. 2008]. Hence the network's channel conditions are relatively more stable as compared to 3G [Chakravorty, et al. 2004], facilitating the implementation of the proposed solution.

4.3.1.2 The threshold circles

In order to facilitate forced handovers and to predict a MN's future movements, the study proposes the introduction of two types of threshold circles in the WLAN coverage – the handover threshold circle and the exit threshold circle as shown in figure 4.1. The handover threshold (HT) circle in the WLAN cell represents the coverage boundary. It is the distance from the AP at which a MN performs a vertical handover to another network. Initially, the handover threshold is assigned a default value by the network administrator during the WLAN setup phase.

$$d_{HT} = d_{initial}$$

After that the threshold radius is based on the average of the last n distance readings stored by the AP at the time the MN performs a vertical handover.

$$d_{HT} = \frac{(d_{HT})_1 + (d_{HT})_2 + (d_{HT})_3 + \cdots + (d_{HT})_n}{n}$$

74

A MN approaching the HT boundary had a clear estimate of the time available to spend in the current cell. This in turn allows it to utilise available resources more efficiently before moving out. Chen and colleagues [Chen, et al. 2004] proposed a somewhat similar concept called the minimum coverage which was the largest circle that could include the real coverage. The problem with the minimum coverage circle was that it also included empty coverage-less areas. Hence the MN would sometimes have to perform a vertical handover before moving out of the minimum coverage circle. A point may be raised that a MN when handing off at the HT distance might do so before reaching the end of the actual coverage at that point, resulting in an under utilisation of available resources. However the reality is that the improved coverage predictability achieved through the precise knowledge of time before vertical handover can actually assist in the efficient allocation of resources in both the current and future network, based on the MN's needs, even allowing it to complete a data transfer session before the MN moves out of WLAN coverage.

The second circle is the exit threshold (ET) circle, introduced to improve the predictability of the MN's change in direction. This simple yet effective concept was adopted from Chiu and Bassiouni's study [Ming-Hsing Chiu, et al. 2000] who proposed the threshold distance (TD) which is a distance smaller than the cell radius that specifies an inner circle. The TD radius is carefully chosen such that a MN once outside it, rarely undergoes a drastic change in direction. This helps to minimise the false submission of reservation requests to a neighbouring cell. The experiments in Chiu's study focused on vehicular movements and were carried out in a cellular environment with a cell radius of 1000 m and a TD radius of 800 m. As this research study targets pedestrian behaviour where MNs are more likely to experience frequent and random changes in direction, the ET distance is dependent on the HT distance and is kept closer to it, within 30-70 m. At pedestrian speeds, this gives the MN enough time to submit new requests to the neighbouring AP. In the case of an AP at the coverage boundary, it allows the MN to prepare for a vertical handover to the next available network. Therefore the main function of the ET circle is the definition of an area beyond which the MN starts the procedure of a vertical handover through the submission of resource requests to the new network if it is available.

4.3.2   Service environment classification

A unique feature of the TBVH-aware client is the ability to recognise the type of environment in which it is located. Based on this new type of context knowledge the client activates the correct environment
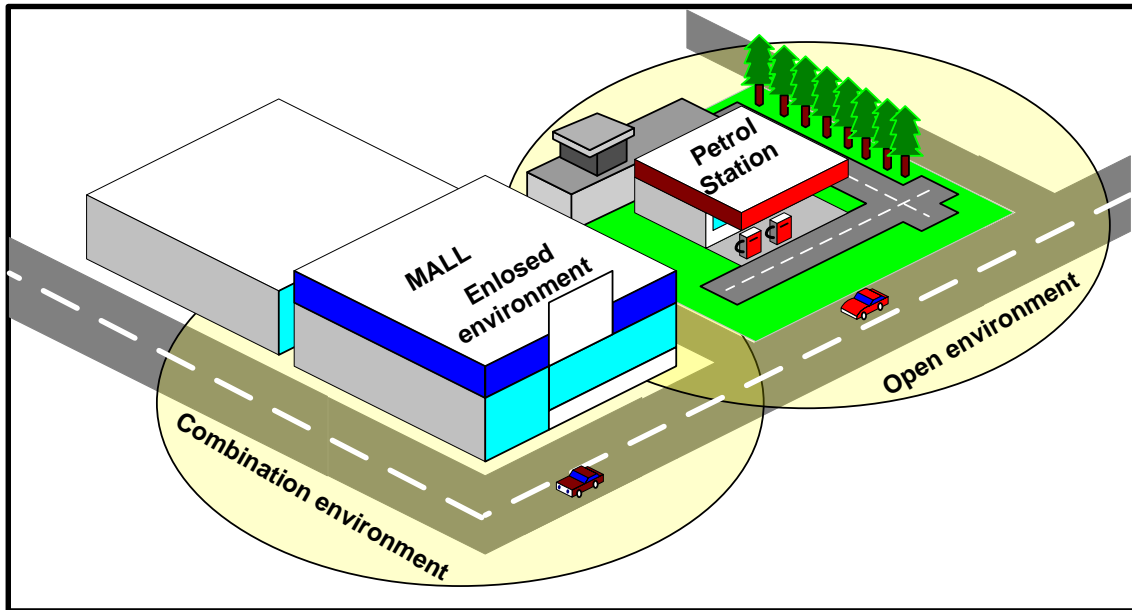
Figure 4.2 WLAN service environments

model to derive TBVH which is further applied in other important tasks such as network selection, handover detection and resource negotiation. The different environments are classified as open, enclosed and combination environments.

4.3.2.1 Open environment

In the open environment the MN's movement is not constrained by physical confines such as walls. The device is free to roam within the coverage area and encounters only small obstacles in its trajectory which do not necessitate a drastic change in direction. More importantly, in an open environment a MN is free to exit from any point on the coverage boundary. TBVH calculation in this scenario is more straightforward and involves consideration of network coverage boundaries only. Examples of an open environment are a playing field or outdoor garden. In figure 4.2, the WLAN coverage provided to customers at the petrol station is said to cover an open environment as there are no walls or boundaries around the coverage that restrict MN movements, particularly at the time of exit.

4.3.2.2 Enclosed environment

A MN roaming under wireless coverage of an enclosed environment has its movements restricted to inside free spaces that are surrounded by physical boundaries such as walls and partitions. These

obstructions can cause a drastic change in the direction of the MN hence it is important to be aware of their presence in advance. In an enclosed environment a MN can enter and exit the coverage area only through a specific set of points which form the physical exits for the enclosed area. An enclosed environment can consist of outdoor or indoor coverage. Examples of the enclosed environment are buildings and outdoor enclosed areas such as courtyards. WLAN coverage inside the mall in figure 4.2 covers an enclosed environment.

Choosing the boundaries for an enclosed environment can have a significant effect on TBVH derivation. For instance, in the case of an organisation which provides indoor wireless coverage within the building and outdoor wireless coverage in its communal gardens, it is easier to consider the actual outer boundary of the premises e.g. the garden's outer limits as the environment's boundary.

When considering the case of equipping the MN with knowledge on the enclosed environment's physical boundaries, security concerns may arise about the suitability of sharing such information with the roaming end users as early as the scanning phase before authentication. However, it is important to understand that the information shared here is only that of the outermost structural boundaries which can be easily obtained by anyone these days through actual physical measurements or with the help of freeware like Google Earth.

4.3.2.3 Combination environment

In a combination environment, the AP coverage falls partially under both open and enclosed areas. An example of this type of coverage environment is a wireless hotspot that covers both a building and the footpath outside the building. Here a wireless user can exit the building (enclosed environment) but still remain under the coverage of the hotspot for a few minutes while walking on the neighbouring footpath (open environment). Alternately, a user initially connected to the open coverage environment outside the building is likely to also experience the enclosed coverage environment upon entering the building. In figure 4.2, a customer communicating using the mall's WLAN while inside the building will remain connected to it for some time even after exiting the building. This is an example of the combination environment.

4.3.3 Enhancements to access point functionality

A key factor in the successful, seamless integration of wireless access technologies is the minimisation of the disruptive effects of vertical handovers through the creation of well-defined boundaries. Recognising the need for this new type of context information, a few studies in the past suggested mechanisms for the identification of service area boundaries. The first step in this process is the recognition of coverage cells that reside at the network's coverage boundary as a MN attached to a boundary cell has a greater chance of performing a vertical handover.

Liu and Maguire [Liu, et al. 1996] defined the concept of a boundary state which was an area or cell located at the boundary of the service area. But the purpose of the boundary state was limited to identifying service boundaries and assigning higher priorities to MNs that had a boundary state in their movement model. An AP at the boundary did not play a significant role in improving proactive decisions during handovers.

Akyildiz and Wang [Akyildiz, et al. 2002] proposed the concept of boundary location area (BLA) controlled by a boundary interworking unit (BIU) located at the boundary between two systems at different hierarchical levels. The BIU was connected to the mobile switching centres and visitor location registers in both systems. A database cache in the BIU stored the roaming information for MNs moving between different systems and allowed both systems to access information on roaming MNs, resulting in a reduction in paging costs. This approach was suitable only in tightly-coupled systems. As the BIU was an independent unit which existed outside the AP, it was not enhanced to handle vertical handovers independently.

As mentioned earlier in Chapter 2, while the IEEE 802.21 standard possessed triggers which informed the device about the current status of a wireless link, the standard did not possess the ability to proactively predict future changes in link conditions, nor did it possess mechanisms that dynamically informed it of surrounding topologies and coverage characteristics.

In TBVH calculation, the AP plays a crucial role in enhancing the context knowledge of a roaming MN by providing information about its signal coverage and the limits of the overall network's coverage boundary. This is achieved through the transmission of a set of specialised packets containing

coverage and topological information to the MN when it registers with the boundary AP. APs are classified into two categories - normal and boundary AP.

4.3.3.1 Normal AP (NAP)

This is the AP that does not fall at the boundary of network coverage and hence is not directly involved in the vertical handover process.

4.3.3.2 Boundary AP (BAP)

This AP falls at the boundary of network coverage and a MN connected to it may experience a vertical handover. Being the main point of contact, it plays a significant role in providing the MN with context information which assists the MN in preparing proactively for a vertical handover. A MN approaching a BAP is automatically aware of the fact that it is approaching the network boundary and may have to perform a vertical handover in the near future.

This simple yet effective classification of APs assists in reducing the computational overhead at the MN as well as enhancing its intelligence as:

1. a MN needs to know TBVH only when it approaches the coverage boundary. Therefore the entire mechanism is needed only when the MN is under the coverage of a BAP or is approaching the BAP from an NAP.
2. a MN connected to an NAP is not in need of a high frequency of location updates.

4.3.4 The MN-AP information exchange mechanism

In order to equip an MN with the intelligence to independently determine TBVH, it is crucial to enrich its contextual knowledge by supplying it information on the surrounding coverage, topology and geographical layout. All APs in a TBVH-aware WLAN hotspot are equipped during the network setup phase with knowledge about the surrounding geographical topology for which they provide coverage. An important issue is to develop efficient communication mechanisms to convey this topological information to the MN.

The main design goals of the AP-MN context exchange communication mechanism are:

1. To dynamically provide context information to the MN as per its needs.

2. To refine the nature of coverage information to be supplied to the MN in a way in which it assists in increasing the MN's awareness of its surroundings. The mechanism also assists in avoiding the increased network overhead that can result from the continuous transfer of large chunks of coverage and topology data.

The fulfilment of these design goals involves the development of specialised MAC layer packet communication mechanisms. The transmission of coverage boundary and geographical topology related information to the MN takes place at specific points in time based on its location within the coverage area. This information transfer takes place at a very early stage during system discovery when the MN communicates with the AP for the first time. The reason for the early supply of information here is to equip the MN with the ability to derive TBVH as early as system discovery and enable it to avoid unsuitable APs, hence playing a key role in eliminating bad choices during the connection establishment phase (discussed in section 2.2.1.2) and the overhead associated with it.

In the WLAN system discovery phase, a MN scans wireless channels either actively or passively as explained in section 2.2.1.3 to detect APs in its vicinity. Each AP responds back with the transmission of the Probe Response beacon frame which is a WLAN Medium Access Control (MAC) layer frame that contains information about the network's characteristics such as data rates and physical layer technology. In the enhanced MN-AP information exchange proposed in this study, clusters of useful topological information are inserted in the probe response frame's body. The next sub-section 4.3.4.1 explains this process in detail.

4.3.4.1 Enhancements to the probe response beacon frame

In WLAN, the MAC frame is designed to carry a large variety of information. Based on the type of information the frame is classified as a control, management or data frame. A device receiving a
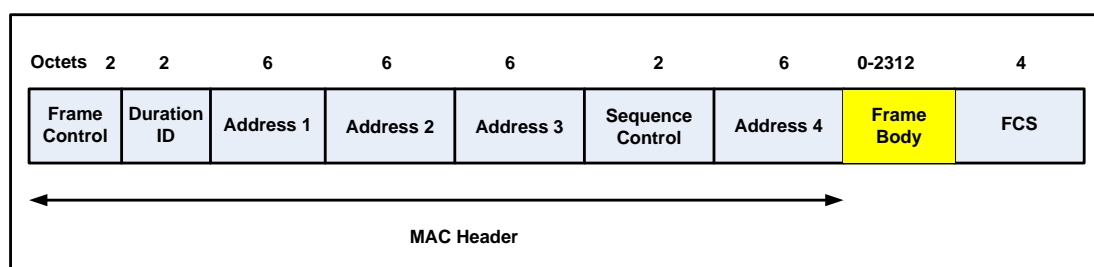
Figure 4.3 WLAN MAC Frame [LAN/MAN Standards Committee. 1999]

Table 4.1 The MAC beacon's frame body field contents [LAN/MAN Standards Committee. 1999]

Order	Information (size in octets)	Notes
1	Timestamp (8)	
2	Beacon interval (2)	
3	Capability information (2)	
4	SSID (0-32)	
5	Supported rates (1-8)	
6	FH Parameter Set (5)	This information is generated by STAs using frequency-hopping PHYs.
7	DS Parameter Set (1)	This information element is present within Probe Response frames generated by STAs using direct sequence PHYs.
8	CF Parameter Set (6)	This parameter set is only present within Probe Response frames generated by APs supporting a PCF.
9	IBSS Parameter Set (4)	This information element is only present within Probe Response frames generated by STAs in an IBSS.

WLAN frame identifies its type based on the combination of bits the 16 bit Frame Control field. Figure 4.3 illustrates the MAC frame format which consists of the header, body and frame checksum. For the sake of brevity this subsection does not delve into the explanation of each field and its contents as this can be found in detail in [LAN/MAN Standards Committee. 1999]. It instead narrows the focus down to the Probe Response beacon frame in the management category and elaborates on the new information introduced in to the Frame Body field. The Probe Response beacon is transmitted on two different occasions - either as an advertisement beacon by the AP at regular intervals during passive scanning to inform a MN about the existence of a new WLAN coverage, or as a response to an explicit probe request submitted by the MN.
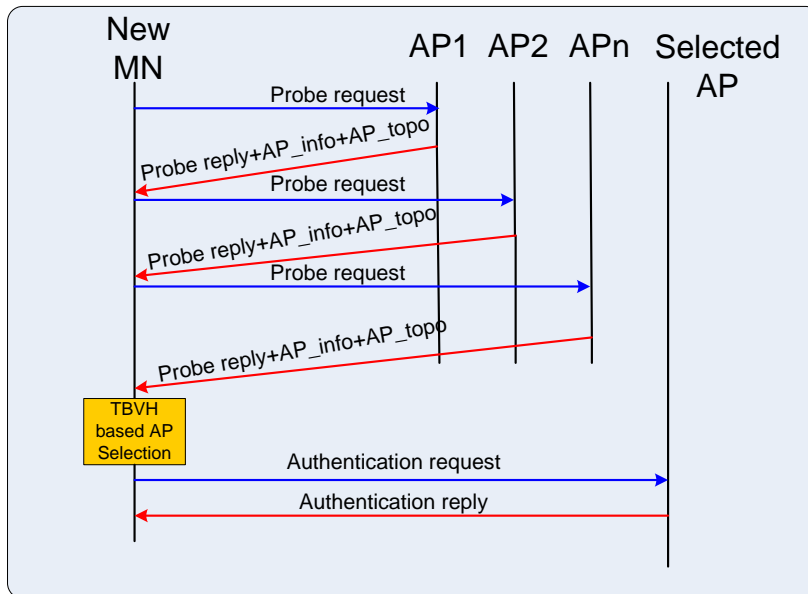
Figure 4.4 Active scanning period in the system discovery phase

As shown in figure 4.3 the maximum length of the MAC frame's Frame Body field is 2312 octets. In the Probe Response MAC frame the frame body contains information on the coverage characteristics as shown in table 4.1. In the beacon frame, the maximum length the frame body can occupy is 68 octets, which leaves ample space for the insertion of additional information if needed. It is here that the AP inserts coverage and topology related information clusters for the MN, supplying it with enhanced intelligence for calculating TBVH in the system discovery period. Let us now look at how this information reaches the MN.

4.3.4.2 The beacon information clusters

In a TBVH-aware WLAN, a group of new information clusters is embedded by the AP inside the probe response beacon frame's body. The three clusters are *AP_info*, *AP_topo* and *AP_nested_topo*. These clusters supply the MN with additional information on the nature of the coverage of APs to which it may connect. They equip the MN with important information for calculating TBVH, enabling it to derive an estimate of the time it has left within the AP's coverage. This allows the MN to eliminate unsuitable APs at a very early stage. The *AP_info* cluster contains information about the coverage related context surrounding a particular AP. The *AP_topo* cluster consists of topological information such as indoor environment boundary and exit coordinates whereas the *AP_nested_topo* cluster contains information on the inner boundaries of a closed structure (explained later in this sub-section). These first two clusters are transmitted on two different occasions:

82
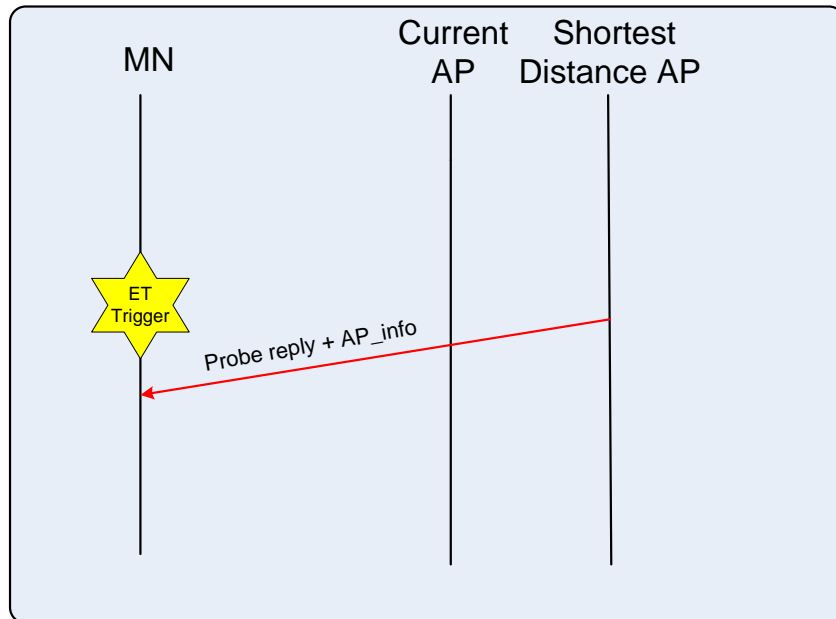
Figure 4.5 Passive scanning of beacon advertisements before handover

1. During the probe phase in system discovery period as shown in figure 4.4 when a MN scans for new APs belonging to a new hotspot to associate with. Upon associating with an enclosed WLAN for the first time in a particular area, the MN receives both the *AP_topo* and *AP_info* clusters.

2. During the discovery of the neighbouring AP as demonstrated in figure 4.5 when the MN moves out of the ET circle of the current AP and towards the neighbouring AP. Once again it receives the *AP_info* clusters from the neighbouring AP in the form of beacon advertisements.

However, as the MN here is simply attempting to connect to another AP within the same WLAN's extended service set, it already is aware of environment boundaries and there is no need to retransmit the *AP_topo* cluster again. Let us look more deeply at each type of information cluster.

AP_info **cluster**

The *AP_info* information cluster shown in figure 4.6 mainly contains AP related information. A description of the cluster fields is as follows:

- **AP type (1 bit)**: this field informs the MN whether the AP is a NAP (value 0) or BAP (value 1).

- **AP location (1 bit)**: This field informs the MN about the geographical location of the AP in the

AP Type	AP Location	Topology ID	AP HT	AP ET
AP X-coordinate			AP Y-coordinate	

Figure 4.6 The *AP_info* cluster

coverage environment, whether it is located in an open (value 0) or enclosed (value 1) environment.

- **Topology ID (8 bits)**: This field is a unique identifier for a particular enclosed environment.

- **AP coordinates (16 bits)**: This 16 bit field stores the X and Y coordinates of the AP. These coordinates are particularly useful if the AP is a BAP. In case of a normal AP, the coordinates assist in calculating TBVH if the next AP the MN is expected to switch over to is a BAP.

- **AP HT radius (8 bits)**: This field gives the HT distance from the AP at which the MN is expected to perform the handover. This may be either a vertical or horizontal handover. The value of this field is history based and is adjusted based on previous handover records.

- **AP ET radius (8 bits)**: This field gives the exit threshold distance.

AP_topo **cluster**

Along with the regular *AP-info* cluster, if the wireless hotspot which the MN wants to connect to for the first time falls under enclosed coverage, the MN receives an *AP_topo* cluster as well along with *AP_info* during the scanning period. This cluster shown in figure 4.7 contains additional topological

Topology Identifier		
No. of Normal edges	Normal edge coordinates x no. of normal edges	
No. of exit edges	Exit edge coordinates x no. of exit edges	
No. of exits	(Exit coordinates + BAP coordinates) x no. of exits	
		Nested topology identifer

Figure 4.7 The *AP_topo* information cluster

information about the enclosed environment which enables the MN to recognise geographical features such as physical boundaries and exits. The *AP_topo* packet can be transmitted by either a NAP or BAP and the information is stored in the MN's buffer as long as it remains connected to the current WLAN's extended service set.

In a real world scenario an enclosed environment may be covered by multiple WLAN service sets (SS) and a MN may qualify to connect to more than one SS. As a result it may be bombarded with the same topological information from different APs belonging to different SS. In order to eliminate this problem each enclosed environment topology is assigned a unique identifier by the administrator during the network setup phase. Once this packet is received the MN notes the unique topology ID and stores the information inside the TBVH module. Upon receiving subsequent beacon clusters, the MN scans their topology identifier field to check whether any new topological information is available. If so then the buffer is updated with the new *AP_topo* cluster information, else the MN ignores it.

It is important to note that the *AP_topo* cluster stores information only on the outermost structural boundaries and exits which fall under the coverage of the wireless network. An enclosed environment may contain many internal physical boundaries such as rooms but as the wireless coverage encompasses the overall enclosed environment, there is no need to include the internal boundary coordinates. This helps to maintain a simple cluster format and to reduce frame packet size. The fields describing the *AP_topo* cluster are as follows:

- **Topology identifier (8 bits)**: This is a unique identifier which is assigned during the setup phase to an enclosed environment.
- **Number of normal edges (8 bits)**: This field gives the number of exit-less edges that form the boundary.
- **Normal edge coordinates**: This 32-bit field gives the x and y coordinates of the start and end points of the normal edge. The instances of this field are a multiple of the previous field.
- **Number of exit edges (8 bits)**: This field gives the number of edges which contain one or more exits.
- **Exit edge coordinates**: A 32-bit field containing the x and y coordinates of the start and end points of the exit edge.

- **Number of exits (8 bits)**: This field gives the number of exits on the exit edge given in the previous field.

- **Exit coordinates**: This 32-bit field gives the start and end points of the actual exits. Once again the number of instances of this field is dependent on the number of exits.

- **BAP coordinates**: This 16-bit field gives the coordinates of the boundary access point that provides coverage to the exit.

- **Nested topology identifier (1 bit)**: This field indicates whether there are further internal structures nested within the enclosed environment.

As for the issue of assigning coordinates to the outer boundary edges, this task is done only once at the network setup phase by the network administrator and involves noting down location coordinates of corners and exits through straightforward measurements. It is important to note that while the inner structure of an enclosed environment may be complex due to internal structural details, e.g. a building which consists of many rooms, inner walls and internal exits, the outermost boundary is usually simple and consists of a small number of main exits. This makes the task of assigning coordinates feasible and straightforward.

As most indoor positioning solutions usually require an initial setup and site survey phase which involves site calibration, a question may be raised as to how necessary is a site survey. This decision depends upon how big and complex the WLAN network is; the more complex the network, the greater is the need for site survey [Bing. 2008]. Site surveys play an important role in the early identification of problems in signal propagation such as dead spots. With the help of a small number of sample readings, it is possible to greatly improve the performance of a Wi-Fi network. The Ekahau technology [Bing. 2008] for instance employs a sample of site specific calibration measurements consisting of sample points from different site locations as input. This collection is done using laptop and calibration software. A sample point consists of the Received Signal Strength Indicator (RSSI) and related map coordinates stored in an area-specific positioning model for accurate tracking. A site specific model is then created very quickly with the help of machine learning algorithms, achieving a positioning accuracy of 1 m [Bing. 2008]. In the case of TBVH models, the sample points themselves can consist of boundary edges and exit coordinates, reducing the overhead associated with additional

Topology Identifier	
No. of Normal edges	Normal edge coordinates x no. of normal edges

Figure 4.8 The *AP_nested_topo* cluster

measurements. The positioning measurements themselves need not be too accurate as they are needed mainly to provide a clear estimate of the geographical context.

AP_nested_topo **cluster**

Enclosed environments may further contain structures nested within them e.g. a building inside the enclosed environment. Assuming that a WLAN SS provides full coverage to an enclosed environment, in this case a MN roaming inside the nested environment will have to first come out of it and into the outer enclosure to experience a vertical handover. Hence there is no need to calculate TBVH as long as the MN resides inside the nested environment. The *AP_nested_topo* cluster contains simple information about the boundaries of the nested environment without any details about exits as shown in figure 4.8. It can be simply sent encapsulated inside a MAC data frame. In order to allay the aforementioned security concerns the transfer of the *AP_nested_topo* cluster takes place at a later stage after successful authentication in the connection setup phase as shown in figure 4.9. The contents of the *AP_nested_topo* packet are as follows:

- **Topology identifier**: This the unique identifier assigned to the nested topology.

- **Number of edges**: This indicates the number of edges in the internal structure.

- **Edge coordinates**: These are the edge coordinates for the internal structure. This field is a collection of all edge coordinates which is a multiple of the previous field.
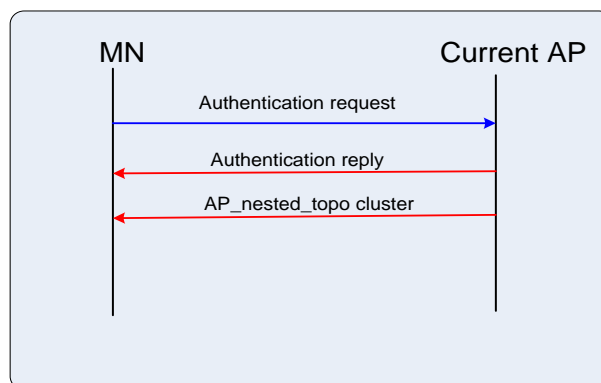
Figure 4.9 *AP_nested_topo* transmission diagram

4.4   Time Before Vertical Handover

This section explains in detail the process of calculating the new context parameter Time Before Vertical Handover (TBVH). TBVH is a parameter derived from the combined processing of a set of context information which includes the MN's location, its velocity and surrounding coverage features supplied to it by the AP. The mechanisms for extracting the MN's location, coverage boundary and context information have been explained in detail earlier in sections 2.5 and section 4.3 respectively. This section elaborates on the extraction of remaining inputs to the TBVH model, i.e. velocity, MN-AP distance and direction. It then explains the development of an extensive set of TBVH models that are invoked dynamically based on the MN's context.

4.4.1   TBVH model inputs

The three important context parameters that form inputs in the TBVH model are:

1. Distance between MN and AP
2. MN velocity
3. MN direction of motion

This sub-section provides an in-depth explanation on the derivation of these components. The knowledge of the MN's geographical position in the coverage area plays a pivotal role in their derivation.

4.4.1.1 AP-MN Distance

In literature, the two popular methods employed for calculating the distance between a MN and the AP it is attached to are:

1. RSS-based
2. Co-ordinate based.

Some earlier studies like [Ming-Hsing Chiu, et al. 2000, Ebersman, et al. 1999, Shen, et al. 2000] and more recent studies like [Xiaohuan Yan, et al. 2008] adopted the RSS-based distance method which calculated the AP-MN distance from path-loss models such as equation 4.1 [Stuber. 2001].

$$RSS_P = E_t - 10\beta log_{10} l_{OP} + \varepsilon$$ Equation 4.1

This method suffers from two drawbacks. Firstly, while it may be possible to calculate the AP-MN distance, this information is not sufficient in itself. For location-aware services it is more useful to know at least which quadrant in the coverage area the MN is currently located in. Secondly, the accuracy of RSS-based methods is largely dependent on RSS measurements which can suffer due to RSS fluctuations. This problem is visible in Yan's study [Xiaohuan Yan, et al. 2008] where the distance estimation error increased by up to 70% due to an increase in the number of RSS samples when the MN's velocity decreased below 10 m/s.

With a huge proliferation in the availability of commercial location tracking solutions offering tracking accuracy up to 1 cm in both indoor and outdoor environments, an increasing number of mobile applications are employing this facility to provide intelligent context-aware solutions. Location-coordinates-based MN-AP distance calculation has been adopted by numerous successful studies in the past [Ming-Hsing Chiu, et al. 2000, Wen-Tsuen Chen, et al. 2004, Vidales. 2005, Song, et al. 2006, Yang, et al. 2001] and is adopted by this research study as well.

While location prediction mechanisms may suffer sometimes from a level of inaccuracy, they can still be successfully applied to derive other useful information such as the direction of motion and speed. A highly accurate prediction of a MN's location for TBVH calculation is desirable but does not form a crucial requirement for successfully implementing the model. This claim is addressed at a later stage in chapter 5.

For outdoor environments, location information can be obtained with the help of standalone location prediction techniques like Global Positioning System (GPS) or network-based location prediction techniques like the Observed Time Difference of Arrival (OTDOA). Once the device co-ordinates are obtained e.g. device A = (lat1,lon1) and device B = (lat2,lon2), the distance between them can be derived from existing formulae such as the Greater Circle Distance formula or other similarly derived ones like:

$$d = 6378.7 \times arccos\left(1 - \frac{((\sin(lat\,1)) - (\sin(lat\,2))^2}{2}\right) + (\sin(lon1) - (\sin(lon2))^2 \qquad \text{Equation 4.2}$$

For the indoor environment, assuming the wireless network coverage covers the entire indoor area, as per the widely adopted approach in indoor positioning techniques it is represented by a grid. This facilitates the calculation of the AP-MN distance by the two-dimensional Euclidean distance formula

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$
Equation 4.3

4.4.1.2 MN Velocity

In mobile wireless networks, the knowledge of the MN's velocity is crucial for efficient channel assignment and resource allocation [Hellebrandt, et al. 1997]. The easy availability of location data for moving MNs has caused an increasing number of studies to adopt positioning data for the calculation of velocity. [Yang, et al. 2001, Cottingham. 2009, Saleh. 2004, Wen-Hong Zhu, et al. 2006] are several such studies that have applied positioning data for the derivation of velocity. Some of the latest devices in market like the iPhone 3GS are even equipped with sophisticated accelerometers that are capable of accurately predicting MN velocity. This research study adopted the positioning data approach and MN velocity is calculated straightforwardly from the time differences in two consecutive position measurements.

4.4.1.3 Direction of motion

In order to calculate the direction of motion for a moving MN the same approach proposed by Xinjie Yang [Yang, et al. 2001] is adopted by this study. According to the author, the knowledge of location co-ordinates for two consecutive positions of a MN enables the calculation of the angle of movement with respect to the AP, at the MN. In figure 4.10, the angle of direction α is derived
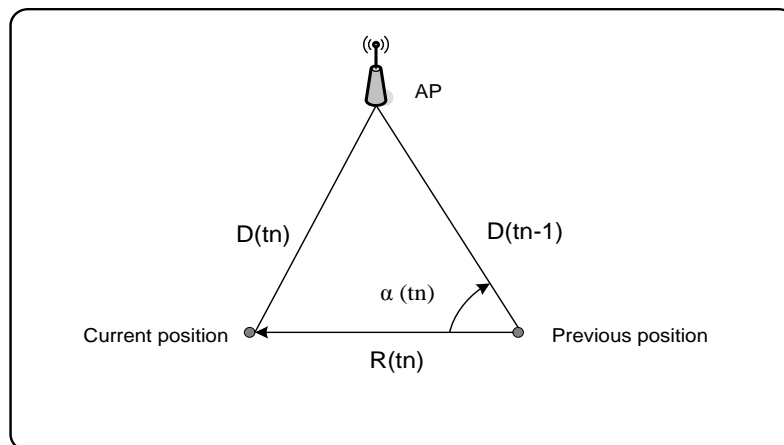
Figure 4.10 Angle of direction for moving MN [Yang, et al. 2001]

90

from the formula

$$\alpha(tn) = \ cos^{-1}\left(\frac{d(t_{n-1})^2+r(t_n)^2-d(t_n)^2}{2d(t_{n-1})r(t_n)}\right) \hspace{4cm} \text{Equation 4.4}$$

where $d(t_n)$ is the estimated distance to AP at time t_n,

$d(t_{n-1})$ the estimated distance to AP at time t_{n-1}, and

$r(t_n)$ the distance between the previous and current position.

4.4.2 Mathematical derivation of TBVH

This section presents a detailed exploration of different scenarios in TBVH calculation that arise due to different MN locations and movements in the WLAN extended service set for both open and enclosed environments. The different cases arise based on the MN's movement with respect to the current AP (CAP) and the type of the next AP it moves towards (NXAP).

4.4.2.1 The Open TBVH Model

CASE I: MN movement from normal AP towards another normal AP

The conditions for this scenario are

CAP = Normal AP

NXAP = Normal AP

$RSS_{CAP}(t_n) < \ RSS_{CAP}(t_{n-1})$

$RSS_{NXCAP}(t_n) > \ RSS_{NXCAP}(t_{n-1})$

$d(MN_{CAP}) > d(ET)$

In this case the MN has reached beyond the exit threshold for the current AP but the next AP that it is about to handover to belongs to the same ESS. As there is no imminent vertical handover detected there is no need to calculate TBVH.

CASE II: MN movement towards inner network
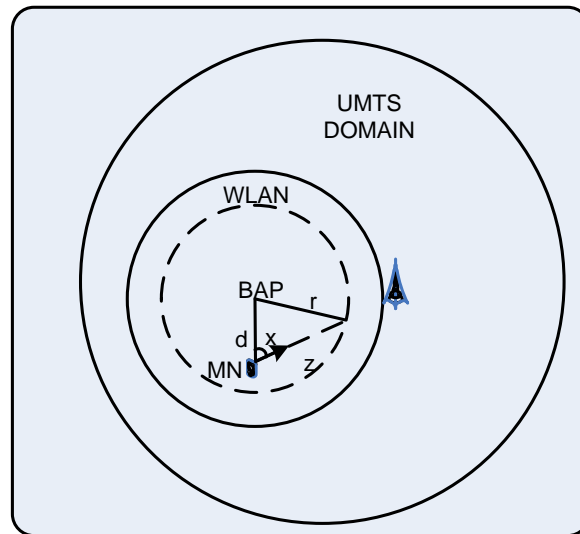
Figure 4.11 Movement of MN in BAP

The MN here is connected to a boundary AP. In this case the MN's movement direction under BAP coverage can give rise to two different cases in TBVH determination. In the case of the movement towards the inner network, once again there is no need to compute TBVH if the next AP it is moving towards is not a BAP. For this scenario,

CAP = BAP

NXAP = Normal AP

$$RSS_{CAP}(t_n) < RSS_{CAP}(t_{n-1})$$

$$RSS_{NXCAP}(t_n) > RSS_{NXCAP}(t_{n-1})$$

$$d(MN_{CAP}) > d(ET)$$

Once again, there is no need for TBVH calculation.

CASE III: Outward MN movement in BAP coverage towards coverage boundary

This scenario considers the case of a MN that is roaming under the coverage of a BAP and is moving towards the boundary with velocity (*v*) as shown in figure 4.11. The inner dotted concentric circle shown in the figure represents the HT circle of radius (r). The conditions are:

$CAP = BAP$

$NXAP = 0$

$RSS_{CAP}(t_n) < RSS_{CAP}(t_{n-1})$

The RSS of any signal picked up by the MN apart from the BAP will be less than a certain threshold RSS_{TH} .

In order to find the TBVH in this scenario we need to calculate the distance z which is the point on the threshold circle where the MN is expected to vertically handover to the higher coverage. As

$$r^2 = d^2 + z^2 - 2dzcosx \hspace{4cm} \text{Equation 4.5}$$

The value of z calculated by solving the quadratic equation 4.5. Therefore, the estimated TBVH for this scenario is:

$$TBVH = \frac{\sqrt{r^2 - d^2 sin^2 x} \pm dcosx}{v} \hspace{3cm} \text{Equation 4.6}$$

In order to accommodate random direction changes and to avoid false handover triggers, while TBVH determined before the MN passes the ET circle can be used for reserving resources in the new network, the parameter's value affects resource allocation in the current coverage only after the MN crosses ET.

CASE IV: Movement of MN from normal AP to BAP

In this scenario, a MN (point C) is under the coverage of a NAP (point A), but is moving towards a BAP (point B) with velocity (v), as displayed in figure 4.12.

$CAP = NAP$

$NXAP = BAP$

$RSS_{CAP}(t_n) < RSS_{CAP}(t_{n-1})$

$RSS_{NXCAP}(t_n) > RSS_{NXCAP}(t_{n-1})$
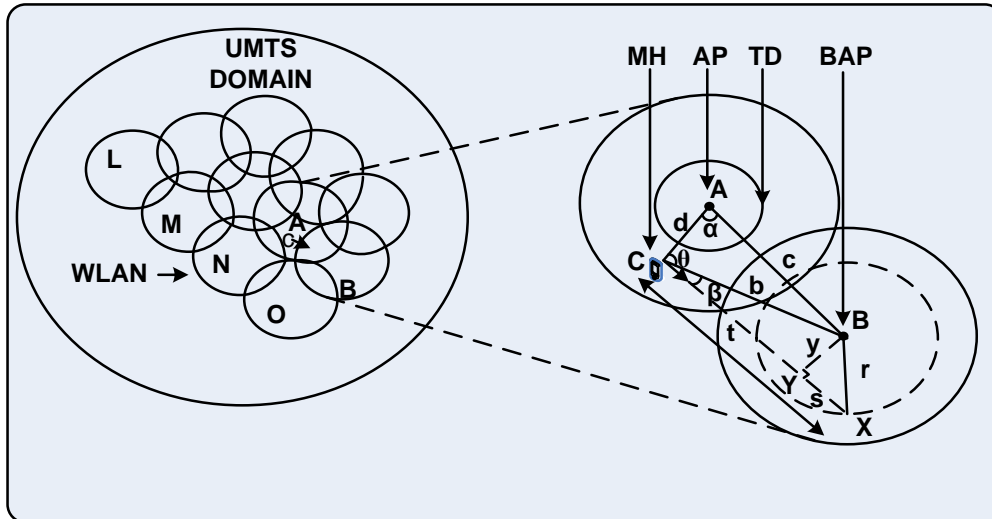
Figure 4.12 Movement of MN in normal AP towards BAP

As the MN is too far from the BAP, depending on which side of line AB point X lies,

$$angle\ \beta = |x - \theta|$$ Equation 4.7

Considering triangle BYC, we have

$$t = bcos\beta$$ Equation 4.8

$$y = b\ sin\beta$$ Equation 4.9

Therefore, in triangle BYX,

$$s = \sqrt{r^2 - b^2 sin^2\beta}$$ Equation 4.10

As

$$z = t + s$$ Equation 4.11

From equations 4.8, 4.9 and 4.10 we have

$$z = bcos\beta + \sqrt{r^2 - b^2 sin^2\beta}$$ Equation 4.12

Thus the TBVH component for this scenario is,

$$TBVH = \frac{\sqrt{r^2 - b^2 \sin^2\beta} \pm b\cos\beta}{v}$$ Equation 4.13

This is similar to the equation obtained in equation 4.6.

4.4.2.2   Deficiencies in the open TBVH model due to physical boundaries

In an enclosed environment, physical boundaries greatly influence the movement patterns of a MN, forcing it to take specific directions. Unlike open environments, a MN located in an enclosed environment cannot rely solely on handover threshold for determining TBVH and demands a new level of intelligence through which it can sense the geographical context. The following scenario highlights the new challenge faced by a MN in enclosed environments.

Consider a MN located in a room and connected to a WLAN as shown in figure 4.13. The room has a single exit and the entire enclosed area falls inside the handover threshold circle represented by the dashed circle, indicating uniform coverage inside the room. The MN in this scenario exhibits random movements characterising pedestrian behaviour which can result in frequent and sudden changes in direction. The figure considers movements of the MN along three different trajectories. Trajectory 1 represents a straight path towards the exit. In this case, the MN stands a good chance of experiencing a vertical handover in the near future once it passes through the exit. In trajectory 2, the
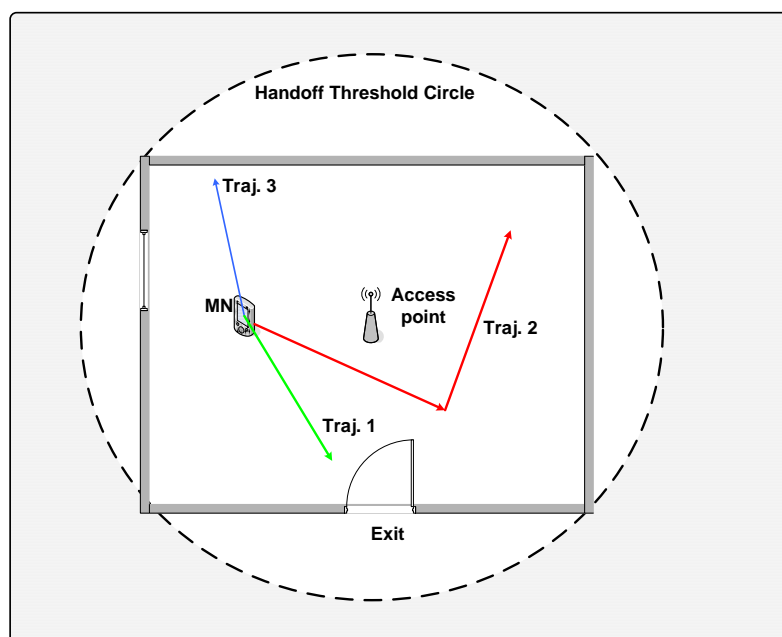
Figure 4.13 False handover triggers due to MN movement within physical boundaries

MN appears to initially move towards the exit but then undergoes an abrupt change in direction and moves away from the exit. In the case of trajectory 3, the MN appears to move towards the edge of the threshold circle but in the direction of a wall instead of an exit which will block any further motion in that particular direction. Hence while the open TBVH model may show a low TBVH value indicating an imminent vertical handover, it is actually a false trigger as in reality the MN can in no way leave the WLAN coverage without changing its direction first. An enclosed area may also consist of multiple exits that need to be taken into consideration. The TBVH value with respect to each exit will be different based on the movement of the MN towards or away from an exit. Not only that, the location of a MN within an enclosed area surrounded by an outer enclosed area can mean further ramifications to the already complex problem. All these factors need to be considered when calculating the time before vertical handover.

All these new issues clearly indicate that for an enclosed environment it is important to develop a handover prediction mechanism that considers the position and movements of a MN through the context of its geographical surroundings and gives a correct indication of the time a MN has before it performs a vertical handover. This study is so far the first one to propose and implement a practical solution which overcomes discrepancies in handover prediction due to geographical confines and provides a quantitative measure of time available before a vertical handover.

4.4.2.3 The Enclosed TBVH Model

The second type of TBVH model developed aims at resolving the issues highlighted in the previous section. This model targets the case of a MN moving inside an enclosed environment which may be indoor such as a building or outdoor, for example a courtyard or garden. TBVH can be predicted with greater accuracy for enclosed environments due to precise definition of coverage from the availability of accurate topological information. These scenarios also facilitate ease in testing in both real-time and simulated systems.

The enclosed TBVH model defines a new parameter called the cosine function based on the direction of the MN with respect to a point of exit. This acts as a weight function which when applied to the calculated TBVH value, assists in eliminating false triggers. The model consists of different case scenarios based on geographical topology, all of which are explained in detailed in this section.

Case I: Enclosed environment with single exit

This scenario explores the case of a MN moving inside an enclosed environment with a single exit which falls inside the coverage of the handover threshold circle of a BAP. Figure 4.14 is the graphical representation of the single exit scenario. The boundary edges and exit are assigned a set of coordinates which are supplied to the MN in the form the *AP_topo* packet by the AP. Equipped with the knowledge of physical confines, the MN then sets on the task of calculating TBVH.

TBVH derivation

Upon observing figure 4.14 it can be noted that the BAP coverage falls under both open and closed coverage categories as the HT circle encompasses both enclosed indoor and boundary-less outdoor coverage. So the first task is to determine where exactly the MN is located within HT coverage. This is done by finding out whether the line created by joining the BAP and MN coordinates intersects with any of the boundary edges. If it does then it means the MN is located outside the closed coverage, otherwise it falls inside closed coverage.

In figure 4.14 the MN takes a random trajectory A-B-C-D-E-F, moving towards or away from the exit until finally it draws close to the exit in section EF. Clearly, calculating TBVH by taking an exit reference point on the handover threshold circle will lead to erratic results as there are times e.g. during section B-C when the MN appears to move towards the coverage boundary but in reality it will
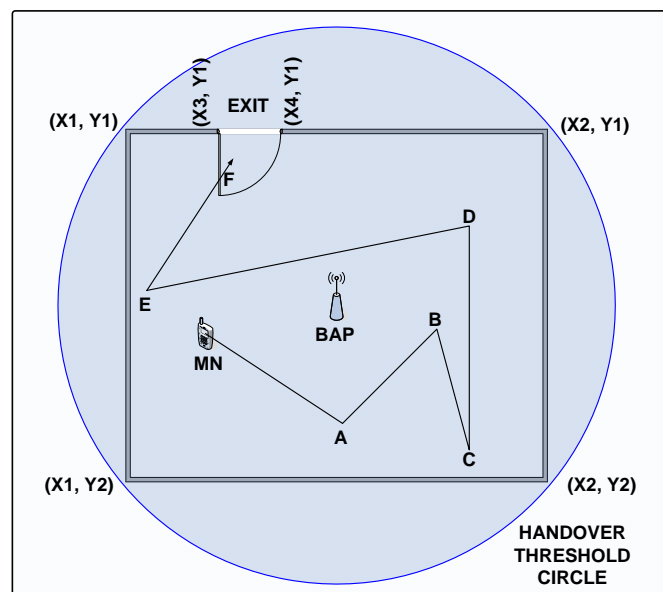
Figure 4.14 Enclosed environment with one exit
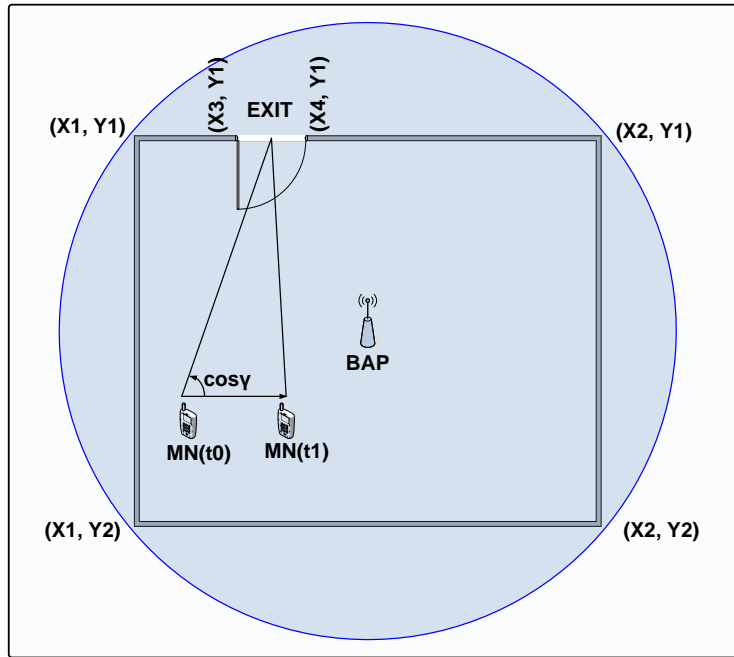
Figure 4.15 The $cos\gamma$ function

be stopped by the wall. Let the device coordinates be BAP (X-bap, Y-bap) and MN (X-mn, Y-mn) respectively. Taking an edge represented by E [(X1, Y1), (X2, Y1)], the task is to find out whether the BAP-MN line intersects with this edge. This is done using the standard formula for determining the intersection point for the two lines [Bourke. 1989]

$$Ua = \frac{(x2-x1)(Ymn-y1)-(y2-y1)(Xmn-x1)}{(y1-y1)(Xap-Xmn)-(x2-x1)(Yap-Ymn)}$$
Equation 4.14

$$Ub = \frac{(Xap-Xmn)(Ymn-y1)-(Yap-Ymn)(Xmn-x1)}{(y1-y1)(Xap-Xmn)-(x2-x1)(Yap-Ymn)}$$
Equation 4.15

where Ua and Ub are the coordinates for the point of intersection for the two lines. In equations 4.14 and 4.15, if

i)   the denominators of Ua and Ub are 0 then the lines are parallel,

ii)   the denominators and numerators for Ua and Ub are 0 then the lines are coincident,

iii)   both Ua and Ub lie between 0 and 1, then the intersection point lies within both line segments.

Therefore the values of Ua and Ub played an important role in determining whether to apply the open or closed TBVH model in a particular situation. TBVH in closed scenarios is the sum of two components t_1 and t_2 where t_1 is the time it will take the MN to reach the exit from a particular point, and t_2 the time it will take it to travel from the exit to the actual handover threshold.

The Cosine Factor
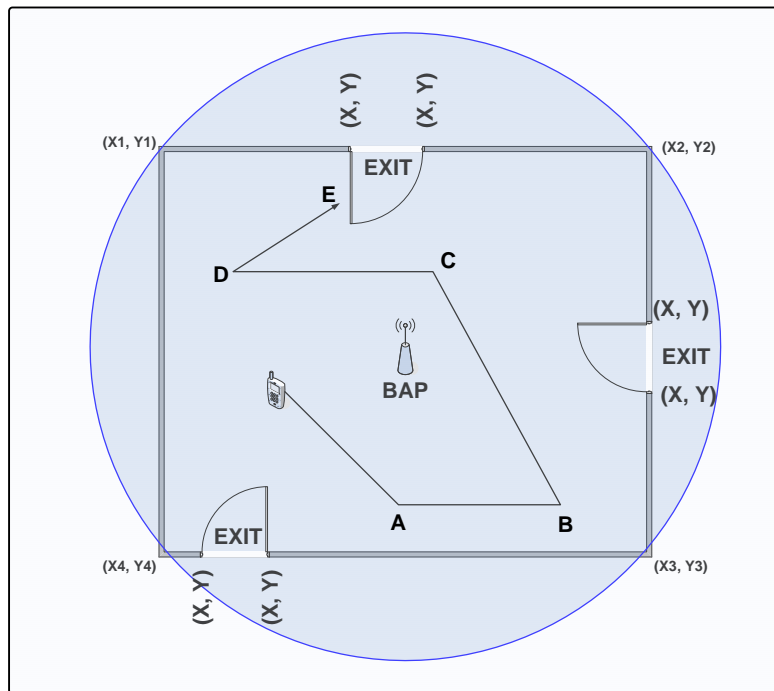
Figure 4.16 Enclosed environment with multiple exits

As mentioned in the previous sections, applying the open TBVH model in closed scenarios can lead to false vertical handover triggers due to geographical confines. Therefore along with the TBVH value, what is needed is another parameter that considers geographical factors along with the MN's movement and assigns like a truth value to the calculated TBVH. The cosine factor is a parameter that complements TBVH in indoor scenarios and plays an important role in the recognition of false TBVH triggers. It is the cosine of the MN direction calculated with respect to a particular point of exit. The cosine value is considered with respect to the exit because it is only after the MN exits through it that it can experience a vertical handover. It is calculated from equation 4.16.

$$cos\ \gamma = \frac{(mid\_distance\ )^2 + (old\_distance\ )^2 - (new\_distance\ )^2}{2 \times (mid\_distance\ )(old\_distance\ )}$$

Equation 4.16

Where

$mid\_distance$ is the distance between the previous MN position and current MN position,

$old\_distance$ is the previous distance between the MN and the point of exit, and

$new\_distance$ is the current distance between the MN and the point of exit.

The closer the cosine value is to 1, the higher is the probability that the MN will pass through the exit.

Case II: Enclosed environment with multiple exits

Figure 4.16 displays a more complex scenario in TBVH derivation. Here the MN moves inside an enclosed environment containing multiple exits. Therefore the MN can move out from any one of these exits. TBVH and the cosine function in this case must be calculated separately for each exit. In the end, when the MN moves close to a particular exit its TBVH and cosine combination will be considered whereas others will be discarded. This is done automatically without any additional functionality as the movement of the MN towards one exit will automatically render the cosine values for other exits negative, resulting in their elimination. All the other calculations in this scenario are similar to those found in the single exit scenario.

Case III: The indoor enclosed environment

This scenario further refines the case of a MN moving inside an indoor area. Figure 4.17 demonstrates the case of a MN inside indoor WLAN coverage such as a building and moving towards the exit. The threshold circle covers both the building and the outer courtyard surrounding it. In this case the courtyard boundary is considered the boundary for the enclosed environment. Thus a MN after exiting the building does not perform an immediate vertical handover but remains connected to
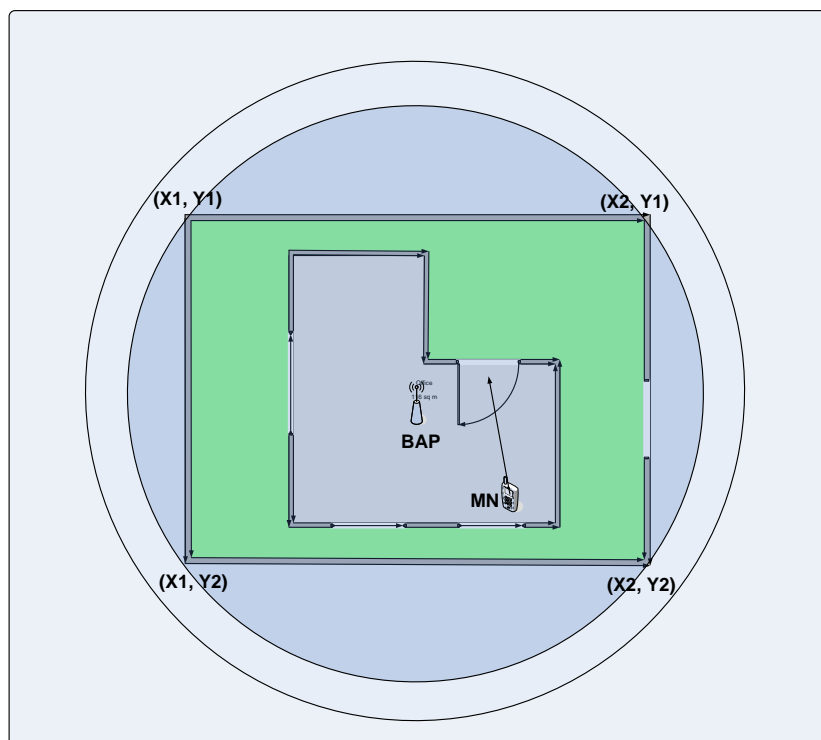
Figure 4.17 Indoor in enclosed environment

WLAN in the outer enclosed area until eventually moving out of coverage at the threshold boundary. This means that as long as the MN is inside the indoor coverage there is no need to calculate TBVH or the cosine function for that matter. The knowledge of this can assist in reducing computational overhead associated with calculations. It can also help to reduce the frequency of location updates at the MN while it falls inside indoor coverage. The intelligence about indoor coverage is supplied to the MN in the *AP_nested_topo* packet after successful authentication and verification. This consists of edge coordinates similar to enclosed coverage and the MN once again checks for the intersection of the BAP-MN line with the edges to find out whether the MN is inside or outside indoor coverage.

4.5   Chapter Summary

This chapter delved deeply into the development of feasible and realistic solutions to tackle the crucial requirement of network coverage availability prediction in 4G networks. It proposed a set of mechanisms for determining the geographical context for a MN, and then explained how this information can be applied along with other MN context information in a unique manner to predict the time before a vertical handover. A set of important conclusions emerged from these proposed solutions. The first is that it is indeed possible for a MN to predict the time it has available to spend in a network before the vertical handover provided it is supplied with a variety of context information about itself and its surrounding context. The second important finding is that it is important not to underestimate the profound effect geographical topology and coverage information can have on all aspects of a MN's behaviour which includes mobility and QoS management, particularly in the event of vertical handovers. This research study is so far the first one that considers network coverage and geographical topology as essential inputs in handover decision.

A noteworthy characteristic of the TBVH parameter is that it is a simple yet meaningful measure that is derived from a large variety of device, network and topological information. TBVH information is simple to understand and it can be applied by a wide variety of device mechanisms that are in need to know the predicted network availability. The next chapter deals with the practical implementation of the theoretical concepts introduced in this chapter along with detailed evaluation and comparison of their performance with RSS-based handover prediction.

CHAPTER 5

TBVH Modelling and Results

5.1   Introduction

This chapter presents the development and simulation of the TBVH-based handover prediction solutions proposed in chapter 4. It also presents and comprehensive comparative evaluation of its performance with the traditional RSS-based techniques. All proposed solutions are modelled in OPNET Modeler 14.5. The chapter begins with section 5.2 which is an introduction on the simulation software OPNET Modeler and specifies the reasons behind its choice for simulating the proposed solutions. Section 5.3 explains the main modelling domains in OPNET and section 5.4 elaborates on the configuration of the evaluation network model to study the behaviour of proposed solutions. Section 5.5 explains the new enhancements made to the standard WLAN node through the introduction of the TBVH module, whereas Section 5.6 focuses on modifications done to the WLAN node's MAC layer to accommodate the exchange of topological information. A detailed explanation of the development and simulation of TBVH models is covered in section 5.7. Detailed simulation results and the performance evaluation of TBVH-based handover prediction solutions with RSS-based handover prediction solutions is covered in section 5.8. Finally, section 5.9 concludes the chapter with a summary of achievements.

5.2   Why OPNET Modeler?

OPNET Modeler is a highly sophisticated simulation software package that enables developers to model communications networks and distributed systems, and allows them to analyse the behaviour and performance of modelled systems through discrete event simulations (DES) [OPNET Training Manual. 1990]. Despite the availability of a large variety of simulation packages in market, OPNET Modeler was chosen as the simulation software due to the following attractive features [OPNET Training Manual. 1990]:

- **Object Orientation:** OPNET adapts the object-oriented paradigm of C++ that allows the creation of objects and classes. Objects inherit the features of classes they belong to. Users are also

allowed to modify these features or add additional features for specific support, or for the development of new objects and classes altogether.

- **Massive communication standards repository**: OPNET users are allowed access to hundreds of standard and vendor specific communication networks, protocols and devices. As many vendor-specific devices have been developed by the vendor committees themselves this ensures a high level of precision in the emulation of their real-world counterparts, adding reliability to obtained results. The readily available standards also help in greatly reducing the time involved in developing simulation environments from scratch, allowing modellers to directly include developed models in their simulations. This is one of the main advantages OPNET offers over other simulation packages such as Network Simulator-2 and MATLAB.

- **High precision in modelling system behaviour**: OPNET provides a large variety of constructs related to communication and information processing, thus providing a high leverage for modelling networks and distributed systems.

- **Hierarchical structure**: OPNET models are hierarchical in structure and closely resemble real communication networks. An object in OPNET may consist of several levels of hierarchy, with each level targeting details for a specific domain.

- **Graphical specification**: In OPNET most of models can almost always be accessed through graphical editors that provide an intuitive mapping from the modelled system to the OPNET model specification. This is another advantage that has resulted in the wide-spread popularity of OPNET.

- **Flexibility in custom model development**: Proto-C is OPNET's high-level and flexible programming language that allows the realistic modelling of all communication protocols, algorithms, and communication networks.

- **Automatic simulation generation**: Models developed in OPNET are compiled into executable, efficient, discrete-event simulations running in C language.

- **Integrated post-simulation analysis tools**: OPNET includes sophisticated data analysis tools for the performance evaluation of simulation results.

These attractive features along with the high-precision simulation ability of OPNET have made it the world's most popular network simulation package with its list of users including many high-profile companies such as British Telecom, AT&T, Cisco, Deutsche Telekom, Ericsson, Lucent Labs, Nortel,

Oracle, Orange, T-mobile, and various US government agencies, to name a few. As the credibility of simulation results conducted correctly in OPNET is accepted in industry, the development and simulation work in this research study was carried out completely in OPNET Modeler 15.0. Moreover, as it was not feasible to conduct the experiments using real multi-interfaced heterogeneous clients due to budget constraints, OPNET provides a relatively cheaper and reliable alternative for model development.

5.3 OPNET's Modeling Domains

This section briefly introduces some of OPNET Modeler's important features which will assist in understanding the methodology involved in the development of the proposed TBVH models. It discusses in particular the three main modelling domains that are responsible for the hierarchical development of all aspects of a network and its devices.

Model specification is the task of developing a representation of the system to be studied [OPNET Training Manual. 1990]. OPNET Modeler contains a set of hierarchically organised editors, designed to capture the various aspects of a modelled system's characteristics. Out of these editors, The Project, Node and Process editors form the main representation of the three domains. The capabilities offered by these modelling domains are a replication of the structures found in real network systems[OPNET Training Manual. 1990]. Further explanation on the three modelling domains can be found in Appendix A1.1.

5.4 Simulated Network Evaluation Model

This section presents a detailed explanation of the simulated network model together with configured traffic that was developed for evaluating the performance of the TBVH models embedded inside a WLAN node.

5.4.1 The Network Topology

Figure 5.1 displays the network topology for the evaluation framework developed in OPNET Modeler. The server side subnet consists of an email and HTTP server connected to the gateway router with 100BaseT links. The client subnet consists of a WLAN access point (AP) connected by a 100BaseT
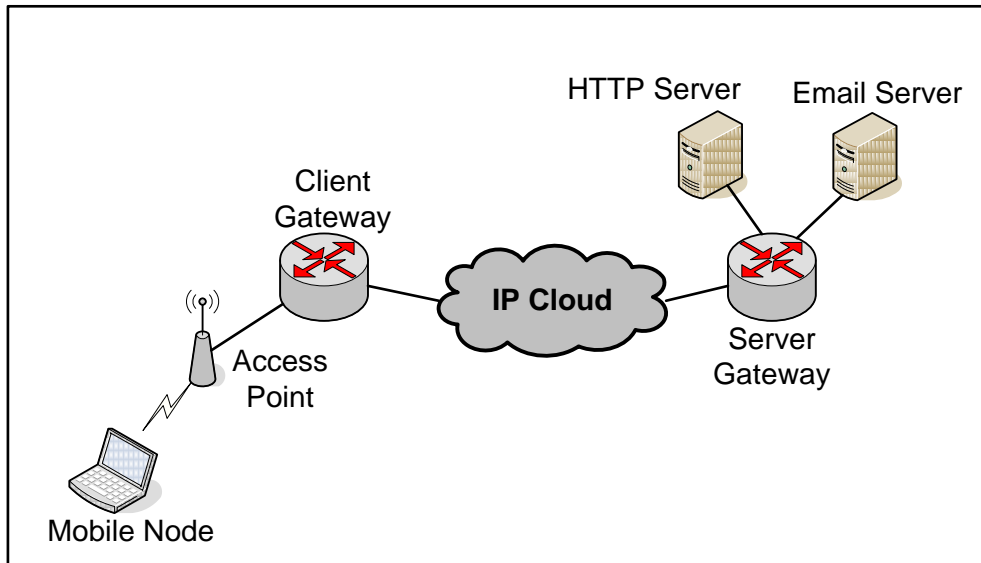
Figure 5.1 Simulated Network Topology

link to the client gateway. Both gateway routers are connected to the IP cloud with Point to Point DS3 links offering data rates up to 44.7 Mbps. The distance between the client and server sites is kept as 4km in order to enable the introduction of WiMAX in future.

5.4.2   The Client Subnet
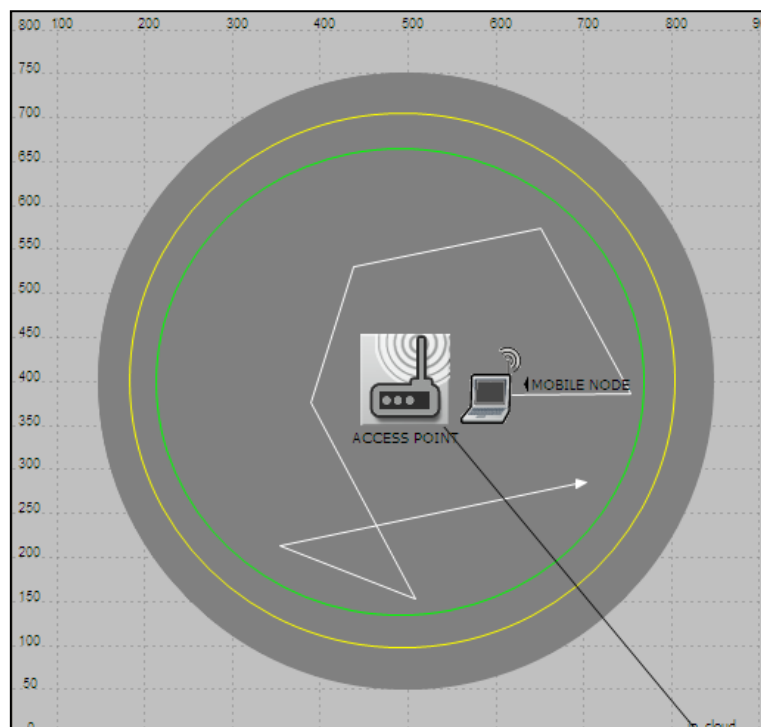
Figure 5.2 The client subnet model

In the client subnet shown in figure 5.2 the WLAN cell's radius is kept at a default value of 300m assigned by the OPNET WLAN model. The MN is assigned a random trajectory on which it travels with a constant pedestrian speed of 0.8 m/s. The grey circle indicates the real coverage, the yellow circle the handover threshold (HT) and the green circle the exit threshold (ET) as explained in section 4.3.1.2. The handover threshold value is initially assigned as 300m.

Once the simulation starts and MNs connected to the AP start performing handovers, after the third iteration the HT value is calculated as an average of the last three available readings. The ET circle is assigned radius (HT-30) m. The AP is configured to assign the Basic Service Set (BSS) ID automatically to any MN within its coverage. The Physical layer technology employed is Direct Sequence Spread Spectrum supporting data rates of 11 Mbps and the channel bandwidth assigned per mobile node is 22 MHz. As the focus of network development is not on achieving maximum data rates within the network through configuration, network characteristics are chosen keeping in mind the need for simplicity in network development and management. The MAC layer access mechanism modelled is Carrier Sense Multiple Access / Collision Avoidance with support for both the Distributed Coordinating Function (DCF) and the Point Coordinating Function (PCF). It also provides support for four different classes of traffic through the Enhanced Distributed Channel Access (EDCA) mechanism, namely, voice, video, best effort and background traffic. MAC addresses are assigned automatically by the simulation kernel. At the Transport Layer the TCP flavour adopted is RENO as it enables fast retransmit and fast recovery. The Duplicate ACK threshold is set to the default value 3 while the maximum number of ACK segments to be sent as 2. On entering the slow start mechanism the window size is set to 1.

5.4.3 Application Traffic Configuration

One of the important goals of the first phase of simulation is the correct configuration of traffic parameters on network nodes together with the correct development of the working network model. It involves the simulation of the Researcher profile on the client node which consists of a combination of heavy web-browsing and light email traffic. This sub-section explains in detail the characteristics for these two traffic types. In the traffic characteristic tables, the description for each attribute is provided in the description column unless the names are self explanatory.

5.4.3.1 Web Browsing-Heavy
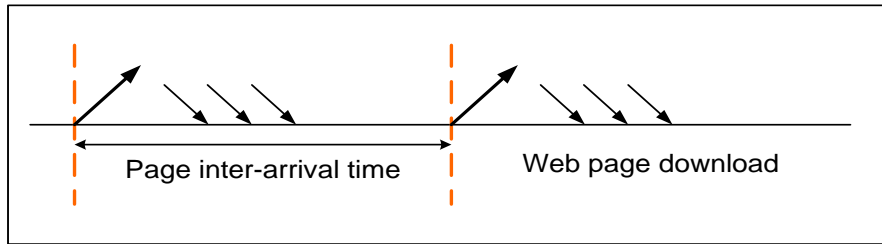
Figure 5.3 Web browsing Application Modeling [OPNET Technologies. 1990]

The HTTP traffic in OPNET Modeler is modelled to be downloaded from a remote server. As shown in figure 5.3, an HTTP page request can result in the activation of multiple TCP connections, one for each object. Table 5.1 lists the main characteristics of this traffic stream.

Table 5.1 Web browsing application characteristics

Characteristic	Value	Description
Name	Web Browsing (Heavy HTTP1.1)	This specified the version of HTTP used.
Page Interarrival Time (seconds)	Exponential (60)	This defined the time between two page requests.
Page Properties		Each row in this compound attribute represented an object found within the webpage.
Page Object		This row represented the page itself.
- Object size	Constant (1000)	
- No. of objects per page	Constant (1)	
- Location	HTTP Server	
Image Object		This row specified the image characteristics.
- Object size	Medium image	
- No. of objects per page	Constant (5)	
- Location	HTTP Server	
Pages per Server	Exponential (10)	This defined the number of pages accessed per website per session.
Type of Service	Best Effort (0)	This defined the priority assigned to the packets of this traffic type by the client.

5.4.3.2 Email – Light
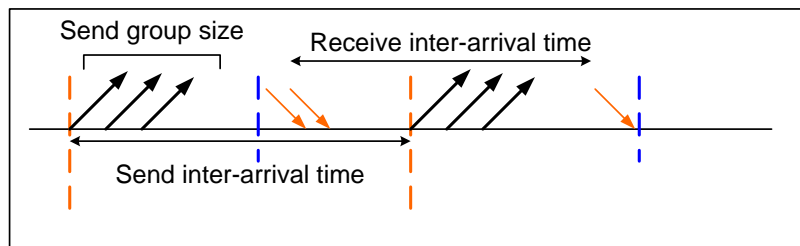
Figure 5.4 Email Application Modelling

Email traffic in Modeller is simulated as a combination of SMTP (Simple Mail Transfer Protocol) and POP (Post Office Protocol). Both protocols use TCP as the underlying transport protocol. Figure 5.4 demonstrates the request-response packet exchange for email traffic and table 5.2 lists the application characteristics.

Table 5.2 Email application characteristics

Characteristic	Value	Description
Name	Email-light	
Send Interarrival Time (seconds)	Exponential (3600)	This attribute defined when the next email was sent with respect to the previous email.
Send Group Size	Constant (3)	This defined the number of queued emails to be sent.
Receive Interarrival Time (seconds)	Exponential (3600)	This specified the time between receiving emails.
Receive Group Size	Constant (3)	This gave the number of queued emails to be received.
Email Size (bytes)	Constant (500)	This defined the size in bytes for an email.
Symbolic server name	Email Server	This gave the email server name.
Type of Service	Best Effort (0)	This defined the priority assigned to the packets of this traffic type by the client

5.4.4 Application Profile Configuration
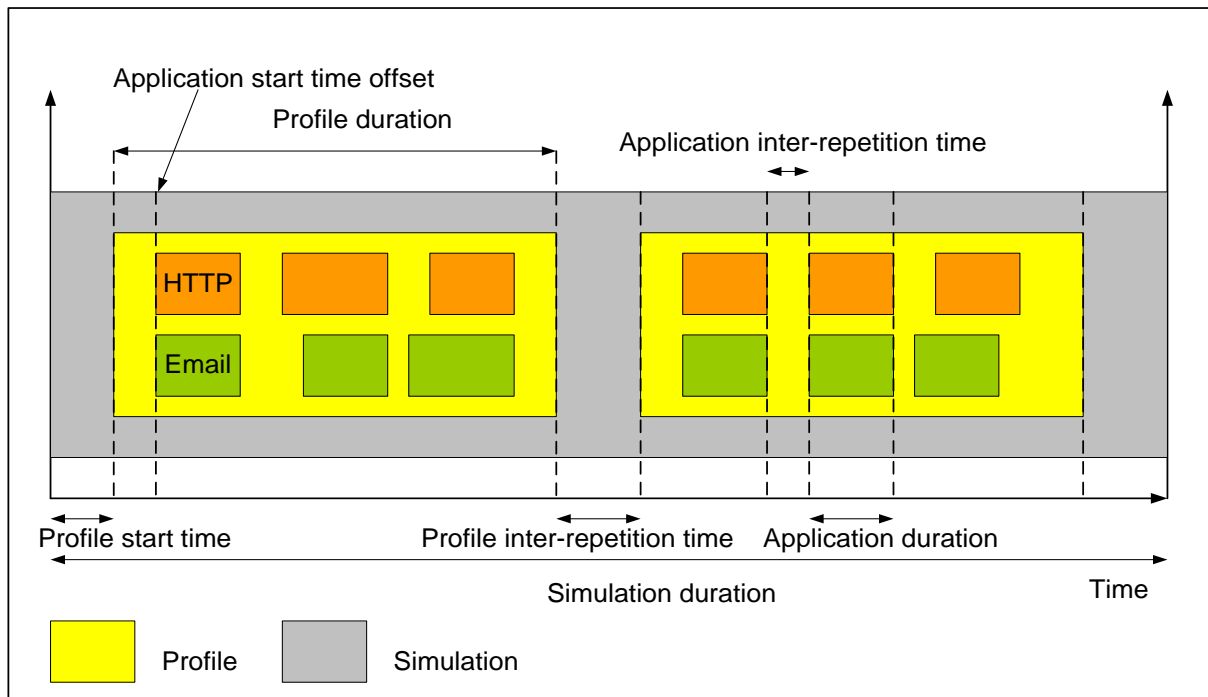
Figure 5.5 Application Profile Diagram

This sub-section deals with the explanation of the next important step which is configuring the profiles for defined applications. Two traffic streams may belong to the same type of application e.g. HTTP but they can differ in their application profiles e.g. the intensity of traffic in different times of the day, the concurrency of user requests, and the repetition in traffic behavioural patterns. The detailed specification of the Researcher profile consisting of HTTP and Email traffic is given in table 5.3. Figure 5.5 gives a simple graphical explanation of the profile information.

Table 5.3 The Researcher Profile

Attribute	Value	Description
Profile Name	Researcher	
Operation Mode	Simultaneous	This means that all other application profiles if present can start at the same time as the Researcher profile.
Start Time	Uniform (100,110)	This specifies when the Researcher profile session starts running on the client after the simulation begins

Table 5.3 The Researcher Profile (continued)

Attribute	Value	Description
Duration (seconds)	End of Simulation	This defines the maximum amount of time given to the profile to run before its end
Applications		This is a compound attribute that defines application characteristics within the profile
- Application Name	Web-browsing (heavy)	
- Start Time Offset (seconds)	Uniform (5,10)	This is the offset of the first instance of the application from the start of the profile
- Duration (seconds)	End of Profile	This is the maximum amount of time allowed for the application session.
- Repeatability		This specifies when the next occurrence of the application will take place
o Inter-repetition time (seconds)	Exponential (300)	This defines when the next session of the application will start with respect to previous one.
o No. of repetitions	Unlimited	This is the number of times the application will be run within a profile
o Repetition pattern	Serial	This means the next session will start after the current one finishes.
- Application Name	Email (light)	
- Start Time Offset (seconds)	Uniform (5,10)	
- Duration (seconds)	End of Profile	
- Repeatability		
o Inter-repetition time (seconds)	Exponential (300)	
o No. of repetitions	Unlimited	
o Repetition pattern	Serial	

5.5   Enhancements to the standard WLAN device node model
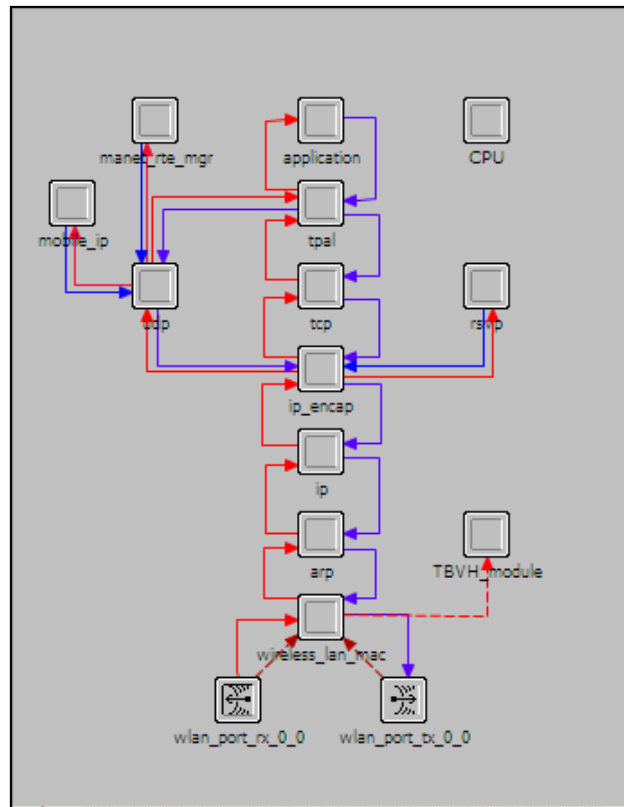
Figure 5.6 The TBVH-aware device node model

Figure 5.6 displays the newly modified WLAN node model of a WLAN client node, equipped with TBVH functionality. The easy availability of the fully functional *wlan_wkstn_adv* node model for the client means that the first phase of modification includes only the introduction of the new TBVH module. A statistic wire originating from the *wireless_lan_mac* module serves as an input for Received Signal Strength measurement triggers in the TBVH module. Along with existing ones, a new set of statistics is defined as follows:

- **Cosine function**: This statistic measures the cosine value with respect to an exit for a MN located inside an enclosed environment. The statistic range is set between 0 to 1. Negative cosine values are ignored.

- **TBVH-open**: This statistic measures TBVH when the MN's trajectory falls in an open environment.

- **TBVH-enclosed**: This statistic records TBVH when the MN moves inside an enclosed environment.

In order to accommodate a BAP whose coverage engulfs both open and enclosed environments, it is important to create two different types of TBVH statistics in order to distinguish between them. Each instance of an exit in an enclosed environment causes the TBVH module to create a new pair of cosine-TBVH statistics for it.

5.6   Modifications to the *wlan_mac* process model

The *wireless_lan_mac* module inside the improved WLAN device model is modelled by a set of process models which provide an in-depth representation of the functionality and behaviour of the complete WLAN MAC layer. The *wlan_mac* process model represents the MAC layer's functionality of Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA).

One of the key enhancements to the *wlan_mac* process is the introduction of topological intelligence in the AP. This involves introducing the proposed topology clusters *AP_info*, *AP_topo* and *AP_nested_topo* within the beacon frame's body field when it is transmitted from the AP to the MN. Topological information entered by the network administrator is stored inside data structures representing clusters and gets bundled out to the MN during the system discovery phase in the connection setup period.
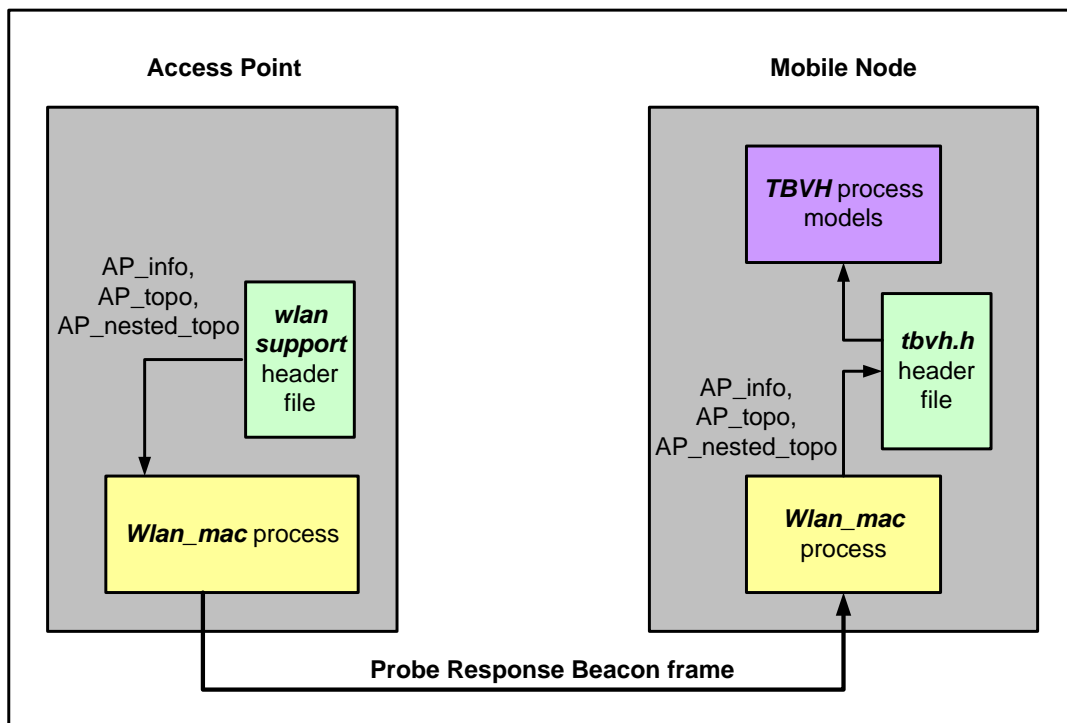
Figure 5.7 Topological information exchange between AP and MN

Due to the immense complexity of the modelled process models representing each OSI layer in the *wlan_wkstn_adv* device model (each layer consists of thousands of lines of code in the process model heirarchy) it is impossible to model the complete encapsulation and decapsulation mechanisms for the proposed clusters at each layer without meandering out of the scope of the study. As this information is mainly required by the TBVH module, a relatively less complex and feasible alternative is to store incoming information at a MN directly in data structures within the header file *tbvh.h* where it is accessible to all child and parent process models. The block diagram flow of topological information from the *wlan_support.h* header file in the AP to the *tbvh.h* header file in the MN is shown in figure 5.7.

The insertion of new topological information involves the modification of code in several key places within different functions in the *wlan_mac* process model. The first modification is in the process model's header file where new topology data structures are declared. A new function *wlan_pk_bbstruct_create()* is then coded for the creation of the new frame body. The proposed data structures are concatenated with the existing frame body's contents. The second modification is at the point where the function decides which MAC frame to create based on existing conditions and requests. It involves populating the contents of the new beacon frame body with the AP and topology information in the function *wlan_prepare_frame_to_send().* The third modification is at the point where the MAC frame is received and decapsulated by the MN. If the node detects this as a Probe response beacon frame it automatically looks in to the frame's body to retrieve topological information. The function *wlan_physical_layer_data_arrival()* is responsible for extracting the *AP_info* and *AP_topo* cluster information and saving it in the structures defined in the module memory *tbvh.h*. The information is now available to all parent and child processes in the TBVH models.

5.7   TBVH Process Modelling

This section provides an explanation of the design and implementation of the process modelling methodology proposed in the development of the TBVH module. The module is introduced at the client side in the *wlan_wkstn_adv* node model. It begins with an introduction and layout of the TBVH process model hierarchy, then goes on to focus on the functioning of individual process models. Simulations results validate the correctness and feasibility of the proposed TBVH concept. An
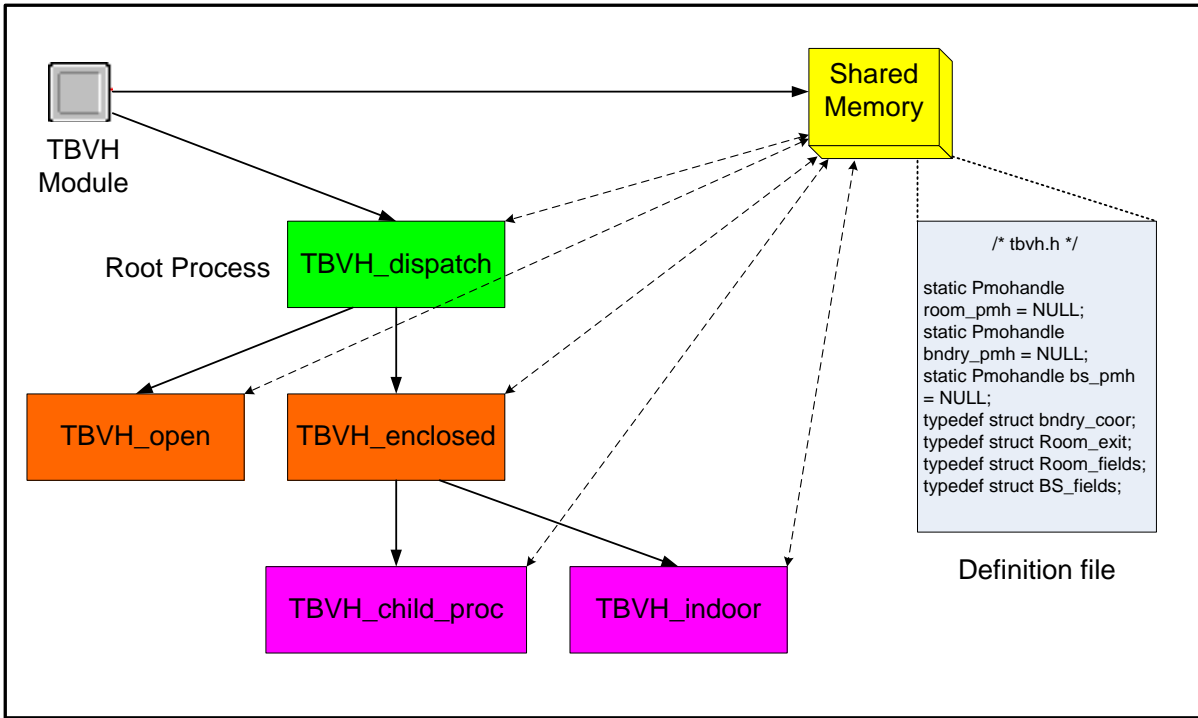
Figure 5.8 The TBVH module's process hierarchy

evaluation of performance of the TBVH-based handover prediction techniques reveals a significant improvement in handover prediction accuracy over RSS-based handover prediction solutions.

5.7.1 The Process Hierarchy

The structural layout of the TBVH module's functional elements is given in figure 5.8. The module is a processor with a hierarchical set of dynamically created process models representing its behaviour. A shared memory block called module memory acts as a common storage location for shared structures. Upon the arrival of the *AP_topo* and *AP_info* clusters at the MN, all the relevant topological and AP related information gets stored in specific structures in the module memory defined as a separate header file '*tbvh.h'.* This facilitates consistency of shared data and the easy referencing of information by all parent and children processes during simulation.

With the invocation of the *begin_sim* interrupt by the root process *tbvh_dispatch* at the start of simulation, the simulation kernel automatically creates an instance of the process. Then based on the information available from the topology clusters on the type of AP and the type of topology, the root process then dynamically creates instances of the first generation child processes – open, enclosed or indoor corresponding to the required TBVH model. Execution of the root process is then

suspended and the created child process is invoked. Upon completion of execution of code in the child process the control is returned back to the parent root process. If the invoked child process belongs to the enclosed TBVH model, it further dynamically creates instances of a second generation of child processes based on the number of exits in the enclosed environment. While the instances of child process models are created at run time, the actual models for all required processes are declared and developed beforehand as part of the process model definition and specified in their respective parent processes. A detailed explanation of the functioning of all developed process models is covered in subsequent sub-sections. All undesired child processes are destroyed once they complete their destined tasks. This helps to avoid the accumulation of unwanted, inactive processes.

5.7.1.1 *tbvh_dispatch* **Process**
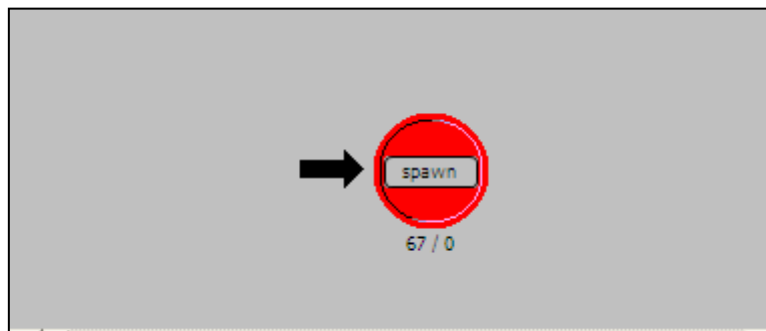
Figure 5.9 *tbvh_dispatch* process
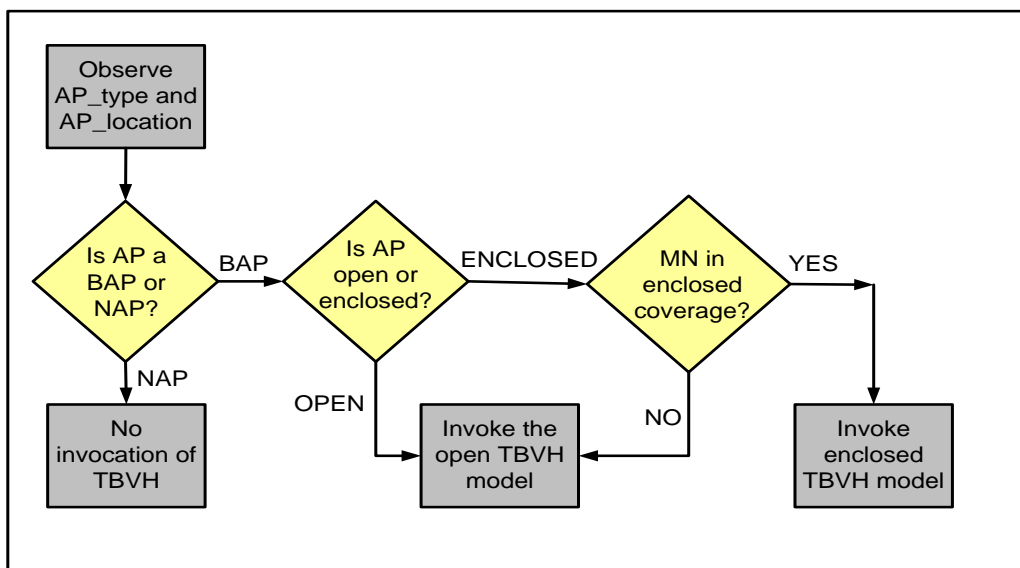
Figure 5.10 TBVH dispatch process decision flowchart

As the root process in the TBVH module (figure 5.9), the main function of the *tbvh_dispatch* process is to scrutinise available information stored in the AP and topology structures in the shared module memory and to choose the correct TBVH model. The process consists of a single forced state called 'spawn'. Based on the values of the *AP_type, AP_location* and *nested_topology* fields, it invokes the respective child TBVH model as shown in figure 5.10. A forced state means that the control is returned back to the simulation kernel only after the complete execution of all code in both the enter and exit executives for the state. The *begin_sim* interrupt is enabled for the process which means that it is invoked as soon as the simulation starts.

5.7.1.2 *TBVH_open_model* Process

The TBVH open model is invoked when a MN is connected to a BAP which provides coverage to an open environment. After calculating the AP-MN and MN(current) - MN(previous) distances, the model invokes the function which determined the point where the MN's line of direction cut the HT circle. These are called the boundary coordinates and are used to calculate the time before vertical handover. Upon the completion of a TBVH iteration the process waits for 5 seconds before re-invoking itself with a self-interrupt for the next iteration as shown in figure 5.11. The waiting time is chosen as 5 seconds in order to be fast enough to capture any sudden changes in MN movements without causing any increased computational overhead which can result from more frequent location measurements.
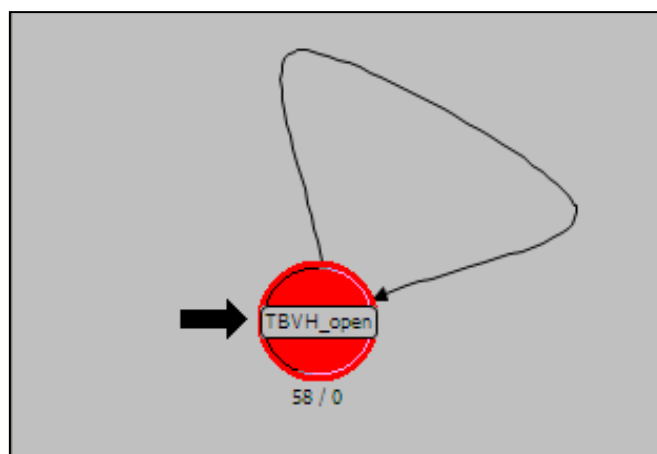
Figure 5.11 The open TBVH model

5.7.1.3 *TBVH_enclosed_model* Process
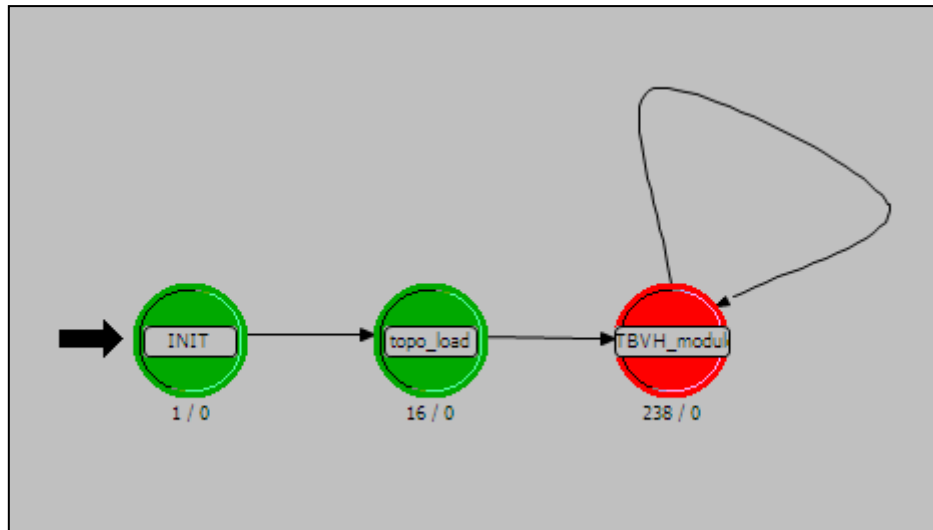
Figure 5.12 The enclosed TBVH model

The enclosed TBVH model process is invoked when the MN starts communicating with a BAP whose coverage falls in an enclosed environment. The State Transition Diagram consists of three states as shown in figure 5.12. Once the kernel control is passed over to this process model, the first state called the INIT state is executed which performs the initialisation procedure. All the required local data structures for storing topology-related information are declared in the function *tbvh_room_init().*

Once initialised, the execution moves on to the *topo_load* state where the room topology is loaded in the declared data structures. The state also creates an array of child processes *tbvh_child_proc* based on the number of exits in the enclosed environment.

After that the execution moves on to the *TBVH_module* state. Boundary coordinates are calculated similar to the open TBVH model. The line intersection formula explained in section 4.4.2.3 is applied to determine whether the line segment made from joining the MN position and boundary coordinates intersects any room edge. If it does then it means the MN is inside the enclosed environment. If not then it means the MN is outside in the open and TBVH is calculated according to the open environment.

If the MN is located inside the enclosed environment and if the topology indicates the presence of an inner enclosed structure, the intersection formula is once again applied to check whether the MN is located inside this inner enclosure. If it is then the *indoor_TBVH* child process is invoked. Otherwise,

117
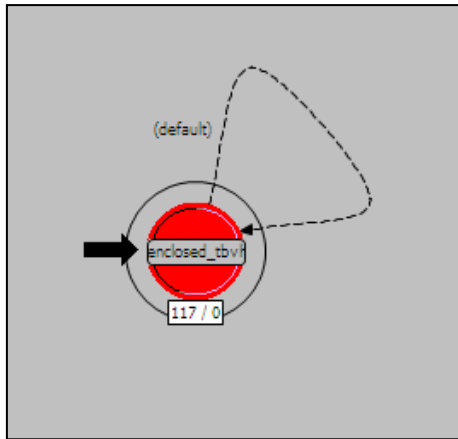
Figure 5.13 Enclosed TBVH child process model

the state declares new pairs of statistics for indoor TBVH and cosine values equal to the number of exits and invokes the same number of instances for the *child_proc_model* process. The self interrupt in this state is once again scheduled for 5 seconds.

5.7.1.4 *child_proc_model* Process

When a MN's trajectory falls inside an enclosed environment, a set of enclosed TBVH child processes are invoked, one for each exit. Once inside the *enclosed_tbvh* state the process calculates the TBVH value and cosine function with respect to a particular exit. When TBVH is calculated for such environments it becomes a combination of the TBVH and cosine values, eliminating the problem of false triggers. Upon the completion of an iteration in each child process, the control is returned back to the parent process which once again performs the condition checks to invoke the correct child processes. This process continues until the MN moves out of the BAP's coverage.

5.7.1.5 *Indoor_proc_model* Process
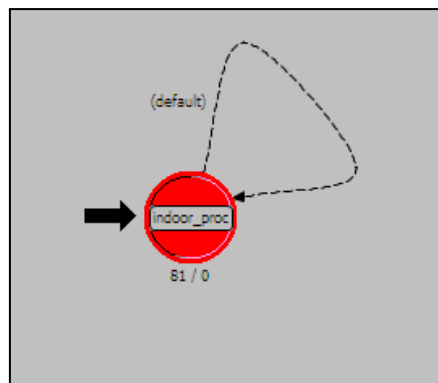
Figure 5.14 Indoor enclosed process model

118

This process does not contain any function for calculating TBVH, it simply checks whether the MN's trajectory falls in double enclosed coverage. If it does then TBVH is not considered until it resides there. Once the MN emerges out of the inner coverage the control is passed back to the parent.

5.8   Simulation Results

This section discusses the simulation results for the models developed in the previous section. After setting up the network scenario as described earlier in section 5.4, a set of simulations is run to evaluate the performance and efficiency of the proposed TBVH mechanisms in predicting vertical handovers.

5.8.1   The Open TBVH Model

This scenario is an implementation of the simulated TBVH model for the open environment. The MN is connected to a BAP with an HT radius of 300m. Keeping in mind pedestrian speed, the ET radius is taken as 270m. The assumption is that the MN is less likely to change its direction abruptly in the last minute. Results from real-world environments [Bernaschi, et al. 2005, Vidales. 2005] have placed handover delays between a range of 1929-4438 ms. The ET to HT travel time is 24 seconds, allowing the MN sufficient time to detect an imminent vertical handover and prepare for it.

In figure 5.15 the MN follows a random trajectory at a speed of 0.8 m/s, changing its direction of motion at various points. The open model means that it is free to exit from the coverage circle at any point. Figure 5.16 represents the graph of plotted TBVH values calculated for the simulated path. Each coloured dot on the graph represents a point on the trajectory when the MN changes its direction, resulting in a change in TBVH as well. Table 5.4 gives the TBVH value at each point on the graph, both before and after change in direction. At point P1 the MN almost crosses the ET circle but then undergoes a change in direction resulting in a large variation in recorded TBVH at that point. This is a typical characteristic of pedestrian behaviour and it is the unpredictability due to these fluctuations that the ET aims to reduce before vertical handovers. A question then emerges whether there is the need to calculate TBVH before ET? The answer is yes. This lies in the fact that a typical heterogeneous MN transmits a large variety of applications ranging from simple data transfer sessions to delay sensitive, live video conferencing sessions. Each traffic type enforces its own
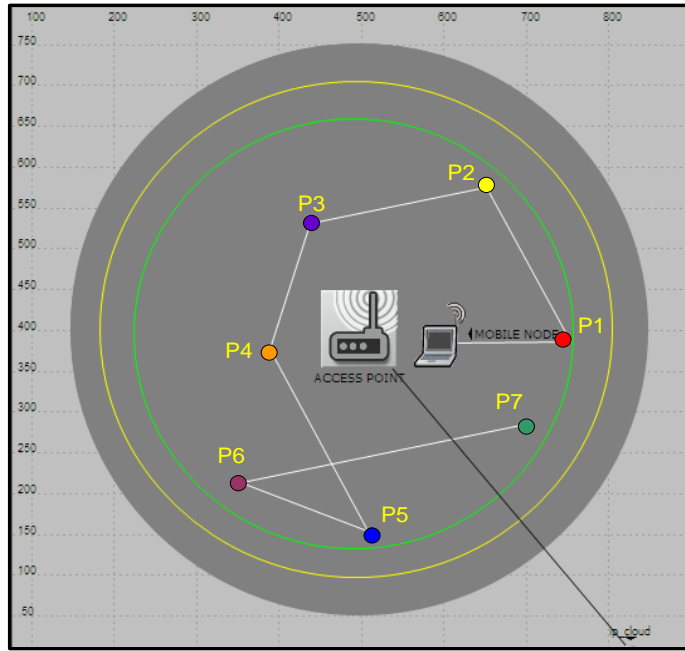
Figure 5.15 TBVH open environment scenario
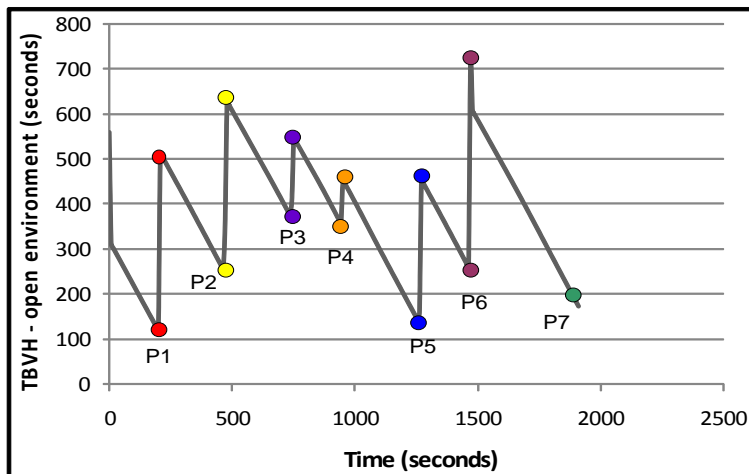
Figure 5.16 Open TBVH graph

Table 5.4 TBVH values at points in graph

Point on Graph	TBVH before direction change (seconds)	TBVH after direction change (seconds)
P1	54	421
P2	168	558
P3	305	481
P4	289	384
P5	75	366
P6	169	643
P7	100	---

unique set of resource and delay requirements. Therefore the knowledge of TBVH is required at different points in time by different layers during the lifetime of a connection.

5.8.2   The Enclosed TBVH model

This scenario captures the TBVH results for a MN's movements in a combination of open and enclosed environment. In figure 5.17 the MN moves inside the enclosed environment for almost the entire duration of simulation before crossing exit 1 towards the end to reach the HT threshold. This means the execution of both the open and enclosed process models based on MN location.

The enclosed environment consists of a room with two exits 0 and 1 with a BAP providing WLAN coverage. Upon receiving the topological information from the BAP during system discovery, the MN applies the available information to calculate TBVH and the cosine value for each of the two exits. Figure 5.18 displays overlapped cosine graphs for exit 1 (E1) and exit 0 (E0) respectively. E1 results are represented by the green graph and E0 by the red graph. It can be observed that the cosine graph successfully captures in a detailed manner the movements of the MN towards or away from the exits. From the start to point P1 the MN moves away from E0 but towards E1, although the direction was not aimed exactly at E1. This behaviour is depicted by negative readings reaching -1 for E0
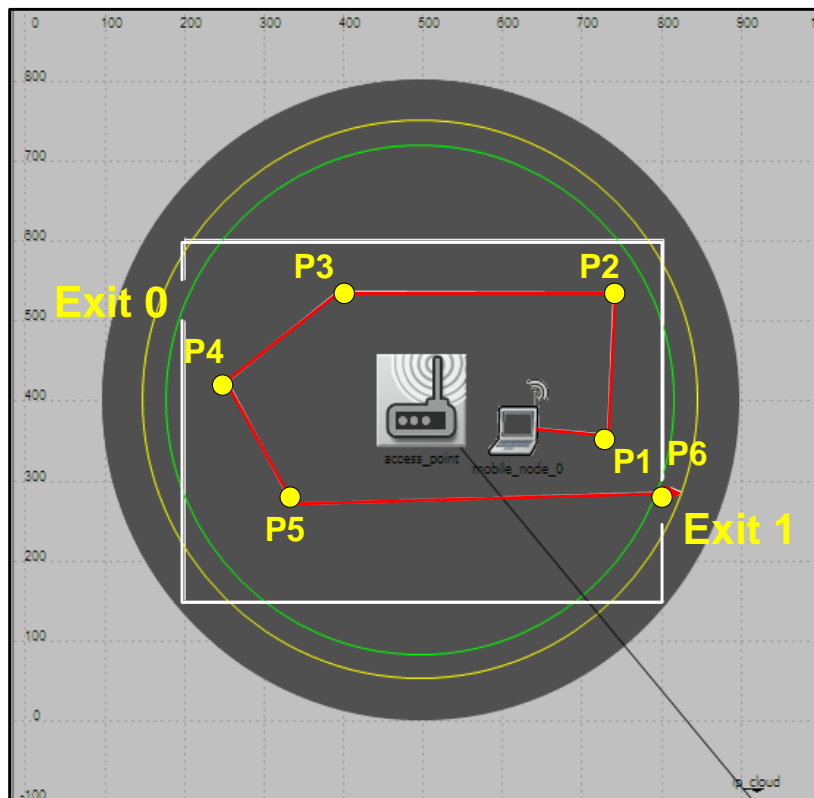
Figure 5.17 TBVH enclosed environment scenario
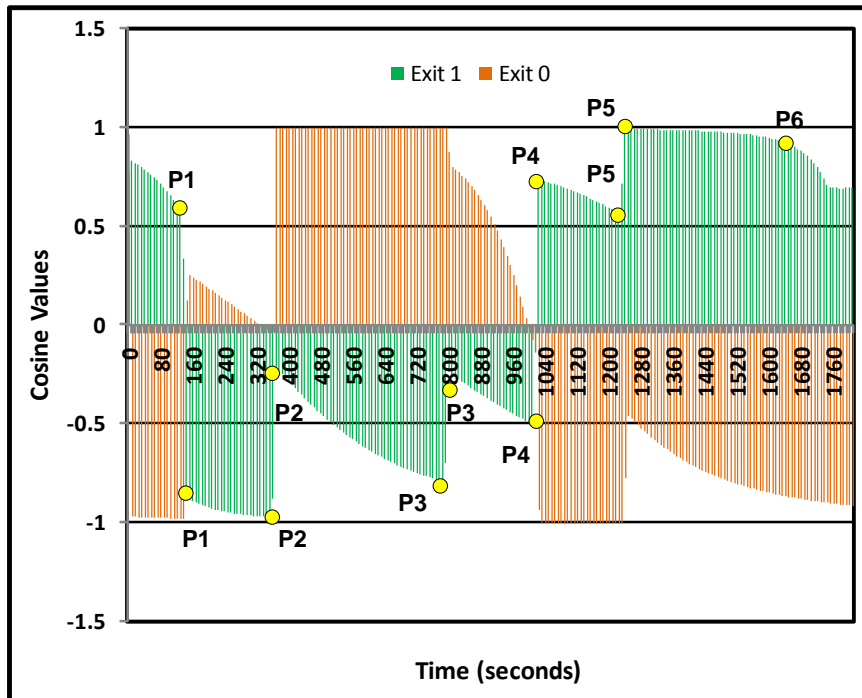
Figure 5.18 Exit 1 Cosine Graph for Exit 0 and Exit1

showing a clear motion away from it. In E1's case the fact that the MN is moving towards but not directly towards the exit is shown by a decreasing positive cosine value. At point P1 the MN changes direction away from E1 hence the green graph records negative readings while cosine E0 goes positive, although the highest value remains below 0.25, indicating that the MN is not moving directly towards it. P2 to P3 is motion towards E0 hence the rise in red values whereas P3 to P4 shows a decrease in both red and green graphs. P4 onwards the MN motion is continuously and directly towards E1 and is shown once again by a rise in the green graph. These behavioural patterns of the cosine graphs prove that the values can be used as a reliable measure for the weight assigned to TBVH when a MN moves inside an enclosed environment.

Figure 5.19 represents the graph of TBVH measured with respect to E1. The next point of interest in the MN's trajectory is beyond point P6 when it moves out of enclosed coverage into open coverage while remaining connected to the BAP. At this point the *tbvh_dispatch* process in the TBVH module invokes the *TBVH_open_model* process to calculate TBVH based on the open model. This behaviour is shown clearly in the graph. The blue line in the graph indicates TBVH calculated based on the enclosed concept whereas the small red line beyond P6 depicts open TBVH. A close snapshot of the
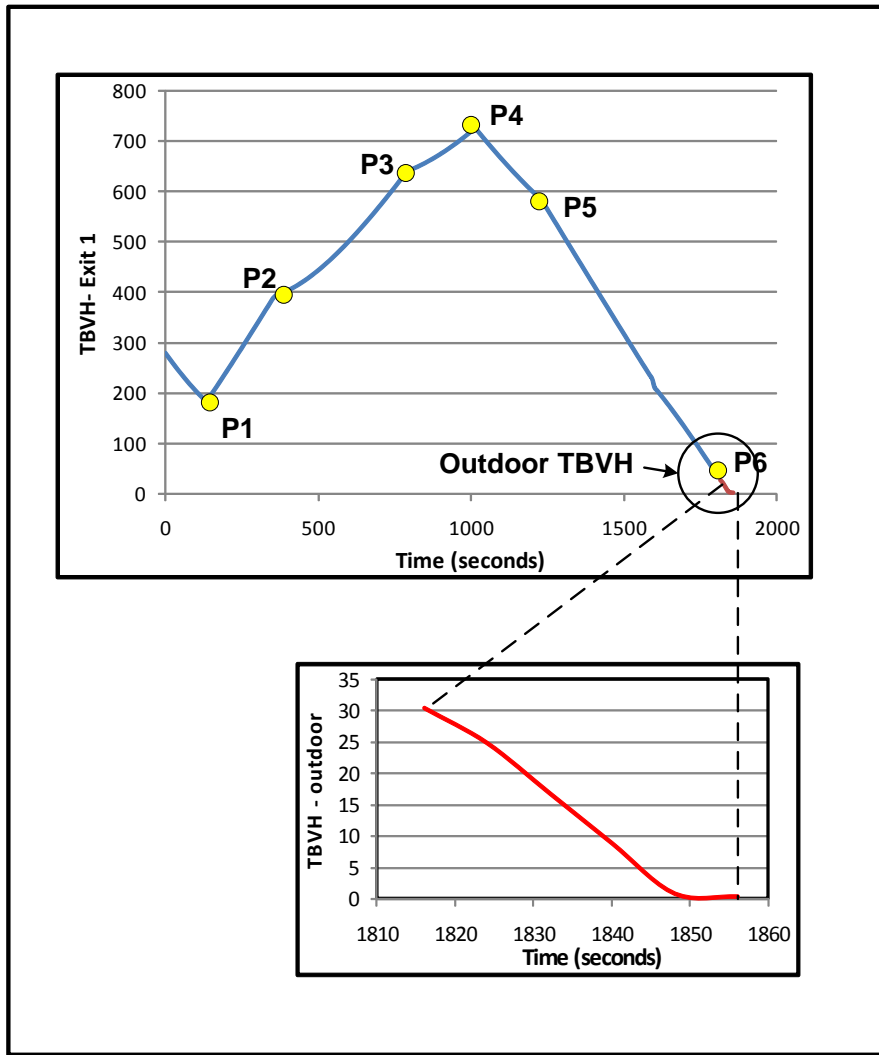
Figure 5.19 TBVH with respect to Exit 1
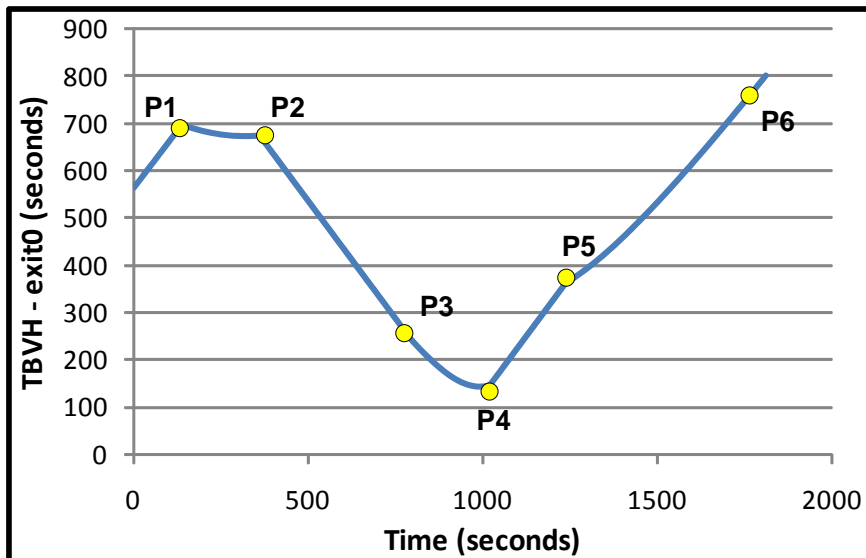
Figure 5.20 TBVH with respected to Exit 0

123

outdoor TBVH line shows decreasing TBVH values which eventually reach zero when the MN reaches the HT threshold. These set of graphs prove that the combination of TBVH and cosine values provides a reliable measure of the MN's time inside enclosed coverage. Figure 5.20 is the TBVH graph with respect to exit E0.

5.8.3   Effect of location error on TBVH accuracy

In order to prove the correct functioning and validation of the proposed TBVH concept, the simulations conducted in the previous sections were based on accurate location coordinate readings. However in a realistic scenario localisation techniques can suffer from prediction errors due to issues such as environmental effects, miscalculations due to unavailability of parameter readings and equipment malfunction. This section discusses and compares the results obtained through the introduction of errors in the location coordinates. The aim is to observe the effect of localisation errors on TBVH prediction accuracy and to investigate the resilience of the proposed algorithms to such errors. The new simulation scenarios are designed to resemble the scenario where a MN suffers from localisation errors for every other reading. Results generated help in discovering some interesting facts about TBVH and cosine functions. Although the conclusions derived from this part of the study are based on
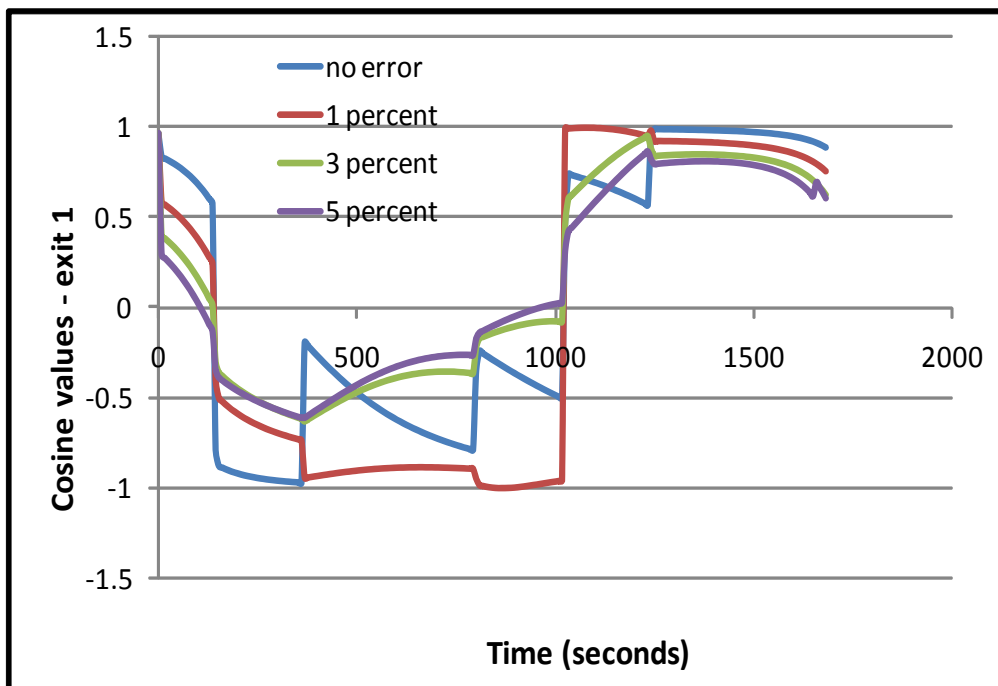
Figure 5.21 Comparison of cosine graphs with varying percentage error

124

Table 5.5 Percent error and corresponding values in metres

Percent error (%)	Deviation from original value (metres)	Time difference (seconds)
1	6	7.5
3	18	22.5
5	31	38.75
10	62	77.5

a large number of different types of graphs generated for different MN trajectories, this section includes a sample each for the main parameters.

Table 5.5 lists the percentage errors in a location reading and how much each error contributes towards the deviation of the new location readings from the original error-free readings. Figure 5.21 shows the comparison graph for different percent error cosine curves. It can be observed that despite different variations in location readings going as high as 30m, the pattern of curves is similar to the original error-free curve. In the cosine graph crests and troughs falling on the same side of the X axis as the original curve. Of course the deviation from the original value cannot be ignored, e.g. at the start of the graph there is a difference of almost 0.6 points between the readings for no error and 5% error. However as they both still fall on the positive side of the axis it indicates a movement towards the exit.
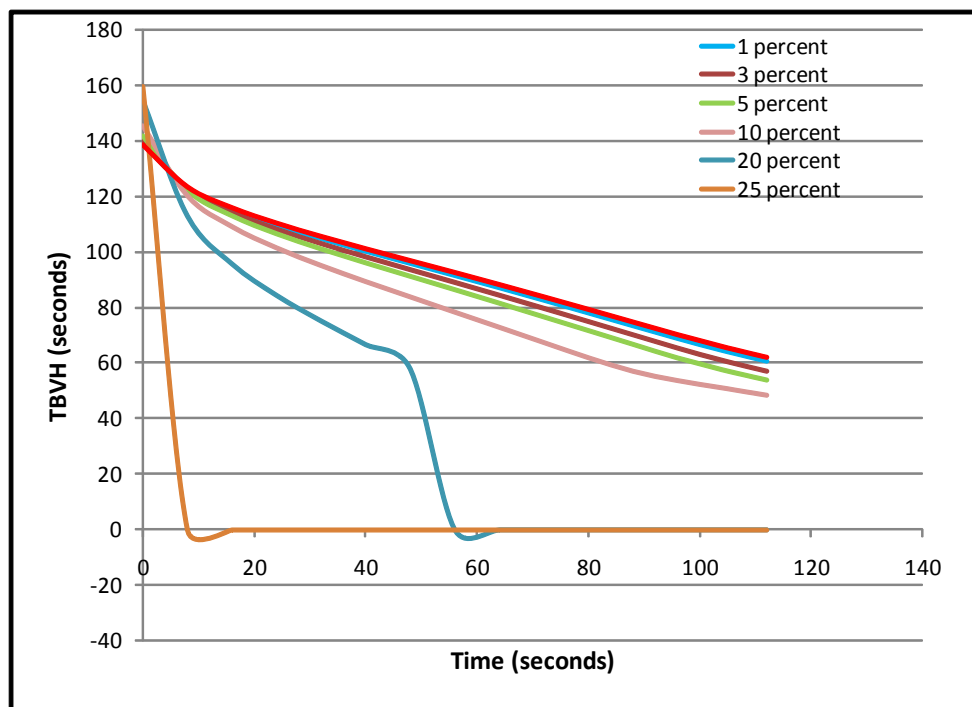
Figure 5.22 TBVH curves corresponding to different error values

The next graph in figure 5.22 compares the different TBVH curves representing the various percent errors. The 20% and 25% error graphs are simply added to depict the limit up to which the algorithms can tolerate errors before the curve coordinates go off the graph limits. An important observation from these results is that all curves until the 5% curve resemble the error-free curve in red with readings falling more or less close to it. Despite the introduction of an error of 30m in the 5% curve, the average difference between the error-induced and error-free TBVH values is 35 seconds which means it can still be accommodated when the device is within the ET circle. The readings still point to an imminent vertical handover, allowing the MN to prepare for it in a proactive manner.

Outside the ET circle a vertical handover can still be predicted however the accuracy of TBVH suffers as the percent error increases. This error can be reduced by reducing the ET radius, which gives the MN more time to spend in the area between the ET threshold and HT threshold, allowing it to adjust to location errors and prepare for the vertical handover. The developed TBVH models thus show resilience to location errors making them suitable for real-world environments.

Having said that, it is important to bring back the attention to the latest developments in the fields of indoor location prediction (section 2.5) where an increasing number of solutions possess the ability to achieve centimetre level accuracy. With location positioning already an indispensible feature in mobile devices and with the commercial deployment of all-in-one wireless solutions like the Ekahau technology [Bing. 2008], one can be optimistic that location errors can be kept below 1%.

5.8.4   Performance evaluation and comparison with RSS-based handover prediction

In a recent study X. Yan and colleagues [Xiaohuan Yan, et al. 2008a, Xiaohuan Yan, et al. 2008b] proposed an RSS-based technique for predicting unnecessary vertical handovers. This study relied on the relationship between RSS and the distance between AP and MN to predict the estimated travel distance between the MN's entry and exit points when it entered a new WLAN coverage. This was a fairly new study whose aim was to rely on RSS thresholds to overcome some key problems found in previous approaches. As it successfully identified some key problems caused due to unnecessary vertical handovers which earlier studies had failed to identify, it forms a benchmark for evaluating the performance of the proposed TBVH-based handover prediction method (referred to as TBVH-M) against the RSS-based handover prediction method (referred to as RSS-M).

5.8.4.1 Dynamic prediction of random mobile movement patterns
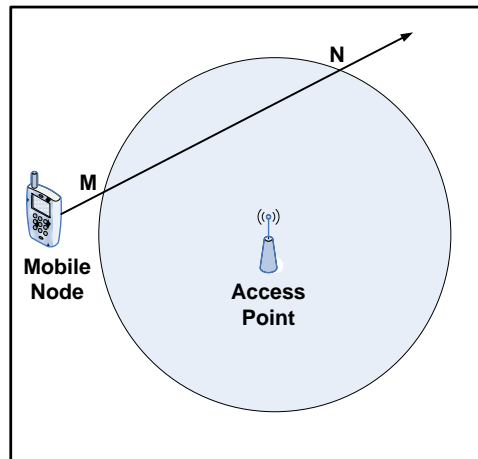
Figure 5.23 Simulated scenario for Yan's RSS-based handover prediction study

The experimental setup for Yan's study is as shown in figure 5.23. The WLAN AP's real coverage is assumed to be circular. When a MN reaches the point of entry M the proposed algorithm utilises the RSS value to predict the AP-MN distance. This distance is then used in geometric calculations to find out the point of exit N. The estimated travel distance between the two points gives an estimate of the time the device is likely to spend in WLAN coverage. In order to compare the performance of the proposed TBVH-M the same experimental scenario was setup and simulated in OPNET Modeler using the new TBVH aware nodes and the performance of the two methods was compared. The new set of simulation results unravelled some important deficiencies with RSS-M. The study had set out to resolve the issue of unnecessary vertical handovers however due to the uncertainty and randomness associated with RSS fluctuations some key issues have not been resolved. TBVH-M on the other hand demonstrates a remarkable improvement in performance in the prediction of unnecessary vertical handovers.

The first deficiency with RSS-M is the technique itself. In order to predict the exit point for calculating the estimated travel distance for an incoming MN it is crucial to know the entry point. Which means that the solution successfully calculates the exit point at the first instance of entry provided the MN does not change its direction of motion. If upon travelling in to the WLAN cell the MN undergoes a change in direction it is no longer possible to find out the exit point as RSS measurements from a single AP alone cannot reliably pin point the exact location of the MN in the cell. This is a common

127

problem found in other pure RSS-based approaches such as [McNair, et al. 2004, Kassar, et al. 2008, Ebersman, et al. 1999].

As demonstrated in previous results, TBVH-M does not suffer from this deficiency and no matter where a MN is located in the cell coverage, it can easily determine how long it has left before a vertical handover. This is also one of the key achievements of the TBVH based handover prediction mechanism over the history-based vertical handover prediction techniques reviewed in section 3.2 as it dynamically captures random MN movements and accurately predicts changes in direction in a particular coverage without relying on old and unreliable stored information. It also does not cause any additional movement prediction related information exchange overhead between the AP and MN as the MN is solely responsible for doing its own calculations. TBVH-M also demonstrates a higher precision in prediction over all discussed history-based prediction techniques as it not only successfully predicts the next cell but goes a step further to successfully predict the time available in the current cell's coverage, taking into consideration topological factors to prevent false triggers. This research study is so far the first one to propose a comprehensive vertical handover prediction solution that considers such a broad variety of information yet applies it in a simplified manner which prevents high computational overhead.

5.8.4.2 Intelligent identification of coverage boundaries

The second deficiency identified in Yan's RSS-M was that it did not specify any method to identify coverage boundaries so it is difficult for the MN to know whether it is exiting just an AP's coverage or the WLAN's coverage. TBVH-M accommodated this issue in a simple yet effective manner through the classification of APs in to boundary APs and normal APs which automatically alerted the MN about approaching WLAN coverage boundaries.

A key advantage of TBVH-M over the coverage-based handover prediction techniques discussed in section 3.3 is that it does not rely on a large amount of coverage information to predict coverage boundaries but instead the whole process is carried out in a more distributed fashion where each AP is responsible for identifying and conveying the MN information about its own coverage and the coverage of any immediate neighbours. This solution does away with the huge time and resource overhead associated with the data gathering phase in coverage based handover prediction. As a MN
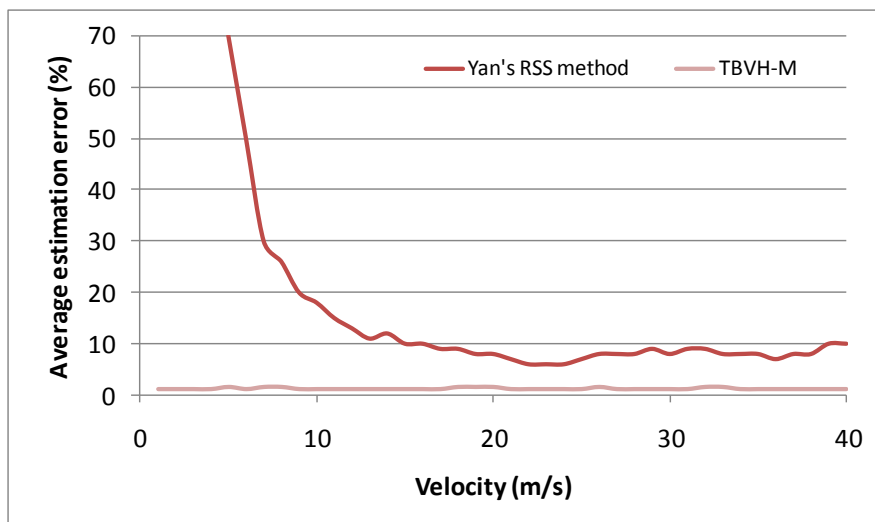
Figure 5.24 Behaviour of estimative error curve for RSS-M and TBVH-M with respect to change in velocity

only needs to be aware of coverage characteristics of networks it is immediately concerned with, it simply obtains this information from the APs to which it attaches itself.

5.8.4.3 Effect of pedestrian speeds on handover prediction accuracy

A third and more serious deficiency with Yan's RSS approach is the failure to predict unnecessary vertical handovers for pedestrian speeds. As shown in figure 5.24, while the estimation error for travel distance for device speeds more than 15 m/s is around 10%, the estimation error for device speeds less than 5 m/s dramatically increases to more than 70%. This problem cannot be ignored because the solution targets WLAN specifically where the majority of users exhibit pedestrian movement patterns. In fact this exact problem associated with large errors due to drastic RSS fluctuations in WLAN environments was identified and acknowledged by Cottingham [Cottingham. 2009] however the issue was not pursued further as it fell out of scope.

As for the improved performance at higher speeds, the issue is debatable as Yan's study does not specify whether RSS sampling frequency was increased with an increase in speed. High error rates at lower speeds can be attributed to the RSS sampling rate as a MN that spends more time in the cell coverage naturally records more values whereas a MN travelling at a high speed does not get the time to collect enough samples which automatically bring down the error rate. TBVH-M however shows no such problems and right from the instant the MN enters the cell coverage until its moment of exit, TBVH is calculated successfully for a large variety of speeds as shown in the graph in figure
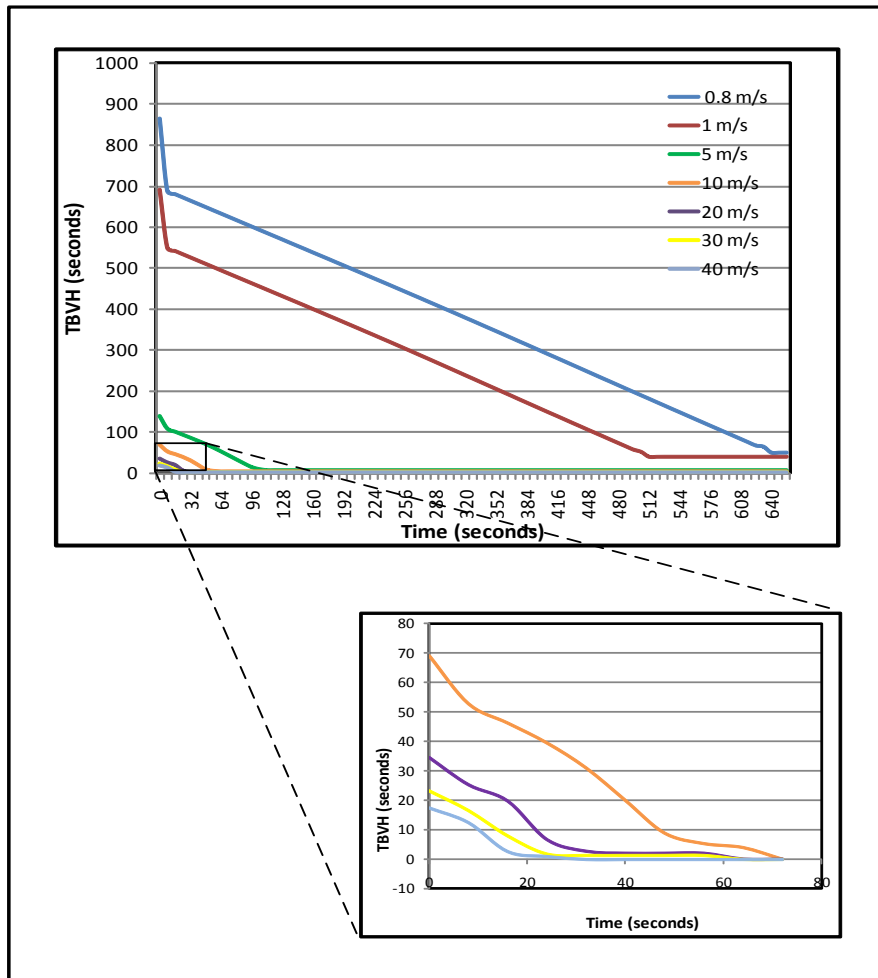
Figure 5.25 Effect of MN speeds on TBVH measurements

5.25. The location update rate for all trajectories in this instance is kept at 8 seconds in order to avoid computational overhead. However an increase in the sample rate simply means a more detailed version of the graph which provides a more refined knowledge of TBVH.

5.8.4.4 Effect of positioning error on TBVH performance

In another set of measurements in TBVH-M, the MN's speed is kept constant at 5 m/s and a range of errors are introduced in location readings. The estimated distance to the exit point is then calculated for each error rating as shown in figure 5.26. The average deviation from the error-free values for each curve is given in table 5.6. Once again the 20% and 25% error curves are plotted to examine the error limit that can be accommodated. The maximum percent error value is 25% which means that an estimation error of 70% obtained in RSS-M will fall completely off the graph due to negative distance values! In Figure 5.24, a realistic consideration of 1% positioning error for TBVH-M still results in a

130
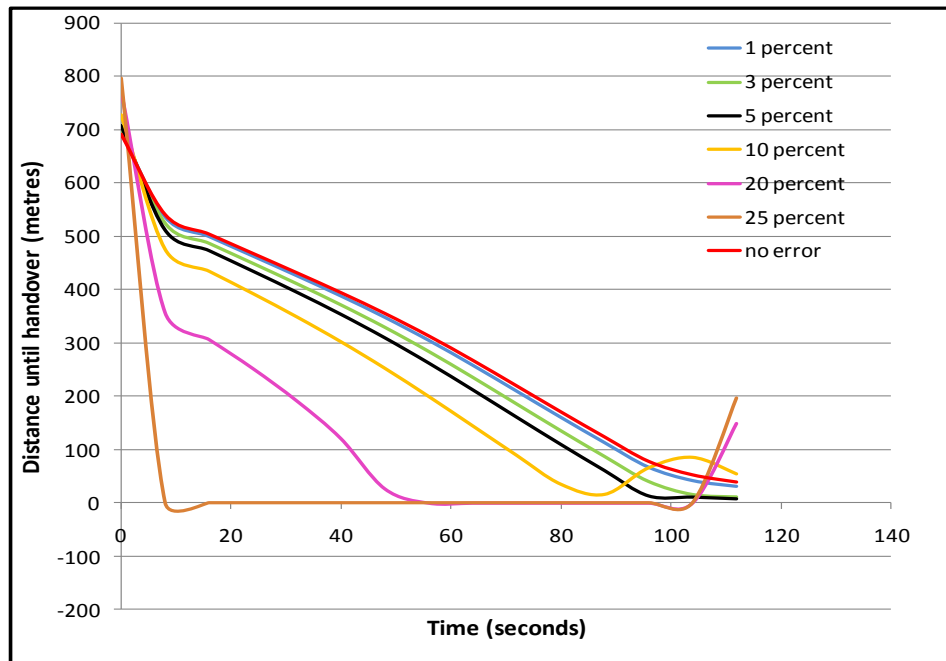
Figure 5.26 Effect of location errors on estimated distance until handover

Table 5.6 Average deviation from error-free readings for different percent errors

Percent error	Deviation of distance (metres)
1%	7
3%	25
5%	42
10%	76

graph that clearly shows that the method successfully captures a wide range of speeds from 1 – 40 m/s.

5.9   Chapter Summary

This chapter presented a practical implementation of the TBVH-based vertical handover prediction theory proposed in chapter 4. The main goals consisted of the validation of proposed concepts through extensive simulations in OPNET Modeler and a demonstration of the successful implementation of the TBVH concept. The work in this chapter consisted of three important parts:

1. Development of a new TBVH module which consisted of a set of processes for modelling the different TBVH models for open, closed and indoor environments.

2. Enhancement to the Probe beacon packet exchange mechanism at the MAC layer to accommodate the introduction of topology-related and coverage-related information.

3. Simulation of developed models and an evaluation of performance together with its detailed comparison with traditional RSS-based handover prediction methods.

Simulation results agreed with intuition and TBVH was successfully derived for all the three targeted environments. The performance was then compared with the RSS based handover prediction technique and TBVH demonstrated significant improvement in performance over the RSS based method in all areas of comparison. A key outcome of the comparison study was the clear proof that RSS alone can no longer be employed as a reliable measure to predict vertical handovers in wireless heterogeneous networks. Results also demonstrated how predicting coverage and calculating time before vertical handover is actually a complex task that requires increased intelligence about coverage, topological factors and device context, something the TBVH parameter successfully achieves and then provides a precise measure of the time before vertical handover. The key strengths of the TBVH mechanism that emerged during performance evaluation are:

1. The ability to provide a clear, quantitative measure of the time a MN had available to spend in a given WLAN coverage before performing a vertical handover, irrespective of its location and movement within the WLAN coverage.

2. The ability to process and convert a large variety of complex context information on device, network coverage and geographical topology into a single parameter TBVH which provides different device layers with a simple measure of the availability of current and future network coverage.

3. The dynamic, client-based prediction of random movements and change in direction for a roaming MN.

4. The spontaneous identification of network coverage boundaries facilitating the accurate prediction of an imminent vertical handover.

5. The ability to accommodate a large range of MN speeds, pedestrian speeds in particular without suffering degradation in performance due to reduction in prediction accuracy. .

This research study so far has focused on the need and derivation of TBVH. The next part of study focuses on the area of downward QoS management in heterogeneous clients. Along with the development of a novel set of mechanisms for the autonomous management of tasks like network selection, resource management and call admission control, the second part also demonstrates how the knowledge of TBVH plays an active role in improving the intelligence of these mechanisms.

CHAPTER 6

Related Work – Network Selection in Heterogeneous Clients

6.1   Introduction

In 4G QoS management one of the crucial tasks is selecting the correct network for an application stream. After that come other related tasks like call admission control and resource allocation. The main purpose of this chapter is to provide a comprehensive literature review of related work conducted in the areas of context-aware, multi-criteria network selection. It summarises the achievements of some previous research works and identifies key deficiencies in existing approaches that have not been addressed so far.

6.2   Network-driven versus client-driven interface selection

In wireless networks, network availability becomes a prerequisite for conducting other parameter checks. A network may offer all the required services a MN desires, but if its availability is limited to only a few seconds, it cannot form a wise choice of selection as it will lead to unnecessary vertical handovers. Many solutions have been proposed in literature to tackle the issue of increased complexity in traffic management in heterogeneous networks. These strategies can fall broadly under the category of either network-based or client-based network selection strategies.

In network-driven interface selection strategies, the responsibility of selecting the correct interface for a particular traffic stream falls mainly on the network components like the BS or other specialised modules. These components perform decision-making on behalf of the MN and hence need to be constantly supplied with context data about the MN's active interfaces. This however can cause a significant increase in network overhead, similar to network-controlled handover management solutions. Network-assisted selection strategies can assist the MN in decision-making as long as their role is limited to supplying it regular information about the networks' status.

In client-driven interface selection strategies, the multi-interfaced MN plays a more strategic role in the selection of the most appropriate network interface for a particular application. Being directly in touch with all available network interfaces, the MN is intimately aware of their medium access, network and transport conditions and can promptly respond to any sudden changes in channel conditions. Acknowledging the superior role of the client in heterogeneous environments, an increasing number of studies have adopted the client-based approach for network selection and vertical handover related decision-making.

Based on the strategies employed in selecting the most suitable interface for a particular traffic stream, network selection solutions can be classified as policy-based, multiple attribute decision-based and fuzzy logic based network selection strategies. Other stray techniques may be found in literature but due to the vastness of the topic this chapter limits itself to only those solutions which are closely related to those proposed in this study.

6.3   Policy-based network selection (with cost functions)

Policies are defined as rules which govern actions [Lewis. 1996]. Policy based resource management in distributed systems has been widely adopted in many different fields and has continued to develop as a solution for monitoring and controlling resources [Moffet, et al. 1993]. In heterogeneous networking, policy-based solutions were mainly adopted in the areas of vertical handover management and QoS management as a means for achieving ubiquitous networking [Hadjiantonis, et al. 2007, Vidales 2005, Verma. 2002]. Some studies went a step further and applied the concept to network selection as well. However a common problem with static policy-based solutions is that they are generally pre-programmed to react in a certain manner to a given set of issues and are usually not able to dynamically accommodate new conditions.

Wang et al [Wang, et al. 1999] in their work developed a reactive policy-enabled handoff system which allowed users to define the 'best' network at any given time while making tradeoffs among network characteristics such as cost (C), bandwidth (B) and power consumption (P). Network selection was based on the weighted cost function

$$f_n = w_b \, . \, ln\frac{1}{B_n} + w_p . \, lnP_n + w_c . \, lnC_n \ \ (\textstyle\sum w_i = 1)$$ Equation 6.1

where the normalized value for each parameter was multiplied by its the weight. This was calculated for each network and the function with the lowest value was selected. This procedure was repeated at regular intervals. Stability period was the minimum duration for which the new network had to consistently display the lowest value of f_n in order to make it better choice for handover. Chen et al. [Chen, et al. 2004] further refined the stability period concept proposed by Wang by introducing adaptability in it. The utility function of a wireless network was the sum of the product of network parameters and their assigned weights.

$$Utility_j = \sum_i w_i \times f_{i,j}$$ Equation 6.2

The utility ratio was the ratio of target network utility by current network utility. The authors then took the ratio of the two utility ratios measured consecutively. The duration of stability period then increased or decreased dynamically based on the decrease or increase in utility ratio.

Both Wang and Chen's studies focused on network selection from an optimised handover point of view. They were reactive schemes which did not take different application requirements into consideration. In a heterogeneous wireless scenario there may be more than one network clearing the stability test when a MN roams under the overlapped coverage of several networks, displaying no need to perform a vertical handover. Moreover, the QoS requirements can change with different service classes and user preferences [Song, et al. 2005]. Applying the knowledge of a requesting application's preferences when selecting the network definitely helps in refining the choice, unfortunately both studies adopted a simplistic method for network selection without taking this factor into consideration.

Murray et al. [Murray, et al. 2003] developed policy mechanisms which addressed access control and network selection issues at both the client and network sides. In this study, mobility management policies were based at the client side and network selection policies at the network side. The study also attempted to address the issue of deciding the correct time to perform a vertical handover based on network QoS parameters available at the client. The drawback of this approach was that it was network-controlled. In the real-world scenario of a loosely coupled heterogeneous environment it is difficult to assign complete control to any single BS for tasks like network selection due to the presence of multiple autonomous service providers who might be unwilling to share information about

their network's performance with competitors. Even if this was made possible, the continuous transfer of a large amount of context information on all active points of attachments will still result in an unnecessary increase in traffic overhead, rendering it unsuitable for vertical handovers.

Zhu and McNair [McNair, et al. 2004] also proposed a policy-based scheme for allocating sessions to different networks. This study again was limited to optimising traffic allocation during vertical handovers by applying a cost function which measured the benefit associated with handing off to a new network. This study did not take user preferences into account.

In another effort [SungHoon Seo, et al. 2007], the authors proposed a reactive policy-based scheme where the choice of a network interface for performing a vertical handover was based on four types of policies – maximising throughput, maximising number of bits per energy, minimising energy consumption, and adaptive configuration. All types of traffic classes were considered and a client-based solution meant the device could independently choose the most suitable interface. However, the deployed network architecture was tightly-coupled and the study focused only on data transmission on the downlink from base station to MN. Being reactive in nature, the proposed solution could not predict future changes in network conditions.

Another study [Marti, et al. 2007] suggested a QoS policy-based access management schema for heterogeneous wireless networks were service requests were granted based on user preferences, service requirements, network conditions and user activity history. Once again the network was responsible for gathering context information and network selection. It thus faced the same problem of increased network overhead as Murray's approach.

In a recent study, Stevens-Navarro and colleagues [Stevens-Navarro, et al. 2008] advocated the use of deterministic Markovian decision process in a policy-based solution aimed at maximizing the total expected reward per connection. Policies consisted of a sequence of Markovian decision rules which suggested a set of actions to be performed in each state. Network selection was based on bandwidth and delay and handover decisions were triggered mainly by measuring RSS. This approach did not take into consideration topological information and application requirements were defined by simple weights. The issue of user preferences for networks was not taken into consideration, neither was the issue of unnecessary vertical handovers.

6.4   Multiple Attribute Decision Making based network selection

Multiple Attribute Decision Making (MADM) [Yoon, et al. 1995] is a preference decision making mechanism that deals with problems consisting of multiple and conflicting attributes. Due to its ability to be applied to a diverse set of problems MADM has been adopted as a decision-making mechanism in all industries. Problems that can be solved by MADM display the following common characteristics:

- **Alternatives**: The options consist of a finite number of alternatives.

- **Multiple attributes**: Each problem consists of multiple attributes or criteria.

- **Incommensurable units**: Each attribute in a problem has a different unit of measurement due to which attributes within a problem cannot be compared with one another.

- **Attribute weights**: Each attribute in the problem is assigned a relative weight that gives its relative importance with respect to other attributes.

- **Decision matrix**: A problem can be represented by a decision matrix where the columns indicate the problem attributes and rows indicate alternatives.

MADM techniques can be classified into a large number of categories, however the category that has gained popularity in the area of 4G heterogeneous wireless network selection is the scoring method were alternatives are assigned scores which facilitate their comparison. The alternative with the highest score is then selected. Even among the scoring method category the classes in score based MADM that are applied to network selection are:

- Simple Additive Weighting (SAW)

- Weighted Product (WP)

- Analytic Hierarchy Process (AHP)

- Technique for Order Preference by Similarity to Idea Solution (TOPSIS)

6.4.1   Simple Additive Weighting

In this method the final score consists of an addition of the contributions of each attribute. Due to the presence of attributes with different units, all attribute values are first normalised to facilitate compatibility. Wang's study explained in section 6.2 is an example of SAW. Due to its simplistic nature, SAW is not employed extensively in network selection in wireless heterogeneous systems.

6.4.2 Weighted Product

In the WP method attributes are multiplied by their weights which act as exponents for the attribute value with which they are associated. This does away with the need for normalisation. Chen's study discussed in section 6.2 is an example of this scoring method. As it suffers from the same simplicity as SAW it does not form a popular mechanism for network selection in 4G.

6.4.3 Analytic Hierarchy Process

The Analytic Hierarchy Process (AHP) was originally developed by Saaty [Saaty. 1980]. Since then the technique has found numerous applications. In this method the MADM problem is structured in the form of a hierarchy where the main goal of the problem forms the top of the pyramid, the multiple criteria defining alternatives the middle, and the competing alternatives the bottom. This can be explained with the help of the example outlining the hierarchy for career choice satisfaction [Yoon, et al. 1995] shown in figure 6.1. Each criterion on level 2 is assigned a score between 1 and 9 based on its importance to a person. A pairwise comparison matrix is then created for the set of criteria in the following manner:

$$
\begin{array}{c|cccc}
 & \mathbf{M} & \mathbf{S} & \mathbf{G} & \mathbf{W} \\
\mathbf{M} & 1 & M/S & M/G & M/W \\
\mathbf{S} & S/M & 1 & S/G & S/W \\
\mathbf{G} & G/M & G/S & 1 & G/W \\
\mathbf{W} & W/M & W/S & W/G & 1 \\
\end{array}
$$
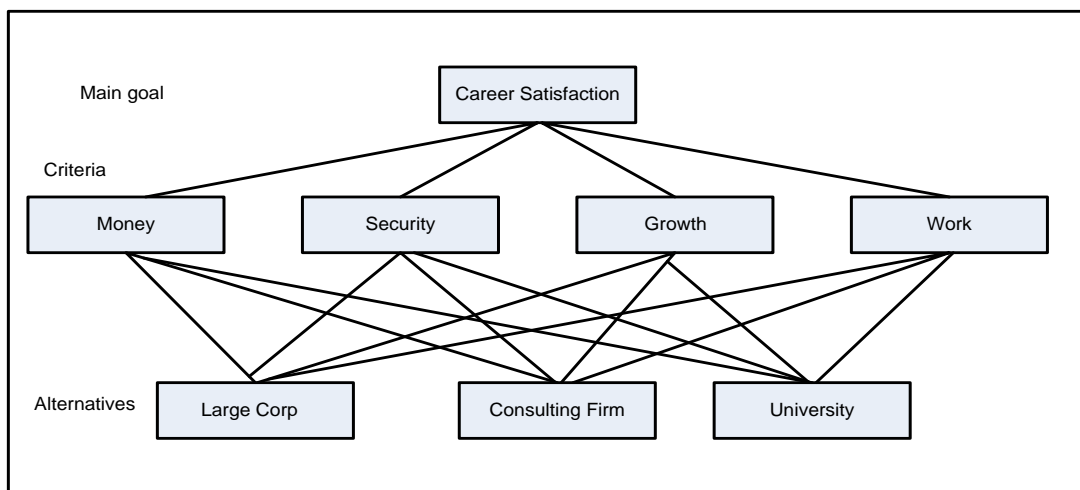
Figure 6.1 Career choice decision problem [Yoon, et al. 1995]

The next step is the creation of the normalised matrix where each element in the comparison matrix is divided by the individual sum of the column. Next the average value of each row in the normalised matrix is taken and this gives the priority for each criteria. A similar procedure is repeated for the alternatives. The overall ranking is then determined by taking the sum of products of each attribute and alternative. The combination with the highest score forms the first choice.

In the area of network selection in wireless networks, P.M.L. Chan and colleagues [Chan, et al. 2002] were the first to employ the MADM approach for network selection in a heterogeneous wireless environment. They employed a combination of analytic hierarchy process (AHP) and fuzzy logic to select the best target network for a vertical handover from among an available set of three networks – GPRS, UMTS and satellite system. They also demonstrated how AHP could be employed to determine the optimum values for certain criteria of an expensive network like satellite and turn it into a popular choice. Since the pioneering work of Chan and colleagues, the AHP process has been employed widely by a number of research studies, either on its own or together with other mathematical processes like fuzzy logic and grey relational analysis for network selection.

Balasubramaniam and Indulska [Balasubramaniam, et al. 2004] developed a vertical handover mechanism which made an attempt to decide the right time to invoke a vertical handover and to which network. This study defined a comprehensive context model which stored both static and dynamically changing information about the network, device and applications. Once the Vertical Handover Decision Process decided that a handover was needed the QoS Mapping Process applied the AHP approach for network selection based on user QoS requirements. The shortcoming of this study was that it was mainly network controlled. All the main decision-making components including the Vertical Handover Decision Process, QoS Mapping Process and Context Repository were located in the core network. The drawbacks of the network-controlled approach have been explained earlier in section 6.2.

In another study, Qingyang and Jamalipour [Song, et al. 2006] integrated AHP and grey relational analysis (GRA) to propose a network selection mechanism based on user preference, service application and network conditions. GRA is a ranking technique that finds a trade-off between alternatives and attributes. Similarly, T. Ahmed et al [Ahmed, et al. February 2006] developed a network selection scheme using AHP where available network options and contexts were compared

with predefined user and application objectives to choose the best network. Both schemes applied a large amount of user and network context information during network selection and were mobile-initiated and controlled. They however did not consider topological information and mobile movements which when introduced in the algorithm can assist further in refining the choice of a network through avoidance of unnecessary vertical handovers. Qingyang's study acknowledged the importance of network availability as a precondition to other deciding QoS factors in network selection. However network availability was described coarsely simply as being available or not available. There was no attempt to provide a quantitative measure of this crucial parameter. T. Ahmed's work did not touch upon this issue.

Li and colleagues [Li, et al. 2007] not only employed AHP and GRA similar to Qingyang but also introduced location information in conjunction with received signal strength in order to predict a forced vertical handover. Location information was mainly used to determine the sector in which the MN was located. The shortcoming of this approach was that it performed network selection first and then went on to evaluate cell coverage upon sensing another imminent vertical handover. The concept of initiating a forced vertical handover proved useful in reducing disruptions to real-time applications but it could not eliminate the problem of unnecessary vertical handovers, particularly with non-real time applications.

Liu Yang [Yang. 2007] gave a detailed analysis of AHP and highlighted some theoretical problems that existed in the approach which could lead to wrong decisions. The first problem was the required independence among all attributes at the same hierarchy level and the second was rank reversal which is a common problem in MADM. In order to overcome these problems, Yang proposed the integration of analytic network process (ANP) (parent of AHP) together with RTOPSIS (derivative of TOPSIS, explained in the next sub-section). While ANP is used for weight elicitation, RTOPSIS was applied for rating candidate networks. This study did not focus on embedding handover optimisation techniques in network selection mechanisms. Moreover, a recent study [Tran, et al. 2008] demonstrated how TOPSIS could actually suffer from ranking abnormalities when alternatives were removed.

6.4.4   TOPSIS

The TOPSIS method is a multi-criteria decision technique that considers m alternatives and n attributes as a geometric system with m points in an n dimensional space. A chosen alternative has the shortest Euclidean distance to the positive ideal solution and the longest Euclidean distance to the negative ideal solution.

Bakmaz and colleagues [Bakmaz, et al. 2007] proposed the TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) multi-criteria decision method for network selection. Network parameters considered were bandwidth, QoS level, security level and data transfer cost. This approach mainly considered network parameters for choosing a network. A simplistic weight based method was employed for users to specify their choice of alternatives and application requirements were not taken into consideration.

Highlighting the problems that arose due to ranking abnormalities in TOPSIS and ranking identification in Simple Additive Weighting (SAW) and Weighting Product (WP), Tran and Boukhatem [Tran, et al. 2008] proposed the novel Distance to ideal Alternative (DiA) approach. Instead of calculating the distance to the ideal solution as in TOPSIS, this approach calculated the minimum Manhattan distance (distance between two points measured along axes at right angles) to the positive ideal attributes. This allowed distances to change uniformly when an alternative was removed out of the candidate list. Parameters considered by this approach were jitter, delay, utilisation, packet loss and cost. Simulation results showed that DiA did not suffer from ranking abnormalities like TOPSIS and was more accurate than SAW in identifying ranks. The focus of this approach was on selecting the best network interface based on available network parameters. It did not consider user specification or application requirements and did not take context information into consideration either.

6.5   Fuzzy logic based network selection

Fuzzy logic is a system which is flexible and capable of operating with imprecise data. It reduces problems faced when comparing dissimilar network access technologies by translating parameter values into common terminology such as 'good', 'bad', 'very good' and 'very bad'. In order to tackle imprecision and nonstatistical uncertainty in parameters, P.M.L. Chan [Chan, et al. 2001] proposed

feeding measured values for each parameter into a fuzzifier which transformed real-time measurements into fuzzy sets by mapping values onto a membership function. As a result an element could have membership in more than one set. In order to obtain the handover initiation factor, fuzzy sets were fed into an inference engine where a set of IF-THEN rules were applied to decide whether a handover was necessary. In order to get rid of multiple results arising due to elements belonging to multiple sets, resultant fuzzy decision sets were converted into a precise value through the application of defuzzification. This gave the handover factor. The shortcoming of this approach was that the authors did not incorporate information like user velocity and detailed network coverage while deciding whether to perform and handover and decisions relied heavily on RSS measurements. The ability to map parameter values on to membership functions makes the fuzzy logic approach itself useful for establishing a common ground for comparing dissimilar parameters for different networks. However, while fuzzy decision answers like Yes or No are straightforward, imprecise answers like Probably Yes and Probably No can have a level of uncertainty and ambiguity which can only be resolved by referring back to actual parameter values.

The authors in [Bari, et al. 2006] proposed a network-assisted solution where the network was assigned tasks like data collection and analysis which the mobile user applied while selecting a network. The authors raised a valid point when they pointed out the reluctance of network operators in providing information to roaming users about network conditions before authentication. However, they also pointed out the drawback of the approach which is the increase in delay introduced due to overhead associated with gathering all decision related information at a single point in the network before feeding the information back to the terminal. The solution is therefore limited to network selection for new connections and is unsuitable for network selection during vertical handovers.

The same authors [Bari, et al. 2007] then considered network selection at the client side and proposed a combination of MADM with Fuzzy logic for the development of a network selection mechanism designed to work with less reliable or unavailable input attribute values. The idea was that unknown values of parameters could be estimated either by relying on past history or by prediction if their relationship with known parameters was known in advance. E.g. regression analysis could be applied to information on network utilisation to derive the values of other missing QoS parameters like delay and jitter. A new concept called confidence level was introduced which gauged the reliability of

derived attribute data based on the sensitivity of the service to this unreliable attribute value, the time since last update and its degree of correlation with the known attribute from which it was derived. The study addressed important issues in network selection management and proposed some novel solutions. However it might not be adequate for direct implementation as parameter prediction was largely dependent on a combined implementation of techniques like previous data history, fuzzification and defuzzification mechanisms, regression analysis and confidence level calculation which can be computationally intensive.

Kikilis and Rouskas in their study [Kikilis, et al. 2007] proposed the introduction of a common server which collected necessary measurements for each radio access technology (RAT) and managed the allocation of resources to clients. This was done either through the allocation of a single RAT to a user or by allocating different portions of the user's requested rate to various RATs. The aim was to maximise user satisfaction through the maximisation of utilities such as Signal to Interference ratio (SIR), Received Signal Strength (RSS), packet delay or cost. The study suggested the interesting idea of maximising resources provided to the user through simultaneous allocation of bandwidth on multiple interfaces. However, allocating the responsibility of network selection to a network-based server can have several negative implications. Firstly, the server will need to be continuously updated with context information for all active interfaces in all MNs, resulting in an increased computational overhead and unnecessary consumption of precious bandwidth. Despite this, the server will still not be in an optimum position to decide correctly as it will always lag behind in its knowledge of the latest network conditions at MNs. Secondly, it is not known how keen competing service providers will be to share sensitive network and customer related information with other competitors, giving rise to a new set of data-sharing issues. These problems also exist in Bari and Leung's work and are common problems in network-driven approaches.

Kassar et al [Kassar, et al. 2008] adopted a combination of Chan's fuzzy logic based approach for handover initiation and T. Ahmed's five stage AHP mechanism for network selection. While the paper attempted to address the issue of handling imprecise data during handovers, it did not differ in its approach on handling handovers as decisions were still dependent on RSS and MN velocity. These parameters may not be precise enough by themselves to form fundamental data on which handover decisions are based, particularly in indoor environments.

6.6   Chapter Summary

This chapter provided a comprehensive literature review of existing network selection mechanisms in the area of 4G QoS provisioning. It examined existing approaches and highlighted important unresolved issues particularly due to the selection of network criteria. Although the importance of knowledge of network availability is acknowledged by studies, they either proposed crude and simplistic methods to state network availability or ignored the parameter all together. This chapter prepared the ground for the introduction of new TBVH-aware QoS management mechanisms in the next chapter.

CHAPTER 7

Stream Bundle Management Layer for Client-based QoS Management

7.1   Introduction

In 4G heterogeneous networking, multi-interfaced heterogeneous clients are expected to intelligently handle the choice of transmission on multiple but changing wireless channels exhibiting varying levels of QoS while benefiting from the best available location-based network facilities. In the critical review presented in Chapter 6, a key deficiency that emerged among existing network selection and traffic management solutions was the uncertainty in optimal resource allocation due to uncertainty associated with the prediction of network coverage.

Unnecessary vertical handovers are a key problem that remains unresolved. The next important goal of this study is to demonstrate how TBVH plays a crucial role in improving network predictability and how this new knowledge can be applied successfully to different vertical handover scenarios to enhance the performance of network selection mechanisms, leading to the avoidance of vertical handovers. This answers the next research question

How can the knowledge of network availability be utilised to

- **assist in the minimisation of unnecessary vertical handovers?**
- **optimise the performance of network selection and QoS management mechanisms in a heterogeneous client?**

Keeping these questions in focus, the study proposes a new intelligent, client-based QoS layer called the Stream Bundle Management (SBM) layer which consists of a novel set of mechanisms for network selection, resource allocation and call admission control. The uniqueness of the SBM layer lies in its refined ability to successfully detect and avoid both unnecessary upward and downward vertical handovers through the intelligent management of available resources.

This chapter covers a detailed description of the SBM layer – its theoretical design, QoS management modules and functioning. It begins with an introduction on the SBM layer's structural framework and

key features in section 7.2. Justification for choosing a benchmark study, its noteworthy strengths and weaknesses which the SBM layer aims to overcome are discussed in section 7.3. Sections 7.4 to 7.9 explain the functionality of the different SBM layer modules and finally section 7.10 concludes with a chapter summary.

7.2   SBM layer structural framework

The Stream Bundle Management layer is a proactive, downward QoS-negotiating layer that resides in the QoS plane at the peripheral side of the YCOMM framework discussed in chapter 2, section 2.6. The layer is constantly involved in active interaction with other layers in an attempt to retrieve relevant information that is required by its different modules. This is a new feature introduced by the YCOMM framework which advocates the point that in order to alleviate the problems associated with seamless provision of network services during vertical handovers it is important that all layers in a device prepare for it in a proactive manner through a regular exchange of information among themselves, not just in a top-down fashion but in a more free manner irrespective of a layer's position in the stack.

Placed above the Transport layer in a multi-interfaced client the SBM layer handles downward QoS management tasks including resource management, traffic scheduling and flow control. The key functions of the QoS layer are summarised as follows:

- Maintenance of an up-to-date knowledge of transport and network conditions and device behavioural patterns for each active network interface through interaction with lower layers.
- Correct mapping and bundling of application QoS profiles for multi-class traffic streams on to the most appropriate network QoS profiles based on application QoS demands, device mobility patterns and prevailing network conditions.
- Prevention of unnecessary upward and downward vertical handovers by studying closely the current and predicted future QoS requirements of application traffic and negotiating resources with the network.
- Employment of intelligent client-based proactive policies for multi-class call admission control for bandwidth negotiation, providing requested QoS to new and existing connections.

- Adaptation of a fine-grained, per-flow approach for the optimal management of network QoS among connections over available channels to avoid their forced termination during channel fluctuations.

- Context-aware resource management for achieving the optimum utilisation of available network resources among different streams.

The successful implementation of context-aware QoS and resource management features in the SBM layer requires the simultaneous and continuous processing of a large variety of information coming in the form of user objectives, application demands, device context and network conditions and availability. In the SBM layer a separate module is dedicated for gathering each type of information as shown in figure 7.1. Outputs from all peripheral modules are then fed into the central network selector
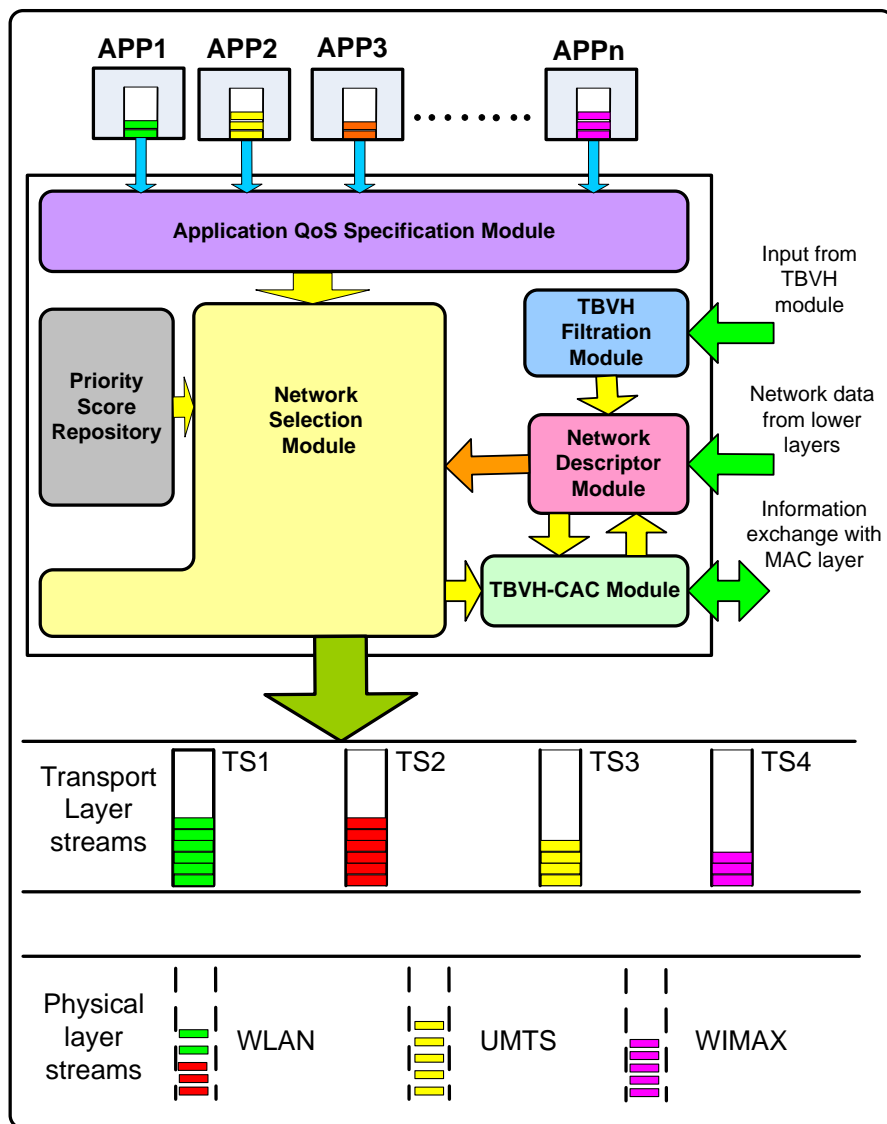
Figure 7.1 Stream Bundle Management Layer

module which runs algorithms to decide the optimal allocation of network channels and related resources to requesting traffic streams. The main functions of the six modules inside the SBM layer are summarised as follows:

- **Application QoS specification module**: this module stores QoS requirements and relative priorities of requesting application streams.
- **Priority score repository**: this contains the scores for application, interface and objective priority sets.
- **TBVH filtration module**: this module filters TBVH values from multiple exits arriving from the TBVH module developed in the first phase in chapters 4 and 5, providing a measure of the future availability of network coverage.
- **Network descriptor module**: this module contains the network descriptor matrix which stores the latest values of network parameters obtained from lower layers.
- **Call admission control module**: this module is responsible for negotiating resources with the AP based on application requirements and TBVH value.
- **Network selection module**: this module contains the AHP-based decision-making algorithms that select the most appropriate network based on the inputs received from other modules.

7.3   Benchmark for Performance Evaluation

In a noteworthy study in the past, T. Ahmed and colleagues [Ahmed, et al. February 2006] presented a strong case in the favour of handover decision mechanisms for heterogeneous networks that did not rely only on received signal strength. They acknowledged the importance of sophisticated handover decision mechanisms where a device was capable of automatically selecting the best option from among several available networks that best suited application requirements and interface capabilities. The authors then presented a new context-aware, network selection and vertical handover decision algorithm for multi-interfaced wireless clients based on the Analytic Hierarchy Process (AHP) (introduced in section 6.4.3).

The AHP based MADM method is a well known mathematical solution which successfully identifies the most suitable choice from a set of alternatives based on predefined objectives [Ahmed, et al. 2006]. The method gained huge popularity in the field of computer networks and has been applied by

a large number of research studies. In a recent study, Liu Yang [Yang. 2007] highlighted two key problems found in the AHP process. The first problem was the required independence among all attributes at the same hierarchy level and the second problem was that of rank reversal. However once these issues are taken care of, AHP is a proven mathematical solution for the successful selection of the best choice from among a group of multiple alternatives.

Among the various AHP based network selection solutions available in literature, this study picked T. Ahmed's method as a benchmark for comparison due to several reasons. The first strength of T. Ahmed's solution was the systematic and organised manner in which it handled a large variety of detailed information from the user, application and network side. This methodology has been adapted and further enhanced in the SBM layer. The second strength was that the method was relatively easy and less resource-intensive to implement as it did not require a huge number of mathematical calculations as compared to other studies that have employed algorithms like GRA along with AHP. The third strength of the study was that it acknowledged the importance of empowering the client in heterogeneous networks and proposed implementing network selection on the client side, an idea that has been promoted by this research study as well. However, the choice of T. Ahmed's method over other methods does not imply that the TBVH mechanism cannot be applied to other methods. On the contrary the positive results generated in network selection optimisation through the application of TBVH strongly advocate its ability to enhance the intelligence of other network selection and QoS management mechanisms as well. This idea forms one of the interesting future areas of work to be implemented.

Network selection in Tansir Ahmed's solution (henceforward referred to as T-NS) consisted of comparing available network contexts with predefined objectives. The objectives were also compared with one another to determine their relative importance. Finally, the best option was selected after combining the results of the two comparisons. As this research study has adopted some of the main concepts of this approach along with other proposed enhancements, the detailed working of T-NS is not explained separately as it forms part of this study's own approach and hence is automatically discussed in subsequent sections. Instead let us first familiarise ourselves with the main features of T-NS. The T-NS solution was implemented in five stages:

Stage 1: Inputs were taken from users that determined relative priorities on three sets of information. Users assigned scores from a range of 1-9. The first set consisted of relative priorities among primary objectives like interface priority, cost and QoS parameters. The second set consisted of relative priorities among network interfaces, WLAN, UMTS, WIMAX. The third set covered relative priorities among applications – streaming, interactive and conversational.

Stage 2: Some QoS parameters like bandwidth and delay tend to be very dynamic in their values. Therefore the aim of this stage was to translate the QoS preferences set by users at stage 1 into continuous values or limits in order to provide more flexibility when comparing them with network QoS parameters.

Stage 3: This stage consisted of comparing the capabilities of active networks with preconfigured user preferences set in stage 2 and assigning a suitable score to each of them.

Stage 4: The AHP method was applied in this stage to calculate the ranking among available networks based on objective priority scores and network scores assigned in stage 1 and 3.

Stage 5: Entire sessions were transferred to the most suitable network selected in stage 4.

7.3.1   Weaknesses of T-NS method

Along with its strengths, T-NS also displayed key deficiencies in the manner in which it handled network selection and resource allocation. They are listed as follows:

7.3.1.1 User-based allocation of scores to primary objectives

In T-NS, inputs on relative priorities for the application, interface and primary objective sets were taken directly from users. This can be both problematic and time consuming if implemented by the layperson. For instance a non-technical user might have no clue about the best combination of application, objective and interface priority scores. Even if a user did possess the technical knowledge to choose the right objective scores it would still be a hassle to specify the choice each time the user had a specific need.

7.3.1.2 Single relative score allocation for primary objectives for all profiles

The second deficiency was displayed in the manner in which relative scores were allocated to primary objectives like cost, interface priority and QoS parameters. As it was the user assigning it, the same objective priority scores were applied commonly for all applications. This is not an effective method in reality. For instance, in the case of the VOIP application were service providers offer ample free minutes in cellular networks these days, cost may not be as crucial as maintaining a high level of QoS as users almost always prefer high voice quality. The relative order of primary objectives in this case would be [QoS, Interface priority, Cost]. However in the case of an FTP application the aim would be to seek the most cost-effective network offering the largest number of resources. Hence the order of objectives would be [Cost, QoS, Interface priority]. This order would again differ largely with user requirements as a student would be willing to compromise QoS to a certain extent for a cheaper option whereas for a busy businessman the priority could be superior QoS but with more time constraints. In this case the order would change to [QoS, Cost, Interface priority]. These different combinations of primary objective scores were not available in T-NS.

7.3.1.3 Lack of network coverage prediction mechanisms

The third and probably the most significant deficiency was the absence of a mechanism that informed the MN about the time available to spend in a particular network's coverage. There was no way in which the network selection mechanism could find out how long a MN had access to a particular coverage before the vertical handover. A network scoring the highest ranking was straightaway assigned a traffic stream despite the fact that it could be available for a very short duration. This problem was a key cause of unnecessary vertical handovers.

7.3.1.4 Lack of resource negotiation mechanism for avoiding vertical handovers

The fourth deficiency was that T-NS did not attempt to actively negotiate network resources like bandwidth with the AP in an attempt to ward off unnecessary vertical handovers but simply allocated available resources to requesting application streams. The fifth deficiency was related to the fourth one in the sense that QoS parameter scores assigned in stage 2 were static and did not change as there was no active negotiation of resources. This was mainly because there was no way of finding out the amount of resources needed to avoid a vertical handover in the first place.

All these deficiencies are addressed by the SBM layer which consists of a set of specialised modules dedicated to resolving each issue. The functionality of each stage is significantly enhanced to eliminate the problems mentioned. Let us now take a closer look at how each one of these modules fulfils its functional objective.

7.4   Priority Score Repository

In T-NS all relative scores for application, objective and interface priorities had to be specified manually by the user which could easily result in wrong combinations due to the user's ignorance of technical details. It has been demonstrated earlier how different users can have different objective priorities for the same application type, a refinement which cannot be captured by the T-NS approach. To solve these issues the SBM layer contains the user profile repository which along with useful priority information, consists of a set of pre-defined profiles tailored to a high level of specification to meet the specific needs of different types of users. It presents a more user-friendly approach for allocating priorities over the one implemented at stage 1 in T-NS and offers two advantages.

1. The first advantage is that instead of placing the burden of specifying relative application, objective and interface priorities on the user who may be unaware of the best combination, the device allows the user to simply choose from a set of pre-configured profiles. Each profile consists of a unique combination of relative application, objective and interface priorities best suited to meet the requirements of each user that can be classified as:

- Student user profile

- Home user profile

- Business/Professional user profile

This can be implemented in two ways. The profile can either be activated on the MN's   Subscriber Identity Module (SIM) card by a service provider, something similar to Orange network's dolphin,

Table 7.1 User profile objective goals

User Profile	Defining Objective
Student	Low cost, acceptable QoS
Home	Low cost, good QoS (value for money)
Business/Professional	Excellent QoS, reasonable cost

racoon and panther price plans [Orange.co.uk. website], or all profiles can exist collectively on the device and the user given the flexibility to choose from among these profiles as per need. The second method provides the user the flexibility to easily switch between two profiles if needed.

2. The second advantage is that as the priority combinations do not change frequently, the AHP method is only applied once after the configuration stage and results of the outcome are simply stored and retrieved whenever needed. This eliminates the need to continuously perform a large number of calculations, bringing down the computational overhead associated with repetitive calculations every time the user specifies new priority scores.

This research study considers three popular profiles, student, home and business. The main defining objective goals for each profile are specified in table 7.1.

7.4.1 Assigning scores to objective sets

Keeping these objectives in mind, the next step is to assign a set of scores in between 1-9 to objectives arranged in the descending priority order. Priority scores are equally spaced integers with a space gap defined by the equation 7.1 [Ahmed, et al. 2006]

$$G = \frac{L_u - L_l}{N_p}$$ Equation 7.1

Where

N_p – number of parameters

L_u – highest score

L_l - lowest score

Table 7.2 Interface priority scores

Scores	Voice	Video-Int	Video-Strm	Data
1	UMTS	WLAN	WLAN	WLAN
4	WLAN	WIMAX	WIMAX	WIMAX
7	WIMAX	UMTS	UMTS	UMTS

This equation is applied to all the three sets of relative priorities – objective priorities, interface priorities and application priorities similar to T-NS but in a more refined fashion where the score is assigned separately per set, per application type in the user profile. This new approach may have been tedious if applied directly to T-NS as it would mean performing a new set of calculations every time a new application resource request came in. However this problem does not exist in the SBM layer's network selection mechanism (henceforward referred to as SBM-NS) due to the storage of preconfigured user profiles.

7.4.2   Assigning application priority scores

SBM-NS considers four different types of applications: voice, video-interactive, video-streaming and data transfer, the order of descending priority. Therefore, applying equation 7.1 to the situation where N_p denotes the number of application types which is 4, L_u is 9 and L_l is 1, we get g = 2, therefore the scores are voice = 1, video-int = 3, video-strm = 5 and data = 7.

7.4.3   Assigning interface priority scores

Let UMTS, WiMAX and WLAN be the three candidate networks available for selection. Interface priority scores are assigned on a per application basis in a manner similar to application priority scores. With G = 2 in this case, the interface priority scores for each of the four applications are listed in table 7.2.

7.4.4   Assigning Objective Priority scores

In SBM-NS, objective priority scores are assigned separately for each application on a per profile basis. For example, the voice application has three sets of objective priorities, one for each user profile. Along with the six primary objectives defined by T-NS, a key new enhancement to SBM-NS is the introduction of a new objective TBVH. As network availability is a prerequisite for network selection which must be fulfilled before any other condition, TBVH is assigned the highest priority. Therefore the seven primary objectives for SBM-NS are:

1. Time Before Vertical Handover (TBVH)
2. Cost
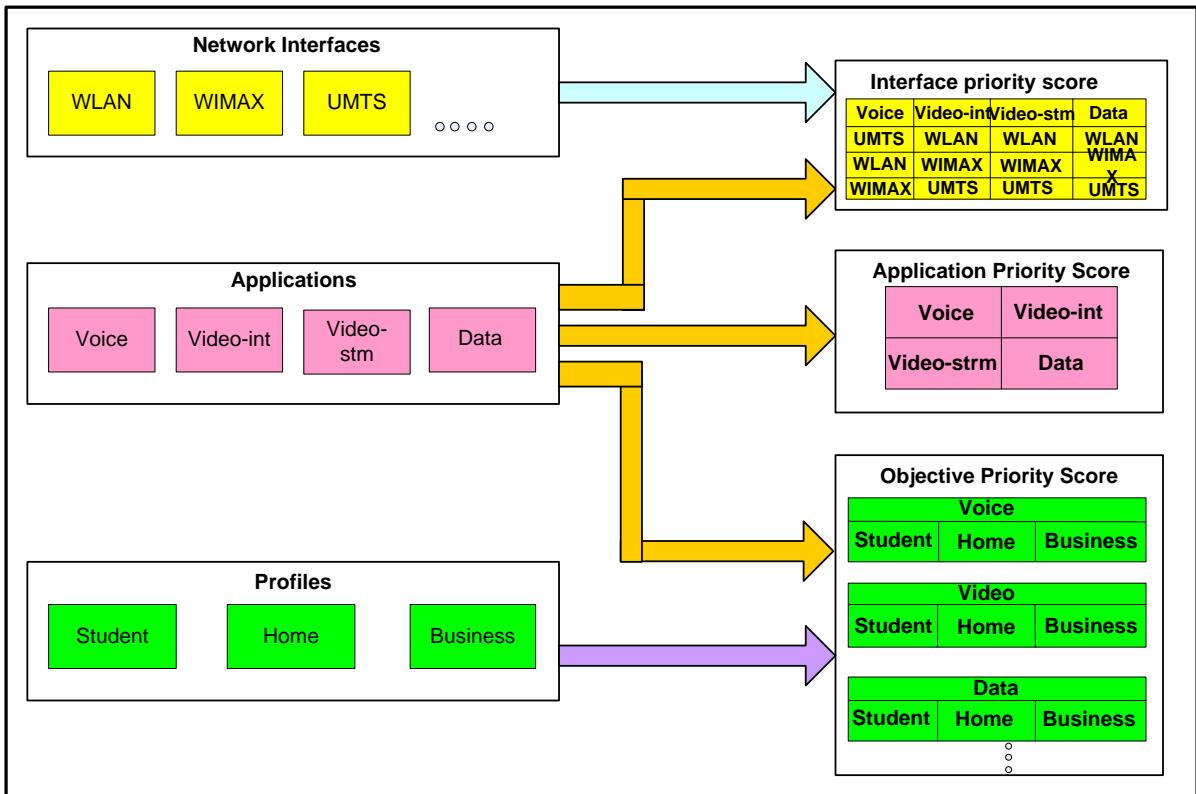
Figure 7.2 Priority score repository

3. Interface Priority (INTP)

4. Mean throughput

5. Delay

6. Jitter

7. Bit Error Rate

In this list, objectives defining QoS can be grouped together under a single objective called QoS. In this case the objective score assigned to QoS automatically becomes the score of all four objectives in the sub-category. Let us take the example of the voice application and consider the three cases in which scores can be assigned to objectives for all three user profiles. In the student profile, the order of primary objectives based on the defining goals listed in table 7.1 is [TBVH, cost, interface priority, QoS] as the user here emphasises on cost effectiveness. For a home user who is almost always connected to the home network (WLAN), the emphasis is on communicating through the home network with an acceptable level of QoS. This results in the objective priority order [TBVH, INTP, Cost QoS]. For the business user who places an emphasis on high QoS, it is the first objective to be considered after TBVH and the objective priority order changes to [TBVH, QoS, Cost, INTP].

To summarise, the priority score repository as shown in figure 7.2 consists of the application priority scores, interface priority scores and objective priority scores defined for each profile. Whenever a user specifies a new profile choice, the scores for all three sets are automatically made available, providing the user with an increased level of flexibility in the spontaneous change of profiles if needed.

7.5   Application QoS Specification Module

The main function of the application QoS specification module is the definition of downward QoS limits for each application type. This task is similar to stage 2 in the T-NS approach where dynamic values of QoS preferences are expressed as upper and lower limits in order to provide flexibility when they are compared with the QoS parameters of available networks [Ahmed, et al. 2006]. Based on the type of QoS parameters it is important to decide for each parameter whether it is best to have the highest or lowest possible value. For example, it is best to have a value as high as possible for
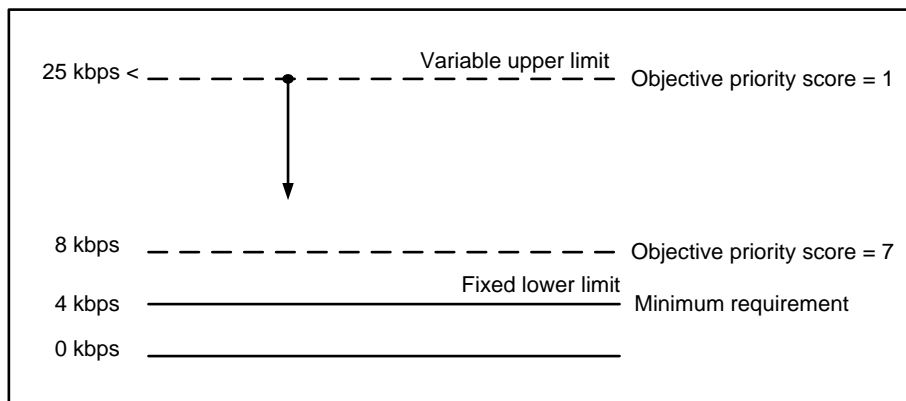
Figure 7.3 Mapping limit values for throughput [Ahmed, et al. 2006]
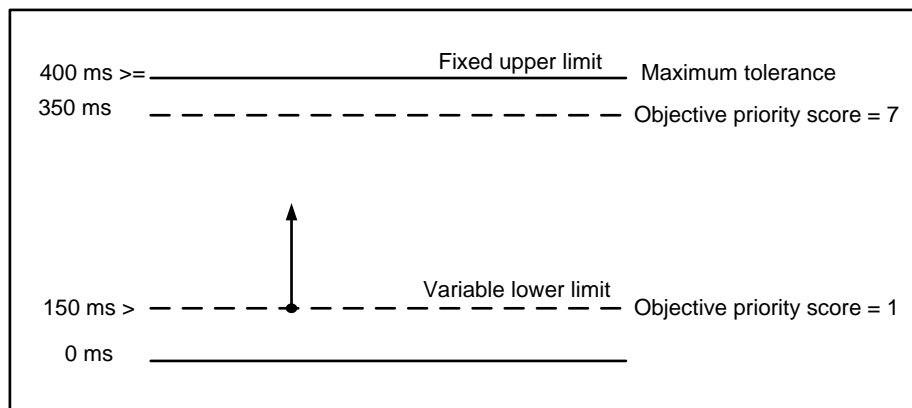
Figure 7.4 Mapping limit values for delay [Ahmed, et al. 2006]

157

bandwidth but as low as possible for delay.

In the case of a parameter like bandwidth which requires a high value, the lower limit is kept fixed. This indicates the minimum value required for the application's data transfer session. T. Ahmed explained this with the help of an example on conversational voice traffic shown in figure 7.3 where the minimum requirement for throughput was >= 4 kpbs. The upper limit varied according to the score of the QoS parameter primary objective assigned at an earlier stage. Hence the objective priority score 1 for bandwidth means a minimum value of 25 kbps. Similarly, figure 7.4 shows the delay value for priorities between 1-7 for one way delay where an increasing priority indicates a decreasing delay. In here an objective priority score 1 means a maximum delay of 150 ms. Values falling outside the window in both cases are by default assigned the highest or lowest priority score.

Therefore for each application type, based on its QoS requirements, the application QoS specification module defines the upper and lower limits for the QoS window for each static parameter which indicates an acceptable level of QoS for that parameter. All QoS parameters except TBVH are categorised as static parameters as their limits are set by the application QoS specification module upon the entry of their request. As for TBVH, its limits are set dynamically. This concept is explained in detail in subsequent sections in this chapter.

7.6   TBVH Filtration Module

The TBVH filtration module is responsible for performing the important task of improving the accuracy of TBVH results supplied by the TBVH module. The module processes available TBVH related data in a manner which makes it easier for the SBM layer to apply the intelligence in a simplified form without having to consider other underlying issues. Improving the accuracy of TBVH values involves the development of strategies to resolve the following three detrimental effects:

• Away-movement effect

• Multiplicity effect

• Ping-pong effect

Section 4.4.2.2 explained how the fidelity of TBVH in the indoor environment is dependent on the cosine parameter which provides a clear indication of whether a MN is truly about to exit the indoor
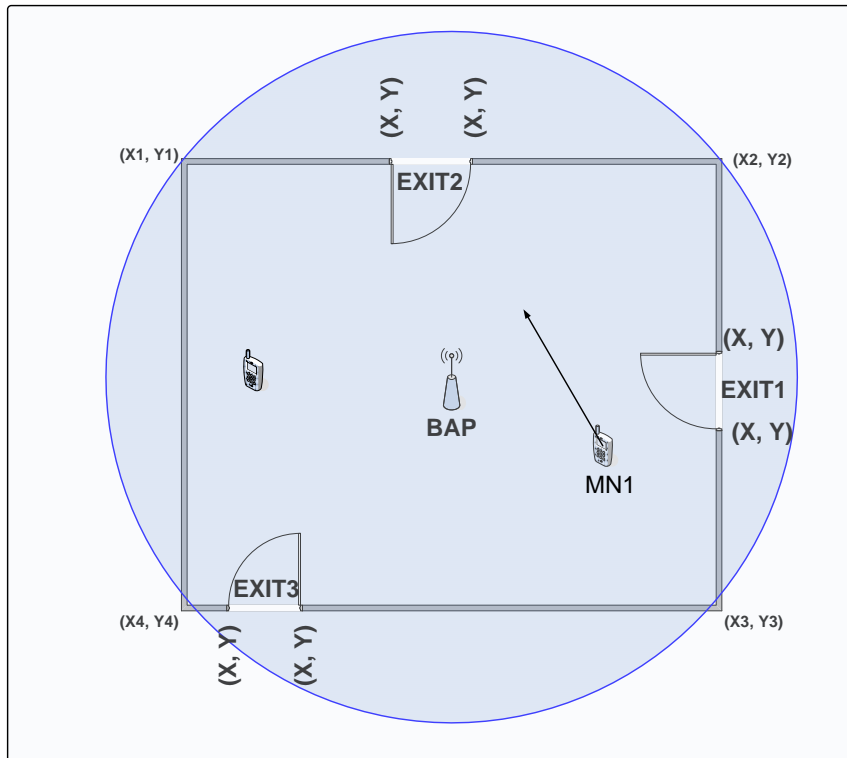
Figure 7.5 Cosine as a measure of fidelity for TBVH

environment, i.e. is moving towards the exit. The away-movement factor examines the urgency associated with a particular TBVH value and is determined with the help of the cosine parameter. Results in section 5.8.2 demonstrated how a cosine parameter value greater than 0.7 indicated a high probability that the MN was going to pass through the exit. However the importance of cosine values is not as important to a MN located in the middle of the room as it is to a MN located closer to the exit. This is because even if a MN may be directly approaching the exit from the opposite side of the room, it still has a considerable amount of time to perform a range of data transfer activities as a vertical handover is not immediately expected. The significance of cosine however increases as the MN draws closer to the exit as it exhibits an increasing urgency for vertical handover with a decrease in TBVH value. A low TBVH value accompanied by a cosine value below 0.6 indicates that although the MN is close to the exit it is not moving towards it and hence is not likely to perform a vertical handover. In this case although TBVH is calculated in order to accommodate sudden changes in direction, there is no need to consider it. The cosine value is hence a measure of the fidelity of a TBVH trigger value in enclosed environments. The case of MN1 in figure 7.5 demonstrates this. As instantaneous TBVH readings from the WLAN interface are rated by a score in between 1-9, a reading with an accompanying low cosine value (less than 0.6) is assigned the highest priority value 1

which indicates that this interface is not likely to experience a vertical handover any time soon, making it the most optimal interface in the TBVH objective category.

The multiplicity effect is considered when a MN has multiple values of TBVH, one for each exit. As multiple TBVH values cannot all be accommodated in decision making by the network selection module, it is important to identify and eliminate the unwanted values and focus on the most accurate one. The policy is to first eliminate the TBVH values accompanied by a negative cosine values. The algorithm then further shortlists all TBVH values with cosine > 0.7, choosing the smallest TBVH value from among the shortlisted ones as the final choice. In MN1's case in figure 7.5 the TBVH module creates two instances of TBVH where each instance is measured with respect to a particular exit. However as the MN is moving towards EXIT2 and away from EXIT1 it means a negative cosine for EXIT1 and high value greater than 0.7 for EXIT2. Therefore TBVH-EXIT1 is eliminated and TBVH-EXIT2 forms the final result.

The third effect that requires elimination is the ping-pong effect of TBVH readings caused due to sudden, rapid and frequent changes in a MN's direction. This effect is present in both outdoor and enclosed scenarios and is particularly of concern when a MN approaches the coverage boundary. In the outdoor scenario due to the ping-pong effect consecutive TBVH values can exhibit large fluctuations consisting of large and small readings as the MN movement fluctuates away and towards the handover threshold. This is because in open scenarios TBVH is measured with respect to the point of intersection between MN direction and HT circle. In order to ensure that a MN is not suffering from the ping-pong effect, if the current TBVH value differs by 30% more or less than the previous value, a set of three consecutive TBVH readings are observed to ensure that difference is not temporary. In this case the two readings after the current TBVH readings should exhibit uniformity in increase or decrease with respect to the previous reading. In the case of the MN that approaches the boundary and crosses the HT, three TBVH readings are observed once again to ensure it will not undergo a change in direction. If all three readings indicate decrease in TBVH, then the MN is said to be moving out and new requests for resources are submitted to other neighbouring networks.

In the enclosed scenario the ping-pong effect results in minor fluctuations of TBVH values as it is measured with respect to the exit. However the cosine value displays large fluctuations between positive and negative values. Therefore the conditions that indicate that the MN is moving steadily out
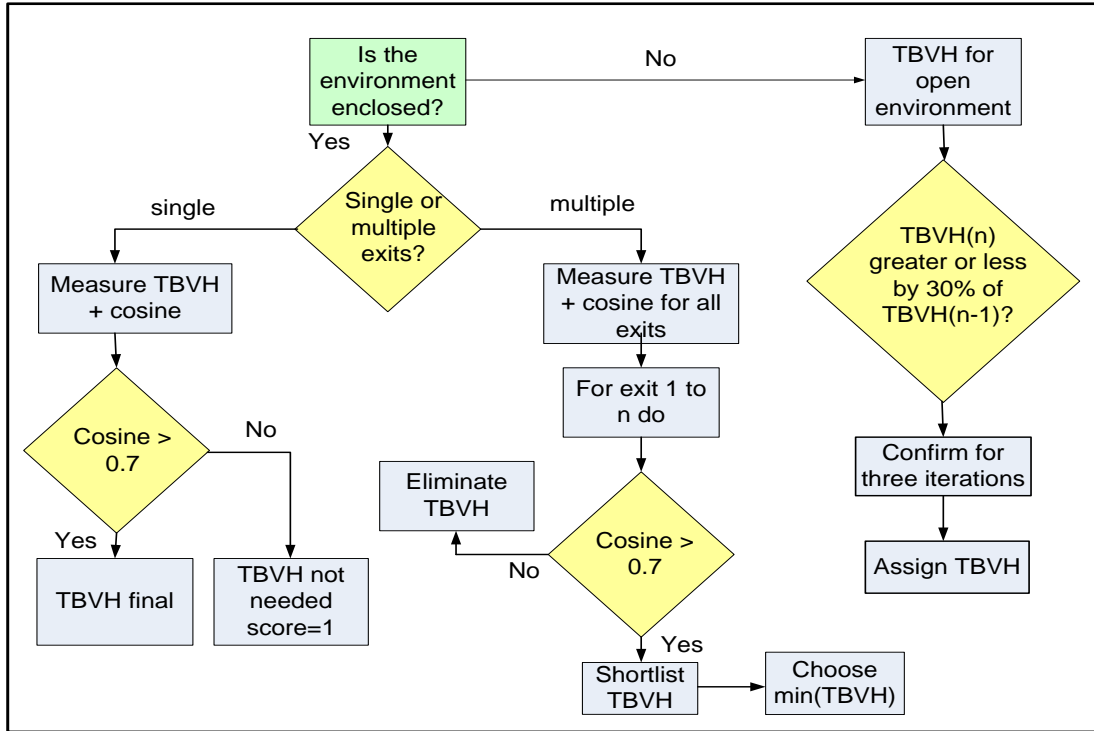
Figure 7.6 TBVH filtration flowchart

are a uniformly decreasing TBVH and a cosine value which does not fall below 0.7 for three consecutive readings. Figure 7.6 displays the flowchart for the different cases in TBVH filtration.

Once the final value of TBVH is obtained it is passed on to the network descriptor module which stores the information in the descriptor matrix along with other network parameters. These are then passed on to the network selector module when required.

7.7   Network Descriptor Module

The network descriptor module is the part of SBM layer that receives and stores regular updates from different layers about the context and available QoS conditions for each active network interface. Figure 7.7 demonstrates the different layers that supply the network descriptor module with regular context information. It mainly consists of a two-dimensional matrix called the network descriptor matrix where each column corresponds to a network parameter and each row to a network interface. Network related information coming in from lower layers is stored in the designated row for each network. A network descriptor matrix representing networks 1-n is given as

$$NWid_1 \quad Status_1 \quad BW_1 \quad Delay_1 \quad Jitter_1 \quad BER_1 \quad TBVH_1 \quad RSS_1 \quad Cost_1$$

$$\vdots$$

$$NWid_n \quad Status_n \quad BW_n \quad Delay_n \quad Jitter_n \quad BER_n \quad TBVH_n \quad RSS_n \quad Cost_n$$
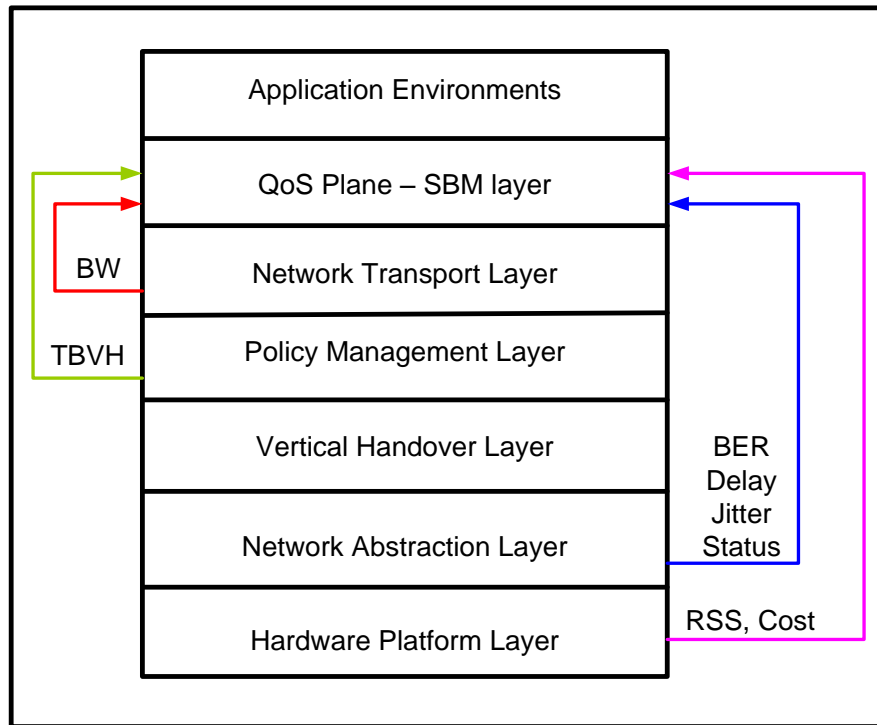
Figure 7.7 Parameter gathering from different layers

In the network descriptor matrix, *NWid* is the network identifier. Each network is assigned a unique identification number. *Status* gives the active status of the network interface, 0 being off and 1 being on. The *BW* field stores the value of the current throughput offered by a certain interface. *Delay, Jitter* and *BER* indicate the one way delay, packet delay variation and bit error rate of the wireless link respectively. TBVH gives the time before vertical handover and RSS the received signal strength which may be needed for supporting calculations. The cost field stores the information on the cost of transmission on each network interface.

7.8   TBVH-aware Call Admission Control in WLAN

In order to facilitate the understanding of TBVH-CAC, this sub-section first provides some background information on the main features of the call admission control (CAC) mechanism in IEEE 802.11e WLAN. It covers the packet exchange that takes place between the AP and MN during the negotiation

of resources, especially bandwidth. It then introduces the SBM layer's TBVH-aware CAC mechanism that submits intelligent resource requests to the AP based on the amount of time the MN is expected to reside in the WLAN coverage. The main aim of the CAC mechanism is to enable the completion of a data transfer for a particular data connection while it is connected to WLAN through the increased allocation of network resources, and hence avoiding the extra computational overhead involved with a vertical handover.

7.8.1   IEEE 802.11e Admission Control

In IEEE 802.11e WLAN standard polling is not limited to the contention-free period but can take place at any other time as well. Hence admission control in this standard [Bing. 2008] takes place both during contention-based and contention-free channel access periods. Each traffic stream has its own polling schedule. A MN requesting resources for a new traffic stream submits an Add Traffic Stream Request (ADDTS) frame, describing the various aspects of transmission and delivery in its Traffic Specification (TSPEC) element. The elements of the ADDTS Request frame are given in table 7.3 [EEFOCUS website]. The TSPEC element contains information on the characteristics and QoS information of a traffic flow. This includes the size of data frames, the average data generation bit rate, the maximum queuing and transmission delay, the maximum service interval and the minimum physical bit rate. The AP in return keeps track of the traffic on the channel and either accepts or rejects the ADDTS request. This communication mechanism differs for contention-based and contention-free channel access.

In contention-based admission control the AP first advertises the access categories in its beacons. A MN that needs to transmit in one of these access categories submits a request using the ADDTS

Table 7.3 The ADDTS Request frame [EEFOCUS website]

Order	Information
1	Category
2	Action
3	Dialog token
4	TSPEC
5-n	TCLAS (optional)
n+1	TCLAS Processing (optional)

frame. If the AP is able fulfil this request it supplies the channel time allocated to the MN for uplink transmission in the ADDTS response frame together with the QoS parameters. If the AP declines the request then the MN can still transmit information for that traffic stream, but using contention-based access and QoS parameters belonging to a lower access category. In case the MN requires additional channel time due to an increase in traffic requirements, it submits a new ADDTS request for an updated allocation.

In contention-free channel access, if the AP is able to fulfil the submitted ADDTS request it responds with a service period schedule which contains a schedule of the delivery of data and polls. If the AP cannot meet the request requirements it will reject the ADDTS request. If the requesting stream is of a higher priority then the AP may terminate the request of a lower priority stream in order to accommodate the higher order request.

7.8.2 The TBVH-aware CAC mechanism

One of the main advantages of TBVH is that it allows the network selection mechanism to adopt a more intelligent approach in reducing both unnecessary upward and downward vertical handovers. Previous studies like Kassar, Qingyang and Yan's studies [Kassar, et al. 2008, Song, et al. 2006, Xiaohuan Yan, et al. 2008] proposed a simplistic method to avoid unnecessary downward vertical handovers simply by considering the speed of a moving MN. None of the studies reviewed in literature including the ones mentioned here have considered a vital factor which is the requesting application stream's resource requirements. Speed alone cannot provide a definitive answer to whether a vertical handover would be optimum or not. This can be explained with the help of a couple of examples.

The concept of exit threshold (ET) was explained in section 4.3.1.2. It is a circle smaller than the handover threshold (HT) in the WLAN coverage beyond which the MN rarely experiences a drastic change in direction. Hence for a MN which crosses the ET it is safe to predict that it is on its way out of the network's coverage. Consider the scenario given in figure 7.8. An ET of 50m means that MN1 moving with a speed of 0.8 m/s has 60 seconds before it exits the HT. Let the TBVH mechanism also predict that MN2 travelling at a speed of 0.8 m/s has 3 minutes available in the new WLAN coverage.
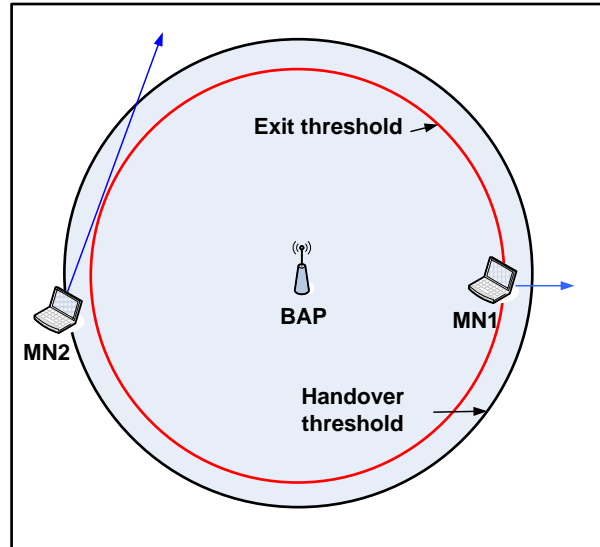
Figure 7.8 MN's movement from ET towards HT

Going back to the case of unnecessary downward vertical handover, 3 minutes is not considered a sufficient time to start a video conferencing session but is enough to start an FTP session transfer of a 10Mb MP3 file at an advertised data rate of 125 kbps. This is a clear indication that an application's resource requirements can form a decisive factor in the handover decision process.

Considering the case of upward vertical handovers, let us assume that MN1 has an ongoing FTP session with an outstanding 15 Mb remaining to be transferred. TBVH in this case is 1 minute for a speed of 0.8 m/s. If the transmission continues with the current allocated bandwidth of 100 kbps it would take another 1 minute and 20 seconds for the transfer to complete which means the MN will have to perform an upward vertical handover for the remaining 20 seconds. On the other hand, if for the remaining 1 minute the bandwidth allocated to the FTP session is increased from 100 kbps to 300 kbps, the MN is most likely to complete the transfer in 50 seconds and hence avoid an unnecessary upward vertical handover all together.

This case once again reinforces the fact stated earlier that a vertical handover is deemed necessary or unnecessary based on the resource requirements of requesting traffic streams in the MN. This establishes the importance of TBVH in the avoidance of unnecessary vertical handovers. Chapter 4 covered in detail how it is possible to supply the MN information on coverage and network resources through the AP beacons as early as the system discovery phase. This means that a new level of
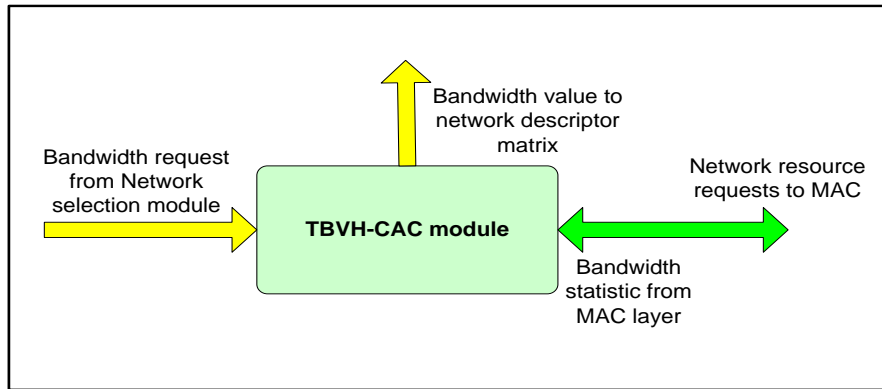
Figure 7.9 Information flow in the TBVH-CAC module

intelligence can be activated in the MN right from the beginning of the connection establishment phase with an AP.

One of the key aims of the SBM layer is to harness the benefits of this new feature and base the intelligence of resource allocation and network selection mechanisms on a combination of new information including TBVH, available network resources, user profiles and application resource requirements. Instead of allocating a network to a requesting traffic stream based on a simple match of required and available resources, the SBM layer works on negotiating an optimum level of resources with the AP/BS through the CAC mechanism with the aim of avoiding unnecessary vertical handovers. This research study is the first of its kind to propose a CAC mechanism which considers the MN, network and application contexts concurrently to propose a new solution for the refined allocation of network resources. Figure 7.9 illustrates the flow of information in and out of the TBVH-CAC module. T. Ahmed's method allocated resources to requesting applications without making an attempt to negotiate for more resources with the AP. Yan's RSS based method predicted vertical handovers but did not take application requirements into account. P.M.L. Chan [Chan, et al. 2002] developed a novel method which predicted the values of QoS parameters that could turn a worst choice network into a best choice network. It however did not consider network availability. Let us now focus on how the proposed TBVH-aware CAC (TBVH-CAC) mechanism works within the SBM layer.

The message exchange for this procedure is displayed in figure 7.10. The resource negotiation mechanism in the TBVH-CAC module is initiated upon receiving a new bandwidth request trigger from the network selection (NS) module in the form of the tuple <application type, file size, network>. The NS module submits this request when it becomes aware through calculations that the specified
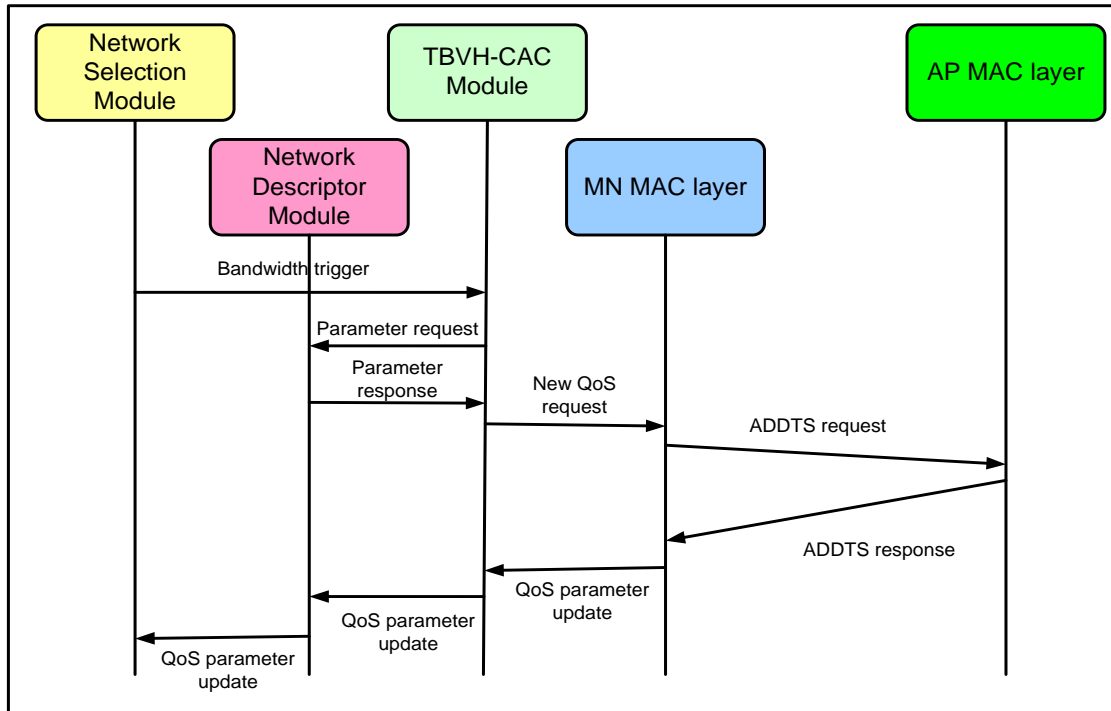
Figure 7.10 Bandwidth negotiation message exchange

application stream's transfer will not complete within the given TBVH. As this research study limits itself currently to TBVH for WLAN, it is assumed that this mechanism is activated by default for WLAN bandwidth requests. Upon receiving the bandwidth trigger, TBVH-CAC sends a request to the network descriptor (ND) module for updated information on the specified network's parameters (WLAN). The ND module which contains the latest values of network parameters responds back with the complete parameter row for that network. TBVH-CAC then sets about calculating the new bandwidth requirement. It first applies equations 7.2 and 7.3 to calculate the time required to transfer the file using the currently available bandwidth and the new bandwidth value if the transfer is to be completed before a vertical handover.

$$T_{required} \; = \frac{File \; size}{BW_{current}} \hspace{4cm} \text{Equation 7.2}$$

$$BW_{required} \; = \frac{File \; size}{TBVH} + 25 \hspace{4cm} \text{Equation 7.3}$$

Then 25 kb is added to the calculated required bandwidth to act like a buffer. This is to ensure that the file transfer will complete within the given TBVH. TBVH-CAC then dispatches the request for this new bandwidth value to the MAC layer. This incoming request at the MAC layer acts as a trigger to make it submit a new ADDTS request to the AP specifying the new bandwidth requirement. The AP's response of acceptance or rejection of the bandwidth request can lead to one of two cases.

- CASE I: The AP accepts the increased bandwidth request and responds with a positive ADDTS-response frame. The new QoS parameters are then conveyed by the MAC layer to the TBVH-CAC module which in turn informs the ND module about it.

- CASE II: The AP rejects the increased bandwidth request and responds with a negative ADDTS-response frame containing the same QoS parameters. The MAC layer conveys this information to TBVH-CAC which forwards then on to the ND module, which forwards them to the NS module. The preparation for vertical handover goes ahead.

7.9 Network Selection Module

The network selection (NS) module's functionality is an enhanced implementation of stages 3 and 4 of T. Ahmed's selection mechanism which consist of assigning scores to available networks and calculating network ranking based on the AHP method. Let us look at each stage in detail.

7.9.1 Assigning scores to available network parameters

For this stage the NS module employs a combination of information coming from the priority score (PS) repository, application QoS specification (APP-QOS) module and network descriptor module to assign scores to available networks. Current network parameter values coming in from the ND module are compared with the preconfigured application and profile preferences set by the PS repository and APP-QOS modules and scores are assigned to each network.

Assigning scores to static parameters such as interface priority and cost is simple. The values specified in the primary objective scores in the PS repository are straightforwardly assigned to the available network. Static parameters such interface priority and cost do not change over time and can be assigned scores in a simple manner. For interface priority, the values given in the PS repository are assigned to each network. As for cost, the available networks are compared with each other and

assigned scores between 1-9 based on equation 7.1 where the cheapest network has a score 1. If the cost of a network is not known it is assigned a score 9.

Dynamically changing parameters such as QoS parameters for available networks are compared with the individual parameter limit values defined by the APP-QOS module. For a given parameter, let u_i and l_i be the upper and lower limits of a particular dynamic parameter and let n_i denote the current parameter value offered by the network. The network score S_i is calculated using equations 7.4 and 7.5. Equation 7.4 is applied when the target value is expected to be as high as possible and equation 7.5 is applied when it is preferred to be as low as possible.

$$S_i = \left(1 - \frac{n_i - l_i}{u_i - l_i}\right) \times 10 \qquad ; \ l_i < n_i < u_i$$

$$S_i = 1 \qquad ; n_i \geq u_i \qquad\qquad \text{Equation 7.4}$$

$$S_i = 9 \qquad ; \ n_i \leq l_i$$

$$S_i = \left(\frac{n_i - l_i}{u_i - l_i}\right) \times 10 \qquad ; \ l_i < n_i < u_i$$

$$S_i = 1 \qquad ; \ n_i \leq l_i \qquad\qquad \text{Equation 7.5}$$

$$S_i = 9 \qquad ; n_i \geq u_i$$

7.9.2 Network ranking based on AHP

This step consists of calculating the ranking among primary objectives and then the ranking among available networks, and finally combining the two ranks to generate the final results. This procedure is performed in three steps.

Step 1: This step calculates the relative scores among the objective priority scores for each profile stored in the PS repository. This is done by applying the following equations:

$$\frac{1}{RS_{ab}} = \left(1 - \frac{S_b}{S_a}\right) \times 10 \qquad ; \ S_a > S_b$$

$$RS_{ab} = \left(1 - \frac{S_a}{S_b}\right) \times 10 \qquad ; \ S_a < S_b \qquad\qquad \text{Equation 7.6}$$

$$RS_{ab} = 1 \qquad\qquad ; S_a = S_b$$

Where RS is the comparison value of two objectives S_a and S_b. Once the relative scores among the objectives are available, the priorities in terms of selecting the best network are calculated using the pair-wise comparison matrix. The dimensions of the pair-wise comparison matrix are dependent on the number of objectives which in this case consist of Interface Priority, Cost, TBVH and QoS.

$$
\begin{array}{cccc}
1 & RS_{12} & RS_{13} & RS_{14} \\[2em]
\dfrac{1}{RS_{12}} & 1 & RS_{23} & RS_{24} \\[2em]
\dfrac{1}{RS_{13}} & \dfrac{1}{RS_{23}} & 1 & RS_{34} \\[2em]
\dfrac{1}{RS_{14}} & \dfrac{1}{RS_{24}} & \dfrac{1}{RS_{34}} & 1
\end{array}
$$

Once the matrix is populated, it is normalised by dividing each element by the individual sum of the column.

$$
\begin{array}{cccc}
B11 & B12 & B13 & B14 \\[1em]
B21 & B22 & B23 & B24 \\[1em]
B31 & B32 & B33 & B34 \\[1em]
B41 & B42 & B43 & B44
\end{array}
$$

The value for each objective is then calculated by taking the average values of each row from the normalised matrix. This gives the priority of each objective with respect to the overall goal.

$$p_i = \frac{B_{i1}+B_{i2}+B_{i3}+B_{i4}}{4} \qquad\qquad \text{Equation 7.7}$$

Step 2: In this step the relative scores among available networks for each objective are compared in a similar fashion using equations 7.4, 7.5 and 7.6. This means step 1 is repeated for finding the relative scores among networks for each of the objectives TBVH, Cost, Interface Priority, and QoS.

Step 3: This step consists of determining the final overall ranking of each network through the sum of the products of results obtained from steps 1 and 2 for each network objective. For i number of networks and j number of objectives, the network ranking is given as

$$R_i = \sum_1^{ij} c_{ij} \, p_j \qquad\qquad \text{Equation 7.8}$$

For instance let's take the example of WLAN. The objective priority for cost obtained in step 1 is multiplied by the final cost score obtained for WLAN in step 2. This is repeated for each objective and finally the overall score of WLAN is calculated by adding up all the product pairs. These final results allow the SBM layer to select the network with the highest ranking for a given application stream.

7.10   Chapter Summary

The aim of this chapter was the development of a novel set of client-based network selection and QoS management mechanisms that efficiently processed a diverse variety of context information to arrive at the best network choice for a particular application traffic stream, ensuring at the same time the avoidance of disruptions like unnecessary vertical handovers. It proposed the client-based Stream Bundle Management layer for downward QoS management which consisted of a set of functional elements that handled important QoS management tasks including QoS negotiation, network selection, resource allocation and call admission control. The presence of new intelligence on network coverage availability through TBVH enabled this research study to progress ahead of its other counterparts to develop enhanced QoS management mechanisms that worked together collectively and proactively to negotiate and manage resources in order to provide the requesting streams with the best network services for the required duration.

A unique feature of the SBM layer was its ability to maintain a detailed awareness of different contexts, and then apply this broad contextual knowledge efficiently to take refined decisions on network selection and QoS management in a manner which was not computationally intensive. The next chapter covers the working of the SBM layer in terms of mathematical results and performance evaluation. A detailed comparison of performance with T. Ahmed's technique is conducted together with a demonstration on the different ways in which the layer successfully detects and avoids

unnecessary vertical handovers. A comparative summary of key differences between T-NS and SBM-NS is given in table 7.4.

Table 7.4 Comparative summary of T-NS and SBM-NS

T-NS Method	SBM-NS Method
Users are required to set relative scores for application priorities, primary objective priorities and interface priorities.	All three sets of priorities are pre-defined in user profiles created specifically for each type of user.
Primary objective priority scores are common for all applications.	Each application had its own set of primary objective priority scores.
Does not possess mechanisms that inform the MN about the availability of network coverage.	Implements the TBVH concept which provides a precise measure of the amount of time a MN had available to spend in the network.
Cannot negotiate resources with the AP but simply allocates available resources to requesting streams.	Implements a unique CAC mechanism to negotiate resources for avoiding unnecessary vertical handovers before allocating a stream to an interface.
QoS scores are static for real and non-real time applications.	QoS limits and scores are assigned dynamically after negotiation with the AP.

CHAPTER 8

SBM Layer- Performance Analysis and Results

8.1   Introduction

This chapter mainly demonstrates the successful implementation of the SBM layer. It demonstrates how the SBM layer efficiently applies the solutions proposed in chapters 4 and 7 to efficiently manage the selection of network interfaces and resources in a manner which minimises disruptions and avoids unnecessary upward and downward vertical handovers. Different functional modules in the SBM layer provide it with detailed information on five key issues:

- The expectations of users in the form of user objectives.

- The QoS requirements of the application streams.

- Most suitable networks among currently available ones for allocating a particular call.

- The current and likely future conditions of these networks.

- The duration for which the networks likely to remain available (time before vertical handover) (in this case WLAN as UMTS and WiMAX are assumed to provide universal coverage).

This information is then collectively applied to determine the best network choice for a particular application traffic stream. The SBM layer's performance is compared with T-Ahmed's method and other proposed network selection approaches. The comparison reveals how the SBM layer succeeds in overcoming some crucial problems that exist in these efforts. Results clearly prove the importance of the knowledge of time before vertical handover in improving the intelligence and performance of handover decision and network selection mechanisms. Overall the findings of this chapter are a clear and strong proof that information on network availability is indispensible in making correct network choices in 4G clients and cannot be ignored. After giving a brief description on the experimental setup, the chapter goes on to explain the practical implementation and working of all SBM layer components, namely the priority score repository, application QoS configuration, call admission control and network selection mechanisms. Network selection results are calculated for a wide variety of vertical handover scenarios and in each case the performance of SBM-NS is compared with the other approaches and key achievements are highlighted.

8.2 Experimental Setup
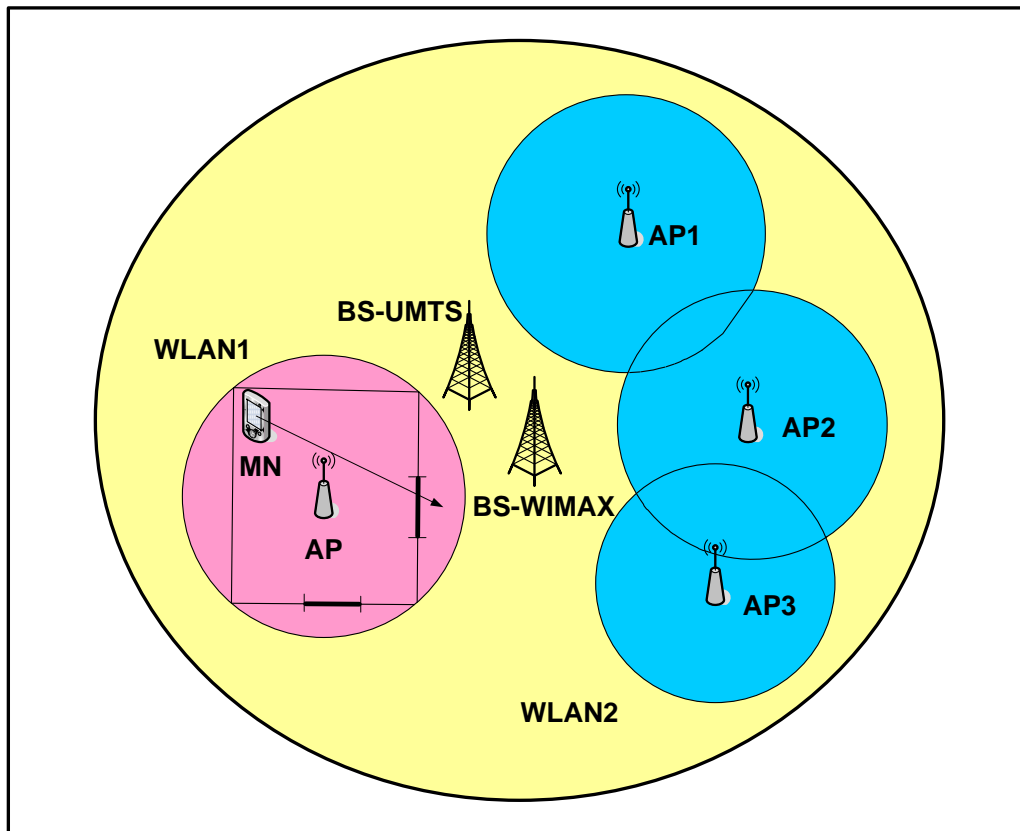
Figure 8.1 Experimental Setup of wireless networks

Figure 8.1 demonstrates the experimental setup for evaluating the performance of the SBM layer's downward QoS management solutions. It consists of a TBVH-aware MN equipped with three interfaces WLAN, UMTS and WIMAX. UMTS and WIMAX are assumed to offer overlapping and uninterrupted uniform coverage while WLAN consists of smaller pockets of intermittent high-speed coverage. The MN moves in and out of WLAN coverage while simultaneously carrying out the transfer of different types of traffic streams, thus exposing itself to frequent vertical handovers. The next few sections list the configurations and content of the SBM layer modules along with a detailed elaboration on their handling of QoS management and network selection in the wake of impending vertical handovers.

8.3 Priority Score Repository configuration

The priority score repository stores static profiles for each application – voice, data, video-interactive and video-streaming, per user profile of type student, home and business. Table 8.1 lists the priority

Table 8.1 Objective priorities scores per profile for voice

Priority Score	Student	Home	Business
1	TBVH	TBVH	TBVH
3	Cost	INTP	QoS
5	INTP	QoS	Cost
7	QoS	Cost	INTP

scores for the voice application objective priorities for each user profile. The objective priority score tables for the remaining three application types can be found in Appendix A1.2. Interface and application priorities per application were listed in chapter 7 and a copy can also be found in Appendix A1.2. For the sake of brevity, results in this chapter focus on the student profile as decision making procedures remain similar for other profiles as well.

8.4   Application QoS configuration

The priority score repository assigns specific scores to each objective based upon its importance for each user profile as shown in table 8.1. For instance for the student profile, QoS is given a priority score of 7 which means that all QoS parameters including bandwidth, delay, jitter and BER are automatically assigned score 7. In terms of actually QoS parameter values it means that maximum expected value for each parameter from the network is that which can be mapped on to a score 7. Therefore the next important step is to translate or map the scores of each of these continuous parameters into tangible QoS measures for each of the four applications in the student profile. This includes specifying upper and lower QoS limits and priority scores for different values in between the two limits as shown in figures 7.3 and 7.4 in the previous chapter. Scores assigned to the QoS objective in the PS repository are translated into actually QoS values with the help of the limit sets. At a later stage, networks exhibiting a certain level of QoS also have their present parameter values translated into scores in order to enable comparison between application QoS requirements and available network QoS.

In both interactive voice and video, the upper and lower limits are assigned to TBVH in a static manner which means these limits are available to network selection module at the start of the decision-making procedure. This is because in the case of interactive traffic there is a real time transfer of information and it is difficult to predict the exact amount of time for which the interactive

connection will remain active. However, as the bandwidth requirements for these applications are constant, once a connection is established with the required bandwidth, negotiating more bandwidth after the current value has reached a certain upper limit threshold will not cause a significant improvement in quality. In the case of streaming and data transfer applications as an increase and improvement in allocated resources like bandwidth can actually speed up the transfer (assuming TCP windows on both sides support the new higher data rate), TBVH limits for these applications are assigned in a dynamic fashion. This is explained later in more detail in section 8.5.2.3.

Tables 8.2 to 8.5, list the upper and lower defining limits for each application in the student profile. In each parameter, upper limits are highlighted in green. For both streaming video and data transfer application which can benefit from an increase in bandwidth, TBVH is not specified for these traffic types at this stage. Instead its upper and lower limits are calculated dynamically soon after the network bandwidth is finalised. QoS values falling outside the defined limits by being too high or too low are assigned a high or low score respectively. Let us take a look at the assignment of bandwidth and TBVH limits for the voice application. For bandwidth the lower limit corresponding to objective score 9 is 4 kbps. This is the minimum requirement below which the call quality will deteriorate badly. The upper limit is kept at 64 kbps for an objective score 1. As for TBVH, the lower limit corresponding to score 9 is a value less than 300 s (5 minutes) which means that if TBVH for a particular network is less than 300 s, a voice call will not be allocated to it. The highest limit is 600 s. With TBVH assigned the highest priority among primary objectives, a network scoring 1 for TBVH is likely to be a strong candidate for selection.

Table 8.2 Defining upper and lower QoS limits for Voice

Score	B/W Kbps	Delay Ms	Jitter Ms	TBVH Seconds
1	64	< 100	< 40	600
2				540
3	32	150		500
4			40	480
5	25	250		360
6				330
7	16	300	70	300
8				
9	4	≤ 400	> 75	< 300

Table 8.3 Defining upper and lower QoS limits for video-interactive

Score	B/W Kbps	Delay Ms	Jitter ms	TBVH Seconds
1	750	< 100	< 30	600
2	650			540
3	512	150		500
4	350		30	480
5	200	250		360
6	128			330
7	64	300	40	300
8				
9	< 64	≤ 400	> 50	< 300

Table 8.4 Defining upper and lower QoS limits for video-streaming

Score	B/W Kbps	Delay Ms	Jitter ms
1	700	< 100	< 30
2	512		
3	400	150	
4	256		30
5	128	250	
6	64		
7	32	300	40
8			
9	< 32	≤ 400	> 50

Table 8.5 Defining upper and lower QoS limits for data

Score	B/W Kbps	Delay ms	Jitter Ms
1	700	≤ 300	300
2	550	370	
3	490	400	400
4	400	430	
5	350	450	
6	270	500	
7	200	520	520
8	100	550	
9	5	600	600

8.5   Network Selection Mechanism

Table 8.6 T-NS Interface priority scores for available networks

Interface Priority Score	Network Interface
1	WLAN
4	UMTS
7	WIMAX

Table 8.7 T-NS Cost priority scores for available networks

Cost Priority Score	Network Interface
1	WLAN
3	WIMAX
7	UMTS

Table 8.8 T-NS Objective priority scores for available networks

Objective Priority Score	Objective
1	Cost
4	Interface Priority (INTP)
7	QoS

This section covers the two important steps in network selection:

- Assigning scores to available networks

- Calculating network ranking based on AHP

In the current evaluation scenario all values of objective, application and interface priorities are kept constant throughout the network selection procedure. Network selection takes place first using the T-NS approach which does not consider TBVH. Next the same set of parameters is applied to SBM-NS along with TBVH. In order to demonstrate the effect of static and dynamic allocation of TBVH and the functioning of the TBVH-CAC mechanism, two types of traffic are considered for evaluation - voice and data in the student profile.

8.5.1 T-NS versus SBM-NS network selection for interactive voice

This sub-section evaluates the performance of T-NS and SBM-NS mechanisms with respect to selecting the best network for an interactive voice traffic stream. In order to facilitate the comparison of application QoS scores and network QoS scores, it is necessary to assign scores to available networks as well based on the quality of QoS parameters they provide. As objectives like interface priority and cost are assigned in a straightforward fashion, tables 8.6 and 8.7 list their scores for

WLAN, UMTS and WIMAX. These scores remain the same for T-NS and SBM-NS. As for the continuous parameters, individual network QoS parameter values are converted into scores using tables 8.2 to 8.5 which automatically assign them a score based on their current values.

8.5.1.1 T-NS based network selection

It is important to remember that in T-NS there is no refined allocation of objective priority scores on a per application basis. Objectives and networks are assigned the same priority score for all applications. As T-NS does not consider TBVH it has three primary objectives with priority scores given in table 8.8. Once scores are assigned to primary objectives, the next step is to perform calculations based on equation 7.6 that establishes the rank of each objective with respect to the other as discussed in section 7.9.2. A pair-wise comparison matrix is then constructed out of the resultant values. The relative scores among objectives Cost, INTP and QoS based on the values listed in table 8.8 are calculated as follows:

RS_{12} = (1- 1/4) x 10 = 7.5

$1/RS_{12}$ = 0.133

RS_{13} = (1- 1/7) x 10 = 8.57

$1/RS_{13}$ = 0.116

RS_{23} = (1- 4/7) x 10 = 4.285

$1/RS_{23}$ = 0.233

The comparison matrix of the above values with rows and columns arranged in the order of objectives cost, INTP and QoS is

1	7.5	8.57
0.133	1	4.285
0.116	0.233	1

Normalising the matrix elements we get

0.8006	0.8588	0.6185
0.1064	0.1145	0.3092
0.0928	0.0266	0.0721

Taking the average of each row we have the final primary objective ranking scores as listed in table 8.9. As cost has the highest score this means the emphasis is on selecting the cheapest network, followed by interface priority, followed by QoS. This establishes the rank of each primary objective on the MN side.

The next step is to take scores allocated to all three networks for each primary objective and construct a pair-wise comparison matrix for each objective using equation 7.6. Which means cost and INTP both have a pair-wise comparison matrix each. The resultant scores for both objectives are given in table 8.10. The network ranking in table 8.10 demonstrates that being the cheapest of the three, WLAN scores the highest in terms of cost whereas UMTS that is chosen as the best interface for voice due to its fine QoS constraints emerges as the best candidate based on interface priority. Establishing both primary objective and network ranking for non-continuous parameters is straightforward. Let us now look at the ranking for continuous QoS parameters.

Table 8.9 Primary objective ranking scores for the client side

Objective	Ranking score
Cost	0.7593
INTP	0.1767
QoS	0.0638

Table 8.10 Relative ranking scores among networks for cost and INTP objectives

Interface	Cost Score	INTP score
WLAN	0.7	0.8
UMTS	0.1787	0.2
WIMAX	0.2	0.1

Table 8.11 Actual QoS parameter values offered by networks

Interface	Bandwidth Kbps	Delay Ms	Jitter Ms
WLAN	64	80	50
UMTS	32	120	30
WIMAX	64	170	60

Now let's consider the actual QoS parameters as they are offered by each network interface for the voice requests. The QoS objective collectively represents bandwidth, delay and jitter parameters. Real-time values are specified in table 8.11. For the sake of brevity, only bandwidth, delay and jitter are considered for comparison. Applying the QoS limits for voice defined in table 8.2, scores are assigned to each individual parameter value using equations 7.4 and 7.5 given in section 7.9.1.

Bandwidth

Upper limit (UL) = 64, Lower limit (LL) = 4,

For WLAN and WIMAX, Ni = 64 which is equal to UL, therefore bandwidth score = 1.

For UMTS, Ni = 32, therefore Si = (1 – (32-4)/(64-4)) x 10 ≈ 5, bandwidth score = 5.

Final bandwidth score: WLAN = 1, WIMAX = 1, UMTS = 5

Delay

UL = 400, LL = 100

WLAN (Ni) = 80 < LL, therefore delay score = 1.

UMTS (Ni) = 120, therefore Si = ((120-100)/(400-100)) x 10 ≈ 1. delay score = 1.

Similarly, WIMAX (Ni) = 2 hence delay score = 2.

Final delay score: WLAN = 1, UMTS = 1, WIMAX = 2.

Jitter

UL = 75, LL = 40

Final jitter score: UMTS = 1, WLAN = 3, WIMAX = 6.

Table 8.12 Final network ranking scores for continuous QoS parameters

Interface	Bandwidth score	Delay Score	Jitter Score
WLAN	0.4	0.5	0.2
WIMAX	0.4	0.1	0.1
UMTS	0.1	0.5	0.7

Applying AHP to obtain the network ranking for each of the three QoS parameters we obtain comparative scores for QoS as given in table 8.12.

Finally, based on equation 7.7 in section 7.9.2, the final ranking of each network is calculated by taking sum of the product of the ranking score for an objective from table 8.9 with the respective network ranking score for that objective as given in tables 8.10 and 8.12. For instance, in order to obtain the final ranking of WLAN in terms of cost, the relative cost objective ranking score from table 8.9 is multiplied with WLAN score for cost from table 8.10. Therefore the final rank of WLAN is calculated as

 (0.7)(0.7593) + (0.8)(0.1767) + (0.4)(0.0638) + (0.5)(0.0638) + (0.2)(0.0638)

= 0.63703.

Similarly,

WIMAX = 0.207

UMTS = 0.35998.

Note that in the case of QoS parameters the QoS primary objective ranking score from table 8.9 is multiplied by each of the network QoS parameter scores from table 8.12.

8.5.1.2 SBM-NS based network selection

Calculations in the SBM-NS approach for the voice application are similar to those in T-NS except that in this approach there are four primary objectives instead of three, including TBVH. As the procedures for assigning relative primary objective priority scores on the client side and obtaining their relative ranking scores based on the AHP and pair-wise comparison matrix have already been explained, for the sake of brevity table 8.13 directly lists the two sets of scores for each primary objective.

Table 8.13 Primary objective scores and their relative ranking scores based on AHP

Objective	Objective priority score	Relative ranking score
TBVH	1	0.7
Cost	3	0.2
INTP	5	0.1
QoS	7	0.04

As the instantaneous QoS parameter values for bandwidth, delay and jitter coming in from networks are kept the same as those in T-NS, they are assigned the same score as given in table 8.12.

The key enhancement in the SBM-NS approach over T-NS is the TBVH score which changes dynamically as the MN moves towards or away from the boundary. As the TBVH values for WIMAX and UMTS are not calculated (out of scope), they both are assigned a default score of 5 which falls in the middle of the score scale. As for the score of TBVH, the lower limit is assigned as 300 seconds (5 mins) which means that if the time remaining in WLAN is close to 300 seconds, the voice stream will not be assigned to it. For WLAN, a set of scores is calculated based on different values of TBVH in the following manner:

UL = 600, LL = 300

WLAN (Ni) = 400. Therefore, WLAN score = 6.

8.5.1.3 Performance evaluation of TBVH in avoidance of unnecessary vertical handovers

The following experiment is conducted in order to demonstrate the refined ability of SBM-NS in recognising approaching boundaries. The MN moving with a constant pedestrian speed and running a voice application is first placed close to HT such that TBVH is less than the lower limit of 300 s. This causes the network selection module to automatically assign TBVH the lowest score 9. With UMTS and WIMAX given a score of 5 each, the module runs the complete AHP process and the final sum of product of objective score ranking and network score ranking is calculated for all three networks. Network ranking in this case is as given in row 1 of table 8.14. As TBVH is the highest priority objective in SBM-NS a low score for WLAN automatically causes the network to get the lowest overall score as well. UMTS is chosen as the best choice in this case for voice traffic due to superior QoS delivery over WIMAX.

Table 8.14 Network ranking scores for dynamic TBVH values

TBVH (Seconds)	WLAN score	UMTS score	WIMAX score
300	0.334	0.462	0.424
360	0.334	0.392	0.354
390	0.1	0.4	0.4
420	0.404	0.394	0.354
490	0.614	0.322	0.0155
520	0.754	0.182	0.144
570	0.824	0.182	0.144
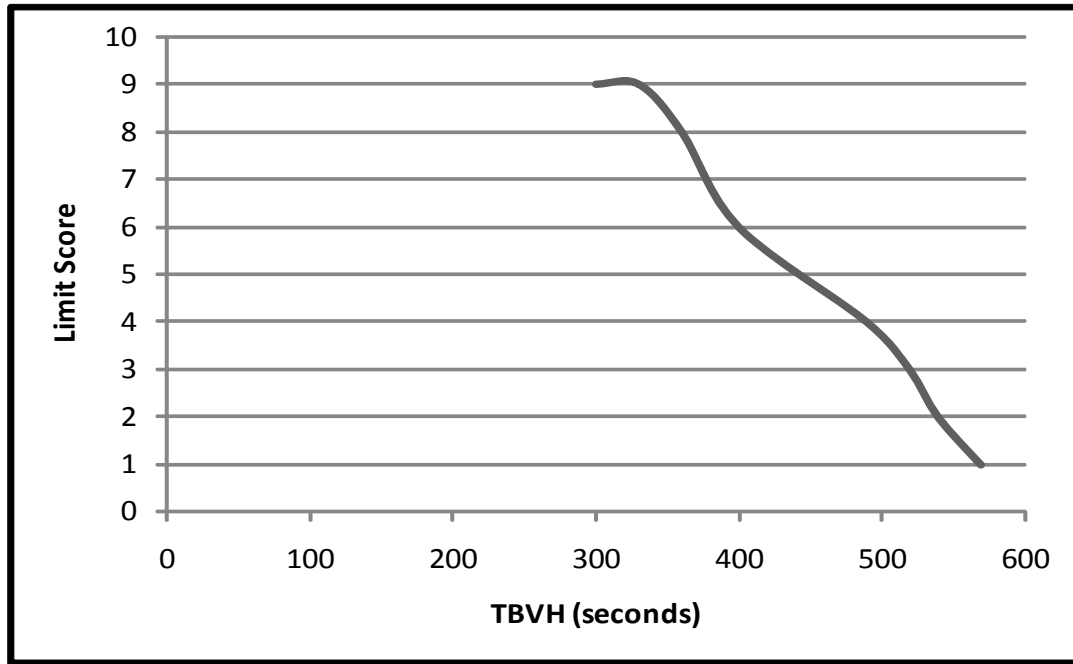
Figure 8.2 Graph of TBVH versus priority score for WLAN

The noteworthy point here is that UMTS is chosen over WLAN despite the fact that WLAN displays higher values for QoS parameters, especially bandwidth, over UMTS. These high parameter values however fail to attract the SBM-NS mechanism as it is aware that these resources are not likely to remain available for the entire call duration. This is one of the major improvements SBM-NS demonstrates over the majority of network selection solutions in literature. For instance, Qingyang's solution (Q-NS) and Kassar's solution (K-NS) are programmed to select the network based on the best set of available QoS services. Although 5 minutes was specified as the minimum threshold limit in TBVH for voice calls for testing purposes, the mechanism can be further enhanced to allow the user to specify this minimum threshold which can be a lot less than 5 minutes.

The MN is now made to move inwards into WLAN coverage. As TBVH increases, the score for TBVH objective in WLAN decreases. A graph tracing this behaviour is plotted to study the relationship between actual TBVH values and corresponding WLAN priority score for the TBVH objective as shown in figure 8.2. From the graph it can be seen that the priority score increases with the decrease in TBVH. This confirms that as TBVH is the objective with the highest rank, its priority score has a significant impact on the choice of network. Final network rankings are calculated at regular intervals. At TBVH = 420 s, WLAN replaces UMTS as the best choice network with a final score of 0.404. As TBVH increases, it continues to remain a strong choice of selection.

184

Let us now observe how K-NS and Q-NS along with T-NS perform in this same situation. A graph of final network choice against TBVH values is plotted for all four approaches including SBM-NS as shown in figure 8.3. In this case WLAN and UMTS are represented by numbers 1 and 2 respectively. The graph demonstrates how all the three earlier methods that specify only QoS parameter superiority as the defining choice go on to select WLAN as the best choice network despite the fact that it will not be available for the minimum duration required for voice applications, causing the device to draw closer to perform an upward vertical handover in the near future.

SBM-NS however recognises this and makes the intelligent choice between UMTS and WLAN based on the time the MN has available to spend in WLAN coverage. As soon as TBVH reaches 400 seconds, WLAN becomes the first choice network for assigning the voice application. This mechanism becomes especially useful when the device has a few tens of seconds to go in WLAN before it moves out of coverage. When SBM-NS selects UMTS in such a case, it not only assigns the voice stream to a more stable network, but by doing so it proactively prevents an unnecessary upward vertical handover and the additional network overhead associated with it. This proves the fact that the
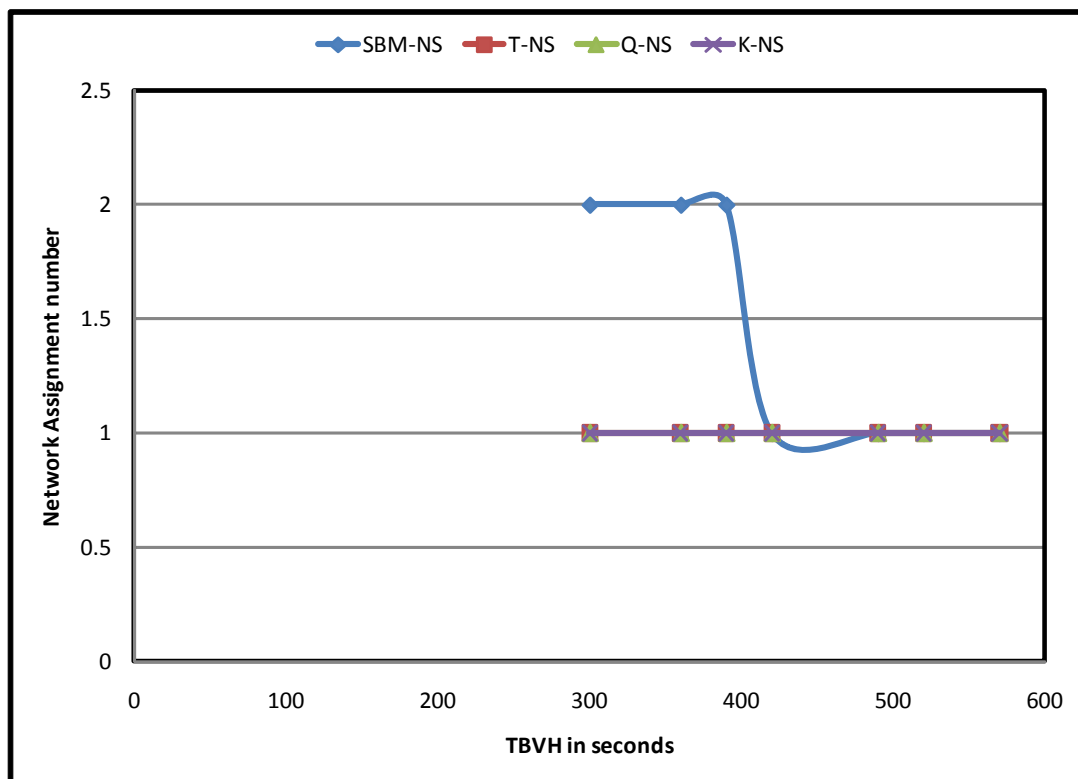
Figure 8.3 Effect of TBVH on network selection in improved performance of SBM-NS over T-NS, Q-NS and K-NS

advanced knowledge of network availability can assist in the avoidance of upward vertical handovers, resulting in a conservation of power and network and device computational resources.

Similarly in the case of downward vertical handovers, when the MN roams into temporary WLAN coverage, the SBM-NS network selection module with the help of available TBVH easily decides whether or not to assign the voice connection to free WLAN coverage. In here it is important to emphasise the fact that unlike Yan, Ylyanttila and Kassar's methods which consider a simplistic and often unreliable combination of only speed and RSS to decide the allocation of a call to WLAN, a technique which can often result in incorrect choices if the actual duration of the MN in WLAN is not considered, the SBM-NS takes into account the actual application's requirements as well along with the duration the MN will remain in WLAN coverage. This enables it to make a more refined and intelligent choice.

8.5.2 T-NS versus SBM-NS network selection for data file streaming application

This sub-section discusses results that highlight the important role TBVH plays in the avoidance of both upward and downward vertical handovers for streaming data traffic. TBVH knowledge is more beneficial in the case of streaming applications as the knowledge of coverage duration can actually play a key role in the enhanced allocation of network resources to a traffic stream. Let's consider an FTP data transfer session requesting an optimum network interface. The efficiency of T-NS and SBM-NS is evaluated in terms of the handling of the request in a manner which assists in preventing both unnecessary upward and downward vertical handovers.

Table 8.15 Upper QoS limits for data transfer traffic

Interface	Bandwidth Kbps	Delay Ms	Jitter Ms
WLAN	500	400	400
UMTS	80	420	300
WIMAX	400	380	300

Table 8.16 Interface priority and cost scores

Network Interface	Interface Priority	Cost
WLAN	1	1
WIMAX	4	4
UMTS	7	7

Table 8.17 Primary objective scores and AHP based relative ranking scores

Objective	T-NS Priority Score	T-NS Relative Score	SBM-NS Priority Score	SBM-NS Relative Score
TBVH	-	-	1	0.7
Cost	1	0.7593	3	0.2
QoS	4	0.1767	5	0.1
INTP	7	0.0638	7	0.04

The current QoS parameter values for bandwidth, delay and jitter for the three networks are given in table 8.15. Table 8.5 has specified earlier the QoS limits for streaming data transfer as defined by the application QoS specification module. The Interface priority and cost scores for each network are specified in table 8.16.

Let's consider the case where the FTP application requests resources for the transmission of a file of size 70 Mb. When network selection is performed by T-NS due to a high priority for cost the mechanism selects the cheapest network with the required QoS. The available network parameters from table 8.15 are converted into scores. Table 8.17 lists the primary objective priority scores and their relative ranking scores calculated with AHP for both T-NS and SBM-NS. This information shows how T-NS focuses mainly on the cost objective for selection and SBM-NS on the TBVH objective. Finally table 8.18 lists the ranking of each network with respect to each objective. The information in table 8.18 is sufficient for T-NS to combine it with the network ranking score for each network, for each objective.

Therefore the final network score based on T-NS is

WLAN = 0.81816

WIMAX = 0.35716, and

UMTS = 0.17643.

Table 8.18 Relative scores

Objective	WLAN	WIMAX	UMTS
Cost	0.7588	0.1767	0.06293
Interface Priority	0.7588	0.1767	0.06293
Bandwidth	0.6094	0.3166	0.07373
Delay	0.4166	0.4166	0.1666
Jitter	0.4651	0.4651	0.0697

Therefore, the order of network selection is WLAN, WIMAX and UMTS. This order remains static throughout the lifetime of the connection, irrespective of MN movements.

In SBM-NS, along with the final objective scores given in table 8.18, network selection additionally requires TBVH score. This objective has been left out purposely from the table as it is calculated dynamically for the data transfer application.

Let us take a closer look at how SBM-NS handles the FTP request. Allocation of resources to streaming applications requires the active collaboration between all modules in the SBM layer. By continuously monitoring network parameters, the SBM layer is always on the lookout for ways in which it can avoid vertical handovers. In the case of the 70 Mb FTP request instead of simply allocating a network to the requesting application, SBM-NS first runs some checks to ensure that the application will actually benefit from the network and experience minimum disruption. As the requesting application is FTP which will clearly benefit from the free WLAN coverage, SBM-NS first checks its file size, available network bandwidth on WLAN and TBVH and determines whether it is possible to complete the file transfer while the MN is within WLAN coverage. Based on the results of these checks there are five different case scenarios that can occur.

8.5.2.1 CASE I: Available bandwidth sufficient to allow completion of transfer within TBVH

Let the current value be TBVH = 180 s

File size = 70,000 kb

Bandwidth = 500 kbps

Time required to complete transfer = 70,000/500 = 140 s

As this indicates that the transfer can easily complete within the current WLAN coverage, the network score for TBVH is set to 1 (WIMAX = 5, UMTS =5). The relative TBVH ranking score for the three networks is

WLAN = 0.8

UMTS = 0.1

WIMAX = 0.1.

After introducing the TBVH score in table 8.18 and calculating the relative priority through AHP, the final ranking score for the three networks is

WLAN = 0.85163, WIMAX = 0.2322 and UMTS = 0.155.

The optimum network for selection in this case is WLAN.

8.5.2.2 CASE II: TBVH indicates WLAN will not be available for required duration of file transfer

In this case the time required is 140 s but let current TBVH be 80 s. One way of making the transfer possible within TBVH is to submit the request for additional bandwidth to the AP. The NS-module submits the new bandwidth request to the TBVH-CAC module by specifying the tuple <FTP, 70,000, WLAN>. Upon receiving the request and parameter values from ND module, the TBVH-CAC mechanism (section 7.8.2) determines through calculations that the additional bandwidth needed is (70,000/80) = 875 + 25 = 900 kbps (25 kbps for buffering purposes). It then submits the request for 900 kbps to the MAC layer which in turn sends out a new ADDTS request to the AP. This situation can again lead to two different cases based on whether the requested is granted or rejected by the AP.

8.5.2.3 CASE III: AP grants request for additional bandwidth

Table 8.19 Dynamically allocated network ranks for TBVH and Bandwidth

Objective	WLAN score	WIMAX score	UMTS score
TBVH	0.8	0.1	0.1
Bandwidth	0.747	0.1939	0.0585

In this case the network scores for TBVH and bandwidth are reallocated dynamically to accommodate new conditions. As the requests can be fulfilled within TBVH, the network score for WLAN is set to 1. The new set of relative network ranking scores for bandwidth and TBVH are as shown in table 8.19. The final network ranking for this scenario is

WLAN, WIMAX and then **UMTS**.

This is a clear proof of how through intelligent resource management the SBM-NS avoids an unnecessary upward vertical handover.

8.5.2.4 CASE IV: AP rejects request for additional bandwidth

Table 8.20 Dynamic score allocation for TBVH

TBVH	Priority Score
116.66	1
102.08	2
87.5	3
72.92	4
58.34	5
43.76	6
29.18	7
14.6	8

As the AP has not granted the request for more bandwidth, the SBM-NS mechanism sets about executing procedures that will minimise the effect of the vertical handover. The remaining file size is first divided by 8 which indicates score range. Therefore, (70,000 / 8) = 8750 kb. The time needed to transfer 8750 kb is 14.58 s with the current number of allocated resources. The total available TBVH is 116.66 s. This is divided into 8 sub durations in order to set the dynamic limits for TBVH. In this case we have the dynamic set of scores as shown in table 8.20.

Using equation 7.4 we have $S_i = (1 - (80 - 14.6)/(116.66 - 14.6)) \times 10 \approx 4$

Applying AHP to obtain relative network scores we get the relative score values

WLAN = 0.5, WIMAX = 0.2 and UMTS = 0.2.

The further application of AHP to derive network ranks gives the final score

WLAN = 0.7547, WIMAX = 0.229 and UMTS = 0.17541.

The order of network selection therefore is WLAN, WIMAX and UMTS.

This is despite the fact that it will result in a vertical handover. This is because results indicate that as most of the file transfer will still take place in WLAN, it remains a favourite choice.

8.5.3.5 CASE V: AP rejects bandwidth request and TBVH is very low

Suppose TBVH in this scenario is 20 s. The dynamic allocation of scores similar to table 8.20 gives WLAN a TBVH objective score of 9. Applying AHP we get the relative TBVH objective scores as WLAN = 0.449, UMTS = 0.449 and WLAN = 0.1. Further determination of final network ranks gives the score

WLAN = 0.3616, WIMAX = 0.4422 and UMTS = 0.3655.

This means the new network ranking order for this scenario is WIMAX, UMTS and WLAN.

Calculations from these five case scenarios successfully demonstrate that unlike the T-NS and other approaches which monotonously assign the same application stream to the same network at all times, SBM-NS makes an efficient utilisation of available information on network coverage together with the QoS availability on each interface to intelligently manage the allocation of resources to requesting streams. This is done in a manner which assists in preventing both upward and downward unnecessary vertical handovers.

These results prove the fundamental point that coverage information can play a crucial role in improving the management of application streams and resources when a device roams freely inside 4G heterogeneous networks.

8.6   Chapter Summary

This chapter has described in detail the functioning of the SBM layer and presented an elaborate explanation on generated results. It makes several important positive contributions to knowledge:

1. It provides additional proofs that reinforce the importance of the knowledge of network availability and the indispensible role it plays towards the correct decision-making for a wide range of mechanisms in a MN, from vertical handover optimisation to QoS and application performance management.

2. Where the popular approach has been to assign an application request to the network interface offering the best resources, the chapter challenges this approach and demonstrates through the performance evaluation of these approaches the detrimental effect it can have on an application's performance and the performance of the device in general.

3. The results put forward in this chapter gave a clear demonstration of how applying the knowledge of network coverage availability in the form of a quantitative measure can actually boost the intelligence and performance of QoS and resource management mechanisms.

4. Results also demonstrated how it is possible to apply available context information collectively in a unique yet simple manner to predict and avoid both upward and downward unnecessary vertical handovers.

Therefore the positive outcomes of this chapter and the previous one (chapter 7) succeeded collectively in providing answers to the second research question:

How can the knowledge of network coverage availability be utilised to

- **assist in the minimisation of unnecessary vertical handovers?**
- **optimise the performance of network selection and QoS management mechanisms in a heterogeneous client?**

The next chapter concludes this thesis by providing a comprehensive summary of proposed solutions together with a discussion on future works.

CHAPTER 9

Conclusions and Future Work

9.1   Introduction

This final chapter provides a succinct summary of the main ideas proposed in this thesis, the results, noteworthy achievements and the future applications of the proposed novel concepts. It captures the main theme of this research study and shows how it succeeded in answering the research questions.

9.2   How were the key research questions addressed?

The research study identified crucial gaps in knowledge arising due to uncertainty in coverage prediction in the field of 4G wireless networking. It revealed how in the handover decision phase it is crucial to obtain the answers to three fundamental questions '*When?*', '*Which?*' and '*How much?*' before connecting to a new network. These questions were then embedded into the two important research questions

- **How can a MN roaming within a network predict future network availability relative to its motion within the network, and determine how long it has before it performs a vertical handover?**

The answer to this research question was the proposal of a new metric called Time Before Vertical Handover which provided a precise measure of the time a device had in a WLAN network's coverage before it performed a vertical handover. Mathematical models for deriving TBVH for open and enclosed environments were proposed in Chapter 5. These models were then validated using simulation and results demonstrated the success of the first concept.

- **How can the knowledge of network coverage availability be utilised to**
  - **assist in the minimisation of unnecessary vertical handovers?**
  - **optimise the performance of network selection and QoS management mechanisms in a heterogeneous client?**

The answer to the second research question involved the development of a client-based Stream Bundle Management (SBM) layer for downward Quality of Service management which applied the new knowledge of TBVH to optimise the performance of network selection and QoS management mechanisms. This was achieved in chapter 7 in the thesis. Chapter 8 then demonstrated how the performance of network selection and QoS management mechanisms actually became more refined and accurate with the introduction of new intelligence on network coverage availability. This played a key role in the avoidance of unnecessary vertical handovers.

9.3   Main Contributions

As introduced in chapter 1, the main contributions of this research study include:

- An exhaustive critical review of existing solutions in the areas of handover optimisation, network selection and QoS management in 4G heterogeneous networks which unravelled important deficiencies that hampered the successful realisation of seamless connectivity and roaming.

- Simulated models for open, enclosed and closed environments that proactively derived new context information on geographical topology and network coverage.

- A new metric called Time Before Vertical Handover (TBVH) which provided a clear, quantitative measure of the duration of a network's availability. It displayed the ability to be applied easily to a wide variety of handover prediction and network selection algorithms to improve their performance.

- The client-based Stream Bundle Management layer for downward QoS that possessed the unique ability to process a large variety of complex context information in an efficient manner to minimise disruptions due to vertical handovers that could disrupt a smooth user experience.

- A network selection mechanism that took into account a large variety of information including user objectives, application requirements, network conditions, device position and mobility patterns, coverage availability and geographical context before bundling traffic streams on to available wireless channels.

- An intelligent resource management and call admission control mechanism that proactively negotiated resources with the network in order to minimise the occurrence of unnecessary vertical handovers.

194

9.4 Elaboration on the main contributions

9.4.1 Identification of crucial gaps in knowledge in the field of handover prediction and 4G QoS management

The study conducted an exhaustive literature survey of related work in the areas of vertical handover prediction and optimisation and downward QoS management. In the area of vertical handover prediction it unravelled how all these studies failed to come up with a tangible means of predicting network availability and the time a MN was expected to spend in a particular network before a vertical handover. In the area of QoS management in 4G heterogeneous client it demonstrated how the lack of TBVH knowledge led to incorrect decisions by QoS management and network selection techniques.

9.4.2 Unique set of mathematical models for capturing coverage and geographical context

This study proposed an improved extension to the forced vertical handover concept which enabled the definition of crisp coverage boundaries. A further enhancement was the development of geographical topology models that informed the MN about hard geographical boundaries within a network coverage which resulted in constrained movements for the MN. The precise knowledge of geographical topology in the form of coverage boundaries together with detailed MN context information in the form of velocity, position and direction was applied to develop a set of mathematical models for a variety of pedestrian environments. These models provided a precise quantitative measure of the time a MN had available to spend in a particular network, in this case WLAN. APs were equipped with the ability to notify the MN about surrounding context, enabling the device to make intelligent decisions about vertical handovers as early as the system discovery phase. With the help of the cosine function for detecting change in direction, the TBVH model unlike Received Signal Strength (RSS) based models further succeeded in capturing sudden changes in MN movements in indoor environments.

9.4.3 A new metric TBVH that provided a measure of network availability

The study demonstrated the application of available context information on coverage and device mobility in a unique manner for the dynamic derivation of a new parameter TBVH that provided a quantitative measure of network coverage availability. It was a common measure that possessed the potential to benefit a wide variety of handover optimisation solutions, specifically solutions in the areas of handover prediction, unnecessary handover avoidance, multi-stream QoS management and resource management. A unique feature of TBVH was that although it was derived from a large variety of device, network and geographical context information, it was in itself a simple measure that could be easily applied by a variety of mechanisms in the device in both real-time and simulated systems.

The proposed mathematical TBVH models were then validated through simulations using OPNET Modeler 15.0. This consisted of modelling a new client-based TBVH module which contained a set of sophisticated process models, one for each type of pedestrian environment. Simulation results proved the successful implementation of theoretical concepts and TBVH was successfully derived for open, enclosed and closed environments. The TBVH based handover prediction method showed key improvements over traditional RSS-based methods in terms of prediction accuracy. While RSS methods suffered from estimation errors up to 70% for pedestrian speeds, the TBVH-method was capable of delivering a good performance and accuracy for all types of speeds. Experimental results also demonstrated the method's resilience to location errors.

9.4.4 Intelligent client-based Stream Bundle Management Layer for downward QoS management

Another noteworthy contribution was the proposal of the client-based QoS management layer called Stream Bundle Management layer which resided in the QoS plane of the client and handled downward QoS. The SBM layer was an intelligent layer that applied the Analytic Hierarchy Process to process a large variety of information on user objectives, application demands, device context and network conditions to predict the best network option for a requesting application traffic stream. The design goal of the SBM layer was to achieve the optimum selection of networks and network

resources for requesting traffic streams in a manner which not only guaranteed the best QoS but which also prevented unnecessary upward and downward vertical handovers.

A unique feature of the SBM layer was its active application of TBVH knowledge to all its decision making procedures including network selection, resource allocation and call admission control (CAC). It simplified the process of specifying objectives for the end user by creating pre-defined user profiles containing the optimum combination of user objective priorities. A user simply selected the profile that best suited its needs. As for the avoidance of unnecessary vertical handovers, instead of adopting a simplistic velocity based approach, the TBVH-aware CAC mechanism adopted a more sophisticated per-flow approach. In order to decide whether a handover was necessary, the mechanism first closely examined the time needed by a requesting application to complete an ongoing transfer based on the number of resources allocated to it. If this time was less than the best network's TBVH the requesting stream was allocated to the network. If the required time was more than TBVH then the CAC mechanism actually negotiated with the AP for more resources so that the requesting stream could still be allocated to the best network without experiencing a vertical handover in the near future.

Results demonstrated the successful implementation of the SBM layer's network selection and QoS management mechanisms. Different scenarios demonstrated how through the correct application of TBVH and other advanced QoS management mechanisms, the SBM layer succeeded in selecting the best network each time along with the simultaneous avoidance of unnecessary vertical handovers, a key success none of the research studies so far have achieved. Overall, the SBM layer results proved the importance of the knowledge of network availability and the indispensible role it played in selecting the best network and resources for a requesting application stream. This successfully answered the second research question.

9.5   Future improvements to solutions from which the study can benefit

This section lists a set of improvements to the proposed solutions and related work upon which this study is based which will result in a further improvement in performance:

- Optimisations to the TBVH setup phase: While the task of deriving information required by the boundary AP to correctly configure coverage boundaries is not complicated for network

administrators, in order to make the technique more user-friendly it is necessary to develop a user interface that simplifies the process of AP setup for the layperson.

- Accurate prediction of MN path: The TBVH models consider realistic pedestrian movements where a MN's path is subjected to random and sudden changes in direction. A combination of cosine direction function, exit circle and average TBVH values are employed to reduce unpredictability associated with random movements closer to the network's coverage boundary. However a MN roaming inside the exit circle's coverage can still undergo frequent changes in direction. Although the cosine function is designed to detect these changes, it can still lead to unstable TBVH readings. Optimisation here means developing a new mobility model that accurately captures pedestrian movements, embedding it in the TBVH model.

- Modifications to beacon frames: the introduction of the topology clusters in AP beacon packets means the introduction of additional code that modifies the behaviour of MNs and enables them to decode the new information arriving in beacons. For commercial deployment it is important to ensure this modification can be accommodated by different WLAN extensions.

- TCP flow control issues: During the negotiation of additional resources by the SBM layer, to increase transmission rates it is assumed that both the sender and receiver side windows can accommodate the increase in data rates. This however cannot always be guaranteed in 4G networks due to the issue of widely varying characteristics of core and peripheral networks together with the issue of varying network characteristics of sender and receiver networks. TCP performance in 4G networks forms a complete new area of research.

- The simultaneous occurrence of vertical handovers at more than one network interface: This case although rare can cause a significant increase in computational overhead at the client and it is important to consider further ramifications for such a rare but valid case.

9.6   How can proposed solutions be applied in the real-world?

The solutions proposed in this study have led to the discovery of a new set of key intelligence which successfully enhances the ability of a MN to take complex decisions about vertical handover. These solutions can improve the seamless connectivity experience of every person with a mobile device in the 4G heterogeneous environment in a transparent and cost-effective way.

9.6.1 Integration with modern mobile devices

In today's urban scenario, Wi-fi access has penetrated in almost every home and institution. Let's take the example of the latest iPhone 3GS [iPhone 3GS website]. This is a device with four different wireless interfaces – WLAN, UMTS, GSM and Bluetooth. It is an intelligent device that is equipped with the ability to gather a wide range of context information ranging from indoor and outdoor location coordinates to user interests. This wireless device has been pre-programmed to automatically choose WLAN as the preferred interface for all file transfer connections above 10 Mb. iPhone users are also supplied with a set of global login details which enables them to automatically connect to free Wi-fi hotspots at home and at an increasing number of public places in the city. On a typical bus journey, it is interesting to note the increasingly large number of free Wi-fi hotspots the iPhone device detects when the user waits at the bus stop or when a bus stops at traffic lights.

In these precise situations the proposed TBVH models will enable the user to derive full benefit of cheap Wi-fi coverage through improved context knowledge. TBVH can specifically enhance the roaming experience of iPhone users by providing an intelligent set of mechanisms through which they are able to connect to the most cost-effective and QoS satisfying network interface for different applications. The knowledge of the duration of Wi-fi availability will allow the device to even make efficient use of temporary coverage if the coverage is available for the desired duration and offers the required QoS.

9.6.2 Application to emergency services

Emergency services are another domain where the improved knowledge of network availability can be of great benefit. A paramedic team communicating using a traditional low speed Wide Area Network can derive benefit from the intermittent yet high-speed availability of additional network resources at WLAN hotspots for the transfer of medical information such as high resolution X-ray scan images whenever their ambulance moves in to the coverage of a hotspot even if for a short duration.

The proposed solutions therefore promise to bring wireless connectivity in 4G networks a step closer to the seamless, hassle-free and pleasant experience for mobile users, both the techno-geek obsessed with the new technical features offered by the device and the non-technical user who

enjoys accessing information over the wireless Internet but dreads the technicality of sophisticated devices.

9.7   What are the future works that can be pursued based on this study?

The successful derivation of the new TBVH parameter for measuring network availability has created new opportunities for improved research activity in different areas of QoS management and handover optimisation. Let's take a further look at how the proposed research ideas can develop further and how other technical areas can derive benefit from the TBVH concept.

9.7.1   Development of TBVH models for other networks

The set of proactive TBVH models proposed by this research study were developed specifically for WLAN networks offering smaller pockets of high speed network coverage which could lead to more frequent vertical handovers. The study assumed an urban environment that offered uniform WAN coverage for networks higher up in the hierarchy like UMTS and WIMAX. TBVH priority scores for both networks were kept constant at the middle value 5 in order to allow a closer study of the WLAN TBVH mechanism. Extension of this work further means the development of TBVH models for UMTS and WIMAX particularly for rural environments. Due to the large cell size in this case it will be necessary to devise new methods for defining crisp cell boundaries as this is a key requirement for calculating TBVH correctly. Once successfully derived for larger cell sizes, TBVH possesses the potential of becoming a key parameter for validating the choice of a network for transmission.

9.7.2   TBVH to achieve maximum bandwidth utilisation for TCP

TBVH can help TCP achieve maximum bandwidth utilisation in the presence of multiple network channels. At present the SBM layer's awareness of the time a connection has access to a particular channel allows it to negotiate more resources with the AP which in turn allows TCP to increase its window size and transmit with a higher throughput. It will be interesting to study the actual behaviour of TCP in the presence of increased resources.

Another interesting front where applying TBVH can result in improved TCP performance is the bundling of one data connection onto multiple channels which will facilitate the exploitation of

situations where a MN experiences a temporary surge in total available bandwidth due to its movement into a temporary high-speed coverage such as WLAN [Radunovic,et al. 2008, Raiciu, et al. 2008]. Instead of performing a vertical handover twice it will be more efficient to start a new connection on the new interface for the duration of TBVH and transmit as many packets as possible. The source in this case first chops the entire file to be transmitted into a number of packets. The packets transmitted on the new interface need not be in sequence with the first connection's packets but need to form a sequence of their own. Upon arrival at the destination all packets are saved in a buffer. Upon the successful completion of the file's transfer the destination arranges all packets in the correct sequence and transfers to upper layers. This mechanism can speed up the transfer of large files without causing disruption to the original TCP connection due to vertical handovers.

9.7.3 Application of TBVH on other network selection mechanisms

Results in chapter 8 have clearly demonstrated how the SBM layer displays enhanced intelligence and improved performance over other network selection mechanisms. This opens a new door in the application of TBVH to other network selection mechanisms which can lead to a similar improvement in performance. This forms another interesting area of future research.

9.7.4 TBVH for enhanced performance of IEEE 802 standards

TBVH possesses the potential to both derive benefit and provide benefit to several IEEE 802.11 extensions.

The 802.11n extension [Bing. 2008] is known to employ multiple-input multiple-output technologies in the form of antennas to achieve data rates up to 600 Mbps or even higher. Instead of having one omni-directional antenna like existing extensions, the standard proposes four directional antennas installed on one AP, each one of them always transmitting in one particular direction. This feature can greatly benefit TBVH performance as it means greater cell sizes with reduced multi-path fading effects and blind spots together with crisper boundaries. TBVH will assist in the more efficient utilisation of bandwidth resource in 802.11n. Due to the high data rates of this standard the correct application of TBVH in call admission control will mean that even when a MN moves under temporary 802.11n, it will still be able to connect and benefit from the large availability of resources.

802.11r [Bing. 2008] proposes new solutions for fast BSS transitions. The extension is designed to provide fast and seamless handovers in order to address the needs of security, minimal latency and QoS resource reservation. This is achieved by allowing the MN to establish a security and QoS state at the new AP before making a transition. Under this standard, MNs can use the current AP as a passage to other APs, allowing the minimisation of disruption caused due to roaming transition. An interesting proposal to apply the concept of boundary AP to 802.11r is to form a mesh of boundary APs belonging to different basic service sets and making them the main point of contact for the transfer of context related information to other APs in the mesh. TBVH can also be used to allow the prompt transfer of context information about a MN from one boundary AP to another boundary AP at the right instant.

The 802.11u extension deals specifically with the issue of improved inter-working with external networks. Although still in an early stage of development, the aims of this promising new extension is to cover areas like improved enrolment, network selection, emergency call support, user traffic segmentation and service advertisement. The extension's aim is to improve the roaming experience for travelling users who will be able to select access to an external network base on provided services and conditions. As this is in line with the SBM layer's aim and objectives, it will be interesting to study how the SBM layer and 802.11u can complement each other to provide a greatly improved performance delivery.

Finally, the refined knowledge of network coverage boundaries through TBVH can actually help in enhancing the intelligence of IEEE 802.21 which is concerned with seamless handovers between heterogeneous wireless networks. TBVH can equip the standard with enhanced proactive knowledge on network context.

9.8   Concluding remarks

This thesis has addressed the key issues of Intelligent Proactive Handover and QoS Management mechanisms using TBVH in heterogeneous networks. We hope that this contribution will play a significant role in the development of future mobile heterogeneous networks.

APPENDIX A1

MISCELLANEOUS USEFUL INFORMATION

APPENDIX A1.1

OPNET's Modelling Domains

A1.1.1   Network Domain

The Network domain [OPNET Training Manual. 1990] defines the network's topology which forms the highest level in the hierarchy. It employs the created instances of nodes and communication links for the creation of communication networks and is the main place for the measurement and analysis of statistical information. The Project Editor is employed to create network models in the network domain. It provides the geographical context and size for the network such as PAN, LAN and MAN, assigning the correct distances to nodes which facilitates the calculation of communication delays. Nodes may be fixed or mobile. In the case of mobile nodes, the editor allows the definition of trajectories that portray the position of nodes as a function of time during simulation. The project editor also allows for the specification of nested hierarchies which helps to break down complexity. These hierarchies can be nested up to any depth. Figures F.1 and F.2 demonstrate the graphical representation of nested topologies and hierarchical communication networks in the project editor. Figure F.1 demonstrates a group of Wi-Fi hotspots deployed within the geographical context of New
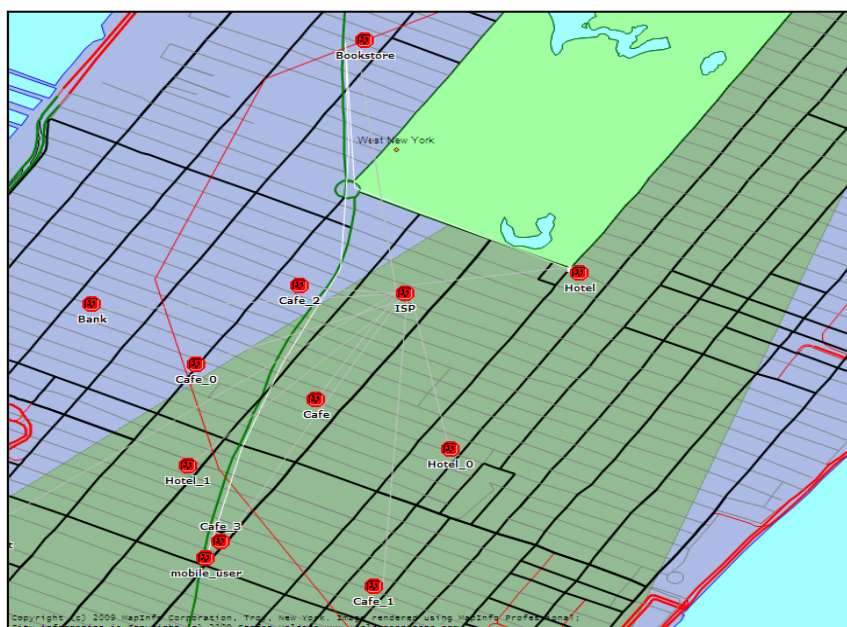
Figure F.1 Highest level network domain representation of WLAN hierarchy
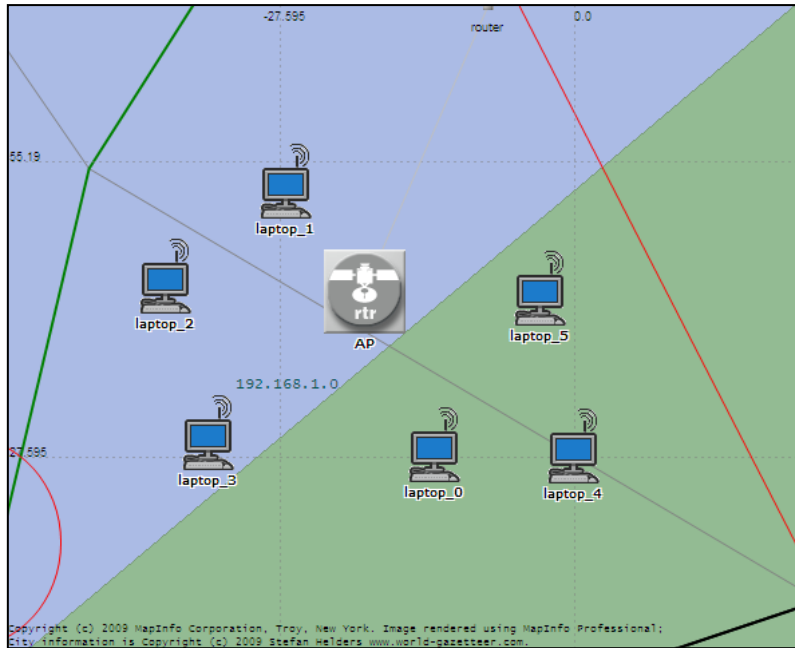
Figure F.2 The WLAN bank subnet at lowest network domain hierarchy level

York City, connected through a common Internet Service Provider. These detailed network models are available in the WLAN example project as part of the project library in Modeler. Figure F.2 represents the bank subnet at the lowest level in the subnet hierarchy which consists of a number of independent nodes communicating through the access point.

A1.1.2 Node Domain

The node domain [OPNET Training Manual. 1990] encompasses building blocks called modules and the connections between them that together form a network node. Instances of nodes are developed using the Node Editor whose modules include transmitters, receivers, processors, queues and sinks. Processors and queues are highly programmable modules that can be programmed to reflect the behaviour of the node's internal components e.g. the OSI layers. This can be done by assigning a process model to the processor or queue. Process models are discussed more in the next sub-section. Modules belonging to a node are connected using various different communication mechanisms called packet streams, statistic wires and logical connections.
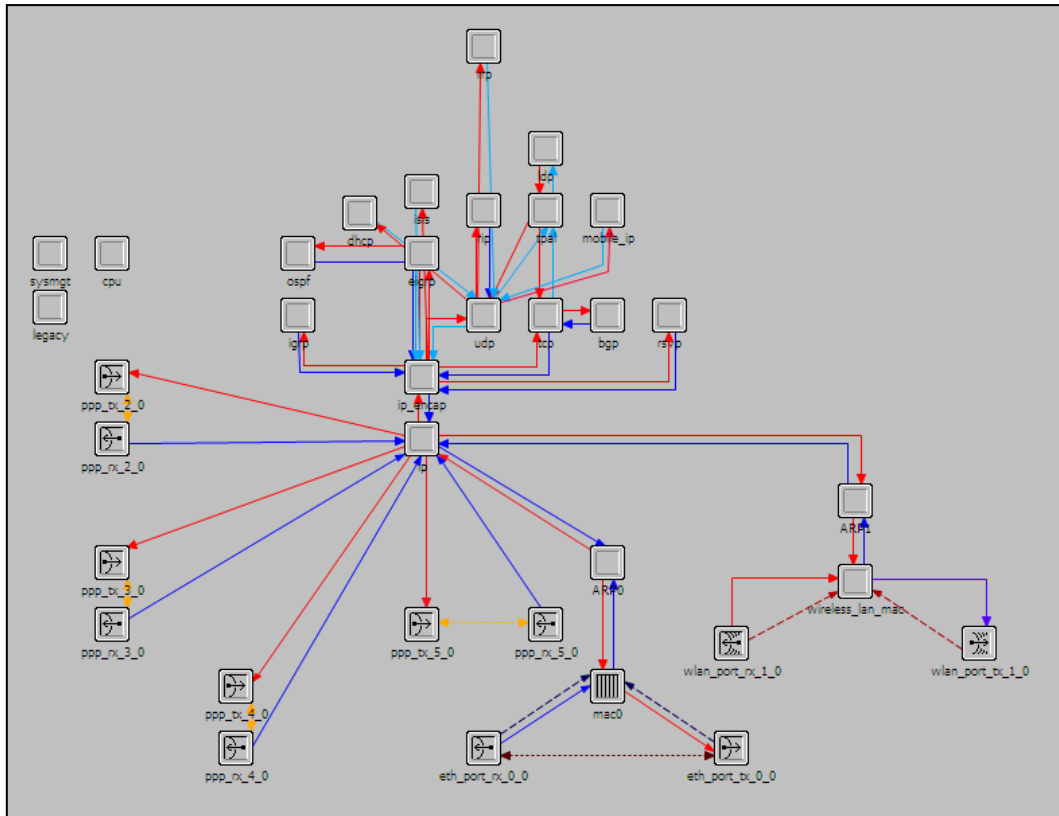
Figure F.3 *ethernet_slip4_adv* access point node model

Figure F.3 represents the node model for the e*thernet_slip4_adv* access point shown earlier in figure F.2. Plain boxes represent processor modules whose functioning is based on the process models assigned to them e.g. the IP module. Boxes with bars symbolise queue modules e.g. the MAC layer module. Boxes with arrows correspond to transmitter and receiver modules. Blue arrows connecting modules represent packet streams and are responsible for the actual transfer of packets between modules. Dashed lines are statistic wires that carry single statistic values from the source to a destination. Orange lines represent logical connections which are used to group sender-receiver pairs. In the access point node model each OSI layer is predefined and represented by an independent processor module. This is the case for almost all device node models in OPNET Modeler.

A1.1.3   Process Domain

The process domain [OPNET Training Manual. 1990] is where process models that define the behaviour of programmable elements like queues and processors are created. A process is the
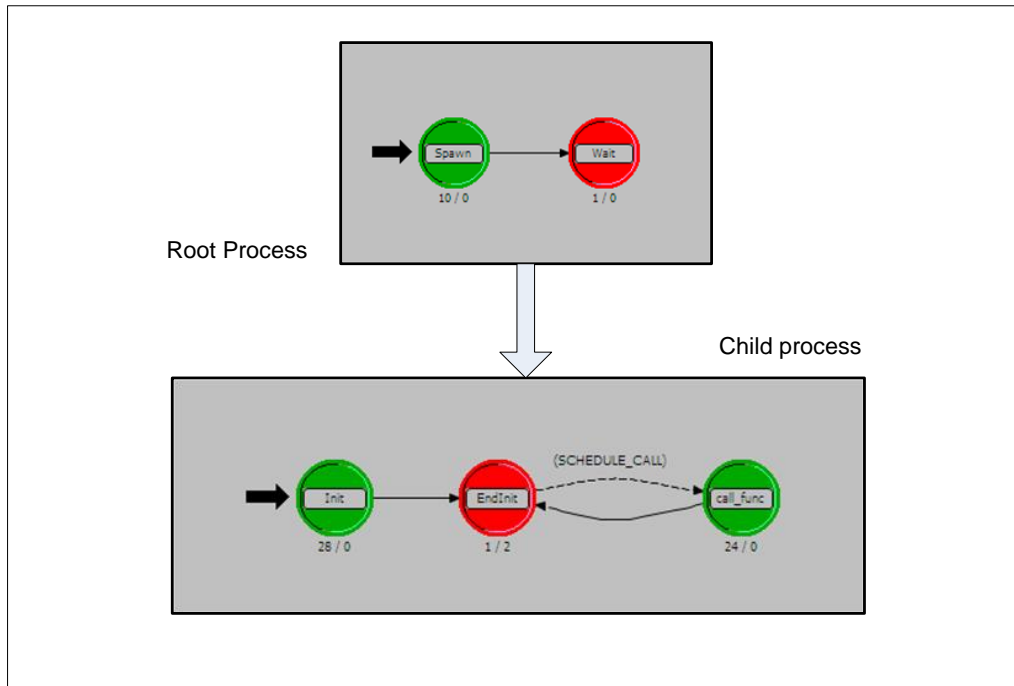
Figure F.4 State Transition Diagrams for the CPU process model

instance of a process model that operates within a module. Simulation execution begins at the root process and continues on to other processes which may be created dynamically as per need.

These newly created processes are called child processes whereas the creator processes are called the parent processes. When an invoked process begins execution, the invoker process is suspended until the invoked process either completes execution or is blocked by an event. The invoker process then regains control of the kernel and continues execution from the point where it had left off. Process models are a powerful feature of OPNET Modeler that allow for the creation of all types of network devices from scratch. A process model is represented graphically by a State Transition Diagram (STD) that captures the behaviour of the module and its response to various interrupt triggers.

Figure F.4 represents the State Transition Diagrams for the CPU module in the *ethernet_slip4_adv* access point. It consists of circular nodes called states which are resting places between interrupts. They consist of code blocks specifying the model's behaviour. These are executed when the control is passed on to the process by the kernel in response to an interrupt. In the figure, a black arrow next to state indicates it is the initial state where execution starts. Transitions between states are arrows that indicate the path the execution will take in response to a particular event. A transition can have a condition statement which specifies the requirements that must be met by the process to follow the transition.

APPENDIX A1.2

A1.2.1 Defining objective goals for user profiles

Table A.1 User profile defining objective goals

User Profile	Defining Objective
Student	Low cost, acceptable QoS
Home	Low cost, good QoS (value for money)
Business/Professional	Excellent QoS, reasonable cost

A1.2.2 Interface priority scores

Table A.2 Interface priority scores

Scores	Voice	Video-Int	Video-Strm	Data
1	UMTS	WLAN	WLAN	WLAN
4	WLAN	WIMAX	WIMAX	WIMAX
7	WIMAX	UMTS	UMTS	UMTS

A1.2.3 Objective priorities per application

Table A.3 Video interactive traffic objective priorities

Priority Score	Student	Home	Business
1	TBVH	TBVH	TBVH
3	Cost	INTP	QoS
5	QoS	QoS	INTP
7	INTP	Cost	Cost

Table A.4 Video streaming traffic objective priorities

Priority Score	Student	Home	Business
1	TBVH	TBVH	TBVH
3	Cost	INTP	QoS
5	QoS	QoS	INTP
7	INTP	Cost	Cost

Table A.5 Data transfer traffic objective priorities

Priority Score	Student	Home	Business
1	TBVH	TBVH	TBVH
3	Cost	INTP	QoS
5	INTP	Cost	Cost
7	QoS	QoS	INTP

A1.2.4 Application priority scores

Voice = 1

Video-interactive = 3

Video-streaming = 5

Data transfer = 7

APPENDIX A1.3

Abstracts of published research papers

A1.3.1   In Proceedings of IEEE ICTTA'04

Critical analysis of high speed wireless standards for transferring telementoring information.

F. Shaikh, A. Lasebae and G. Whitney

School of Computing Science, Middlesex University, London.

The main focus of this paper is to investigate the requirements for developing a feasible solution for the transmission of Telesurgical (mainly Telementoring) audio and video information over high speed wireless links. This requires an improvement in the quality of service of the wireless link between two sites in terms of throughput, delay, delay variations, stability and security. As the high level features of current broadband wireless technologies can be employed to provide the high-speed link between the two sites, this paper also critically analyses and evaluates the functionality issues of current and upcoming wireless standards such as Spectrum Management, Media Access Technologies, and Authentication Access and Control. It also investigates their suitability for the transmission of Telesurgical information. It then classifies the wireless technologies into different categories based on their ability to transfer information over long and short distances. The analysis carried out in this paper would lay out the guidelines for selecting an appropriate wireless standard for transferring Telesurgical information.

A1.3.2   In Proceedings of MIT'05

SBM layer for optimum management of co-existing telemedicine traffic streams in heterogeneous networks.

F. Shaikh, A. Lasebae and G.Mapp.

School of Computing Science, Middlesex University, London.

Heterogeneous networks promise users ubiquitous and seamless networking anytime, anywhere, with access to multimedia and real-time applications. As these networks facilitate easy and cost-effective penetration of medical advice in both rural and urban areas, there exists a tremendous potential in this field that can be harnessed to increase the radius of availability of health-care. However, disparate characteristics of different wireless networks lead to noticeable variations in network conditions when users roam among them e.g. during

vertical handovers. Telemedicine traffic consists of a variety of real-time and non real-time traffic streams, each with a different set of Quality of Service requirements. To ensure satisfactory user experience while roaming among different networks, it is vital to maintain a stable and balanced flow of multi-class traffic across a wireless channel under varying network conditions. This paper discusses the challenges and issues involved in the successful adaptation of heterogeneous networks by wireless telemedicine applications. We propose the development of a Smart Bundle Management (SBM) Layer for optimally managing co-existing traffic streams under varying channel conditions in a heterogeneous network. The SBM Layer acts as an interface between the applications and the underlying layers for maintaining a fair sharing of channel resources through implementation of internal priority management algorithms. Coloured Petri nets are used to present a framework of the basic design of the SBM layer along with analysis and evaluation of its functioning and correctness.

A1.3.3   In Proceedings of IEEE ICTTA'06

Client-based SBM Layer for Predictive Management of Traffic Flows in Heterogeneous Networks.

F. Shaikh, A. Lasebae and G. Mapp

School of Computing Science, Middlesex University, London.

In a heterogeneous networking environment, the knowledge of the time before a vertical handover (TBVH) for any network is vital in correctly assigning connections to available channels. In this paper, we introduce a predictive mathematical model for calculating the estimated TBVH component from available network parameters and discuss the different scenarios that arise based on a mobile host's trajectory. We then introduce the concept of an intelligent Stream Bundle Management Layer (SBM) which consists of a set of policies for scheduling and mapping prioritised traffic streams on to available channels based on their priority, device mobility pattern and prevailing channel conditions. The layer is also responsible for the maintenance of connections during vertical handovers to avoid their forced termination.

A1.3.4   In Proceedings of WINSYS'06

An Architectural Framework for Heterogeneous Networking

G. Mapp (1), D. Cottingham (2), F. Shaikh (1), P. Vidales (3), L. Patanapongpibul (4), J. Baliosian (5), J. Crowcroft (2)

(1) School of Computing Science, Middlesex University, London.

(2) Computer Laboratory, University of Cambridge, Cambridge.

(3) Deutsche Telekom Laboratories, Berlin..

(4) Samsung Electronics Research Institute, Staines, Middlesex.

(5) Network Management Centre, Ericsson, Ireland.

The growth over the last decade in the use of wireless networking devices has been explosive. Soon many devices will have multiple network interfaces, each with very different characteristics. We believe that a framework that encapsulates the key challenges of heterogeneous networking is required. Like a map clearly helps one to plan a journey, a framework is needed to help us move forward in this unexplored area. The approach taken here is similar to the OSI model in which tightly defined layers are used to specify functionality, allowing a modular approach to the extension of systems and the interchange of their components, whilst providing a model that is more oriented to heterogeneity and mobility.

A1.3.5　　In Proceedings of WICON'07

YComm: A Global Architecture for Heterogeneous Networking

G. Mapp (1), F. Shaikh (1), J. Crowcroft (2), D. Cottingham (2).

(1) School of Computing Science, Middlesex University, London.

(2) Computer Laboratory, University of Cambridge, Cambridge.

In the near future mobile devices with several interfaces will become commonplace. Most of the peripheral networks using the Internet will therefore employ wireless technology. To provide support for these devices, this paper proposes a new framework which encompasses the functions of both peripheral and core networks. The new architecture is called Y-Comm and is defined in a layered manner like the OSI model.

A1.3.6　　In Proceedings of IEEE NGMAST'07

Proactive Policy Management using TBVH Mechanism in Heterogeneous Networks.

F. Shaikh, G.Mapp, A. Lasebae

School of Computing Science, Middlesex University, London.

In order to achieve seamless interoperability in heterogeneous networking, it is vital to improve the context-awareness of the mobile node (MN) so that it is able to predict future network conditions with sufficient accuracy. In this paper, we introduce a predictive mathematical model for calculating the estimated Time Before Vertical Handover (TBVH) component from available network parameters. The model is practically implemented in OPNET and our simulation results confirm the validity of the concept. We then demonstrate how the knowledge of TBVH along with other network parameters can be applied by downward Quality of Service management policies which bundle multi-class traffic streams on to available network channels based on application QoS, device mobility patterns and prevailing channel conditions.

A1.3.7 In Proceedings of IEEE ICTTA, 08

Proactive Policy Management for Heterogeneous Networks

F. Shaikh, A. Lasebae, G. Mapp

School of Engineering and Information Sciences, Middlesex University, London.

Context-awareness is a vital requirement of heterogeneous devices which allows them to predict future network conditions with sufficient accuracy. In this paper we present a proactive modelling-based approach for policy management which allows the mobile node to calculate Time Before Vertical Handover for open and closed environments. The paper explains how the knowledge of this component can improve the manner in which multi-class traffic streams are allocated to available network channels. Simulation results confirm the feasibility of the concept.

A1.3.8 In Proceedings of EUPS'09

Exploring Efficient Imperative Handover Mechanisms for Heterogeneous Wireless Networks

G. Mapp (1), F. Shaikh (1), M. Aiash (1), R.P. Vanni (2), M. Augusto (2), E. Moreira (2)

(1) School of Engineering and Information Sciences, Middlesex University, London.

(2) ICMC, University of Sao Paulo, San Carlos, Brazil.

The Next Generation Internet will provide ubiquitous computing by the seamless operation of heterogeneous wireless networks. It will also provide support for quality-of-service (QoS), fostering new classes of applications and will have a built-in multi-level security environment. A key requirement of this new

infrastructure will be support for efficient vertical handover. Y-Comm is a new architecture that will meet the challenge of this new environment. This paper explores the design of efficient imperative handover mechanisms using the Y-Comm Framework. It first looks at different types of handovers, then examines the Y-Comm Framework and shows how Y-Comm maps unto current mobile infrastructure. It then explores support for different handover mechanisms using Y-Comm. Finally, it highlights the development of a new testbed to further investigate the proposed mechanisms.

APPENDIX A2

SELECTED RESEARCH PAPERS

Appendix A2.1 Paper published in EUPS'09, August 2009

Exploring Efficient Imperative Handover Mechanisms for Heterogeneous Wireless Networks

Glenford Mapp, Fatema Shaikh and Mahdi Aiash
School of Engineering and Information Sciences
Middlesex University, Hendon, London NW4 4BT
Telephone: +44 208411 6871
Fax: +44 208411 4594
Email: g.mapp, f.shaikh, m.aiash@mdx.ac.uk

Renata Porto Vanni, Mario Augusto and Edson Moreira
ICMC, Univerity of Sao Paulo,
San Carlos, Brazil
Telephone: +55 16 33739699
Fax: +55 16 33739751
Email: rporto.usp, marioeaugusto@gmail.com, edson@icmc.usp.br

Abstract—The Next Generation Internet will provide ubiquitous computing by the seamless operation of heterogeneous wireless networks. It will also provide support for quality-of-service, QoS, fostering new classes of applications and will have a built-in multi-level security environment. A key requirement of this new infrastructure will be support for efficient vertical handover. Y-Comm is a new architecture that will meet the challenge of this new environment. This paper explores the design of efficient imperative handover mechanisms using the Y-Comm Framework. It first looks at different types of handovers, then examines the Y-Comm Framework and shows how Y-Comm maps unto current mobile infrastructure. It then explores support for different handover mechanisms using Y-Comm. Finally, it highlights the development of a new testbed to further investigate the proposed mechanisms.

Index Terms—Next Generation Internet, Vertical Handover, Architectural Framework, Mobility

I. INTRODUCTION

The Next Generation Internet (NGI) will provide ubiquitous computing via the seamless operation of heterogeneous wireless networks including WLAN, 3G, WiMax, Ultrawideband, etc. Using these networks, users will be continuously connected to the Internet as they move around. Vertical handover [1] which allows mobile nodes to seamlessly switch their connections from one network to another is a key mechanism that must be supported in NGI. In order to do this effectively, it is necessary to gather extensive information about various system parameters including the state of individual wireless network interfaces as well as the state of transport connections.

In a broader context, the widespread use of wireless technologies has highlighted a significant evolution in the architecture of the Internet. In terms of performance, it is now possible to divide the Internet into two distinct parts: a core network and edge or peripheral networks. The core network consists of a super-fast backbone and fast access networks which are attached to the backbone. Peripheral networks will be dominated by the deployment of different wireless technologies. This means that the characteristics of the core network will be very different to the peripheral wireless networks on the edge.

This change needs to be reflected in a new networking architecture which attempts to clearly define the functions,

their order and the interlocking relationships that are necessary to support heterogeneous networking. Recently, a new architecture called Y-Comm [2] has been designed to capture this new reality. This paper shows in some detail how different types of imperative handovers can be supported using the Y-Comm architecture. The rest of paper is structured as follows: Section 2 looks at a classification for vertical handovers. Section 3 looks at previous work while Section 4 describes the Y-Comm Framework. Section 5 shows how Y-Comm maps unto current mobile infrastructure while Section 6 discusses vertical handovers in Y-Comm. Section 7 explores current work and the paper concludes with a section of conclusions and future work.
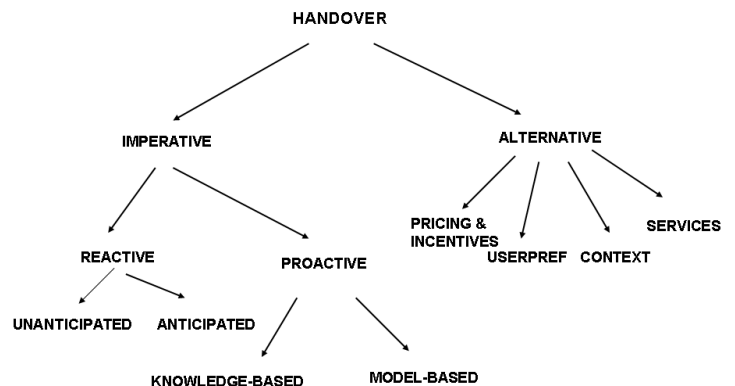

Fig. 1: Vertical Handover Classification

II. VERTICAL HANDOVER - A DETAILED CLASSIFICATION

In this section we take a detailed look at vertical handover to pinpoint its different classes. Firstly we introduce the concepts of hard and soft handovers. Hard handovers occur when the current attachment is broken before the new connection is

established while in soft handovers, the current connection is broken after the new connection is established.

Another important operational factor is the entity that makes the decision to do a vertical handover. The options basically are network-controlled handover in which the decision to implement handover is taken by the network(s) to which the mobile node is currently attached. The second is called client-based handover in which the client is the deciding entity. Though Y-Comm can facilitate network-controlled handover, client-based handover is favoured as a more elegant solution [3]. This is because client-based handover is more scalable as the mobile node can easily monitor the necessary parameters from its wireless interfaces. In addition, client-based handover allows the mobile node to look at other issues such as the state of its TCP connections. A general classification of handover is shown in Figure 1.

Imperative handovers occur due to technological reasons only. Hence the mobile node changes its network attachment because it has determined by technical analysis that it is good to do so. This could be based on parameters such as signal strength, coverage, the quality-of-service offered by the new network. These handovers are imperative because there may be a severe loss of performance or loss of connection if they are not performed. In contrast, alternative handovers occur due to reasons other than technical issues [4]. Hence there is no severe loss of performance or loss of connection if an alternative handover does not occur. The factors for performing an alternative handover include a preference for a given network based on price or incentives. User preferences based on features or promotions as well as contextual issues might also cause handover. Finally there may be other network services that are being offered by certain networks. In this paper we concentrate on imperative handovers.

Imperative handovers are in turn divided into two types. The first is called reactive handover. This responds to changes in the low-level wireless interfaces as to the availability or non-availability of certain networks. Reactive handovers can be further divided into anticipated and unanticipated handovers [5]. Anticipated handovers are soft handovers which describe the situation where there are alternative base-stations to which the mobile node may handover. With unanticipated handover, the mobile is heading out of range of the current attachment and there is no other base-station to which to handover. These handovers are therefore examples of hard handovers.

The other type of imperative handover is called proactive handover. These handovers use soft handover techniques. Proactive handover policies attempt to know the condition of the various networks at a specific location before the mobile node reaches that location. Proactive policies allow mobiles nodes to calculate the Time Before Vertical Handover (TBVH) which will allow the mobile node to minimize packet loss and latency experienced during handovers. Proactive handovers therefore represent a mechanism that could be used to support seamless handover as it allows the system and applications more time to deal with handover issues. Presently, two types of proactive handovers are being developed. The first is knowledge-based and attempts to know by measuring beforehand the signal strengths of available wireless networks over a given area such as a city. This could involve physically driving around and taking these readings [6]. The second proactive policy is based on a mathematical model which calculates the point when vertical handover should occur and the time that the mobile would take to reach that point based on its velocity and direction [7].

III. PREVIOUS WORK

Work on handovers has been going on for sometime. Most of the research done by the mobile operators focused on network-controlled horizontal handover where handover is done between adjacent cells of the same network. The development of models to understand whether handover should be done given the relative load on individul cells based on on-going calls, new calls being made within the cells and incoming calls due to handover from nearby cells was a major goal.

With the introduction of Mobile IPv4 and Mobile IPv6 [8], client-based handover began to be investigated. For these mechanisms, handover latency is high because they only work at the network level as they are based on Router Advertisements (RAs) which are relatively slow. In order to reduce this latency, Fast Mobile IPv6 (FMIPv6) [9] makes use of L2 events and triggers to reduce handover latency.
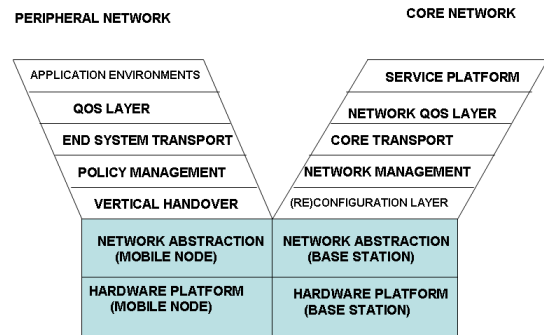
Fig. 2: The Y-Comm Framework

The study of vertical handovers was greatly enhanced with the deployment of the Cambridge Wireless Testbed [10], which was the first testbed to study client-based vertical handovers. The testbed used the Vodafone 3G Experimental network, with Home and Foreign WLANs and a wired IPv6 LAN. Using the testbed, PROTON [11], a policy manager for reactive handovers was developed. PROTON was implemented using a 3-layer structure. Y-Comm is a direct follow-on from the Cambridge Wireless Testbed. It should improve the handover process by dynamically supporting all types of handover

including proactive ones which allow the system to acquire resources long before handover will occur and so prevent a loss of performance.

IV. THE Y-COMM ARCHITECTURE

The Y-Comm Architecture is a new architecture to support heterogeneous networking. It uses two frameworks. The first is called the Peripheral Framework and deals with operations and functions on the mobile node. The other framework is called the Core Framework and shows the functionality required in the core network to support the Peripheral Framework. The structure of the Y-Comm architecture is shown in Figure 2. A brief explanation of Y-Comm is now attempted starting with the lowest layer. A more detailed explanation can be found in [12], [13].

A. The Peripheral Framework

The Hardware Platform Layer (HPL) is used to classify all relevant wireless technologies. Hence different wireless technologies which are characterised by the electromagnetic spectrum, MAC and modulation techniques make up this layer. The Network Abstraction Layer (NAL) provides a common interface to manage and control all the wireless networks. These first two layers for both frameworks are similar in functionality. In the Peripheral Framework, the Hardware Platform and the Network Abstraction layers run on the mobile to support various wireless network technologies while in the Core Framework these two layers are used to control the functions of base stations of different wireless technologies.

The Vertical Handover Layer (VHL) executes vertical handover. So this layer acquires the resources for handover, does the signalling and context transfer for vertical handover. The Policy Management Layer (PML) decides whether and when handover should occur. This is done by looking at various parameters related to handover such as signal strength and using policy rules to decide both the time and place for doing the handover.

The End Transport Layer (ETL) is used to provide network and transport functions to the mobile nodes in peripheral networks. It allows the mobile node to make end-to-end connections across the core network. The QoS Layer (QL) in the Peripheral Framework supports two mechanisms for handling QoS. The first is defined as Downward QoS. This is where an application specifies its required quality-of-service to the system and the system attempts to maintain this QoS over varying network channels. The other definition is Upward QoS where the application itself tries to adapt to the changing QoS. This layer also monitors the QoS used by the wireless network as a whole to ensure stable operation. The final layer of the Peripheral Framework is called the Applications Environments Layer (AEL). This layer specifies a set of objects, functions and routines to build applications which make use of the framework.

B. The Core Framework

As previously mentioned, the first two layers of the Core Framework are engaged in controlling base-station operations.

The third layer is called the Reconfiguration Layer (REL). It is a control plane to manage key infrastructure such as routers, switches, and other mobile network infrastructure using programmable networking techniques [14]. The Network Management Layer (NML) is a management plane that is used to control networking operations in the core. This layer can divide the core into a number of networks which are managed into an integrated fashion. It also gathers information on peripheral networks such that it can inform the Policy Management Layer running on mobile nodes about wireless networks at their various locations.
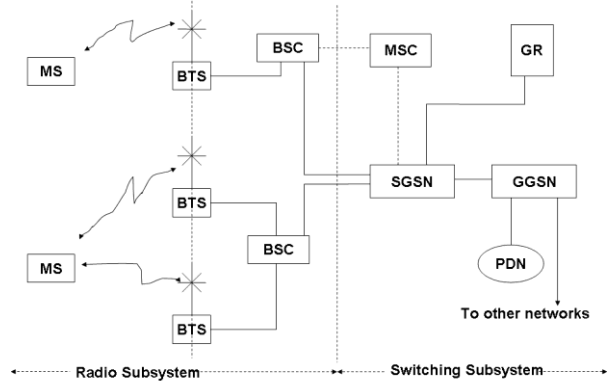

Fig. 3: The GSM/GPRS Network

The next layer, called the Core Transport System (CTS), is concerned with moving data through the core network. Where the peripheral networks join the core network is called a core endpoint. Core endpoints are usually situated in access networks and several peripheral networks may be attached to a single core endpoint. CTS is concerned primarily with moving data between core endpoints with a given QoS and a specified level of security.

The Network QoS Layer (NQL) is concerned with QoS issues within the core network especially at the interface between the core network and the peripheral networks. A main concern of this layer is to prevent overloading. In this regard, admission control techniques are applied by the NQL to prevent new streams or mobiles doing vertical handovers from overloading core endpoints or associated peripheral networks. Finally the Service Platform Layer (SPL) allows services to be installed on various networks at the same time.

V. MAPPING Y-COMM ONTO MOBILE INFRASTRUCTURE

In this section we show the relationship between Y-Comm and current mobile infrastructure. We believe that Y-Comm can easily be mapped onto well-established networks such as the GSM/GPRS architecture [15]. The GSM architecture was developed by the European Telecommunications Standards Institute (ETSI) and remains the most popular mobile infrastructure ever deployed. The GSM/GPRS infrastructure is shown in Figure 3. The mobile node runs the GSM/GPRS protocol

stack while the required network functionality is distributed using several core entities. The Base station transceivers (BTS) interact directly with the mobile node using specified radio channels. Each BTS is controlled by a Base Station Controller (BSC) while each BSC is controlled by a Mobile Switching Centre (MSC) for voice traffic or a Service GPRS Support Node (SGSN) for data traffic. The Gateway GPRS Support Node (GGSN) serves as a gateway to other networks with the help of the Gateway Register (GR).
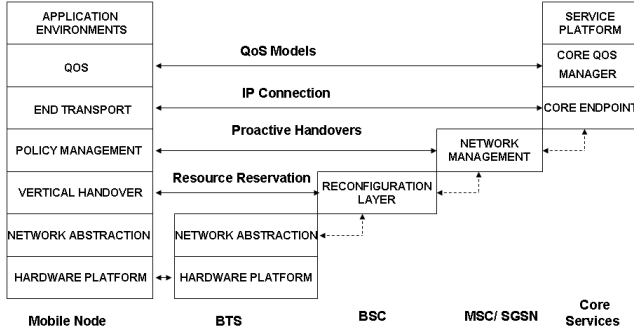

Fig. 4: Mapping Y-Comm onto Mobile Infrastructure

We can now show how the functions of Y-Comm can be mapped onto the GSM infrastructure making possible the transition from GSM to Y-Comm. This is shown in Figure 4. The Mobile Node (MN) runs the entire Peripheral Framework as shown. The Core Framework is distributed throughout the core network in a similar way to the GSM/GPRS infrastructure. The Hardware Platform and Network Abstraction Layers run in the Base Transceivers. Y-Comm however, supports BTSs of different wireless technolgies including 3G base-stations, Wi-Fi and WiMax APs, etc. The Reconfiguration Layer of Y-Comm runs in the Base-station Controllers for GSM or access routers in WLANs and LANs. This layer uses programmable techniques on the Network Abstraction Layer to control the resources on individual BTSs. It is expected that each Y-Comm BSC would control one wireless technology. The Reconfiguration Layer on the BSC allocates resources to do a handover to a particular BTS.

The Network Management Layer (NML) manages different wireless networks and runs at the level of the MSC/SGSN level in current mobile infrastructure. In Y-Comm, a local NML manages all the BSCs in a local area and knows the status of each wireless network and its topology. This information can be shared with the Policy Management Layer on the mobile node. The core endpoint is used by the mobile node to connect to the wider Internet. For a given connection, IP packets to and from the mobile node are tunnelled through the core network using core endpoints. Finally when an application on
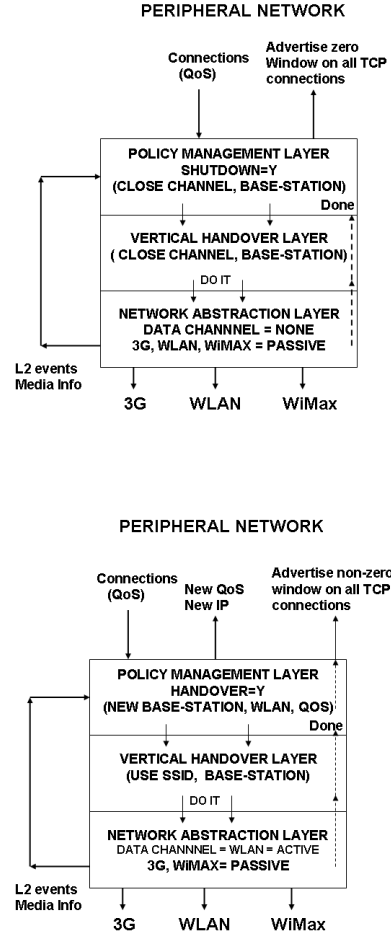
Fig. 5: Unanticipated Handover: Stages 1 and 2

the mobile node wishes to make a connection through the core network, the QoS layer running on the mobile node interacts with the QoS manager in the core network with regard to QoS requirements for the new connection. The QoS manager will return two core endpoints which can be used for the new connection.

VI. VERTICAL HANDOVER IN Y-COMM

In this section we look at the various layers of Y-Comm that are involved in vertical handover. Y-Comm supports both reactive and proactive handovers.

A. Reactive Handovers

With reactive handovers, the main inputs into the Policy Management layer are the L2 events and Media Information from the Network Abstraction layer which monitors the different network interfaces. The state of ongoing TCP connections and their required QoS are also monitored. The mobile node makes the decision to handover based on these factors only.

1) Unanticipated Handovers: Unanticipated handovers result in a two-stage interaction. The first stage occurs when the Policy Management Layer is informed that the signal strength from the current base station is fading fast and there is no other base station in the vicinity. The PML will first instruct the TCP connections to advertise a zero receive window on all its connections. This stops senders from sending data while the mobile node is unconnected from the network. It then instructs the Vertical Handover Layer to close the present channel. This is shown in Figure 5.

In the second stage of the unanticipated handover occurs when the mobile node finds a new base-station as its next point of attachment. It first signals to the vertical handover layer to acquire a channel if it is a 3G base station or obtain the SSID of the WLAN network. It is worth noting that for reactive unanticipated handovers the core network cannot be involved as there is no current connection to the core network. Hence the mobile node must acquire an unreserved channel on the BTS. Once this is done, a new IP address and a new QoS are communicated to the upper layers. The system then tells the TCP connections to advertise a non-zero window. This is shown in Figure 5.

2) Anticipated Handovers: In this section we explore anticipated handovers in which there are alternative base-stations to which to handover. In Figure 6, we look at an anticipated handover to a WLAN workstation. In this example, it decides to handover to a WLAN base-station from which it is already receiving beacon frames. The Vertical Handover Layer which is responsible for the handover, uses the Reconfiguration Layer to obtain the valid SSID for the network. The REL uses its programmable interface to get the SSID key from the base-station. It passes it back to the vertical handover layer which then does the handover. This sequence can be represented by the message exchange diagram shown in Figure 7.

The new QoS and new IP address are signalled up to the End Transport and QoS layers. When the handover has occurred, each layer signals to its upper layer that handover has been completed. In order to quickly restore communications in anticipated handovers, the transport level keeps a copy of the last TCP acknowledgement for each connection. After handover, the previous channel is released and the mobile node implements a fast retransmit algorithm, i.e., it sends the last TCP acknowledgement 3 times and this results in fast retransmission where any packets lost during the handover are retransmitted.

B. Proactive Handovers

This section looks at proactive handovers. This is shown in Figure 8. Since proactive handovers attempt to determine when and where handover should occur, it is necessary to have a knowledge of networks in the local area where the mobile is located. In addition, in order to perform vertical handover using a mathematical model approach, it is also necessary to know the topology of these local networks. In Y-Comm, this information is managed by the Network Management Layer in the Core Framework. The mobile node therefore polls the
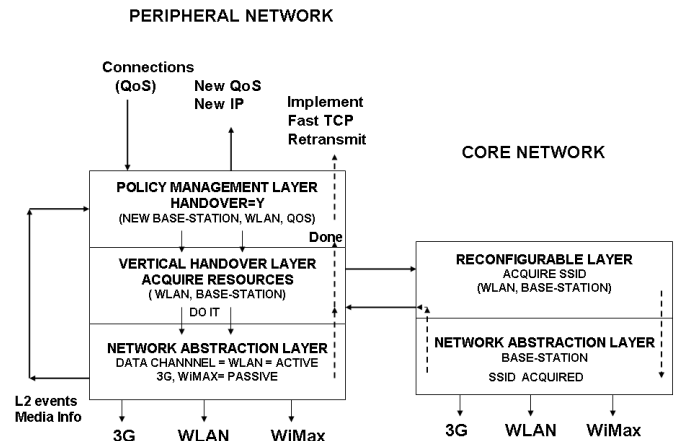

Fig. 6: Anticipated Handover
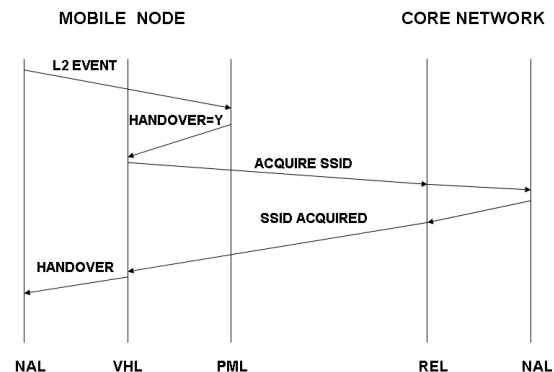
Fig. 7: Anticipated Handover Sequence

NML to obtain information with regard to all local wireless networks, their topologies and QoS characteristics. This information along with the direction and speed of the mobile as well as the QoS of on-going connections are used by the Policy Management Layer to determine where and when handover should occur. The PML calculates TBVH - the period after which handover will occur. This information is communicated to the Vertical Handover Layer which immediately requests resources to do a handover. Even though the resources are acquired early, handover actually takes place when TBVH expires. This sequence is shown by the message exchange diagram shown in Figure 9.

In addition, once the PML decides to handover, the new IP address, the new QoS as well as TBVH are communicated to the upper layers. Given TBVH, the upper layers are expected to take the necessary steps to avoid any packet loss, latency or slow adaptation. For example, it may be possible for the End-

Transport Layer to signal an impending change in the QoS on current transport connections and to begin to buffer packets ahead of the handover. After handover, the previous channel used by the mobile node is released.

VII. CURRENT WORK

Reactive policies were explored using the Cambridge Wireless Testbed. However, proactive policies have only been investigated using simulations. Detailed results were presented in [16] and show improvements in handover performance. The proposed mechanisms need to be tested in a real environment. Work has begun on a Y-Comm testbed which will be used to test algorithms and mechanisms needed to implement the Y-Comm architecture including the vertical handover mechanisms discussed in this paper. The testbed will initially support vertical handover between WLAN, 3G and LAN systems.
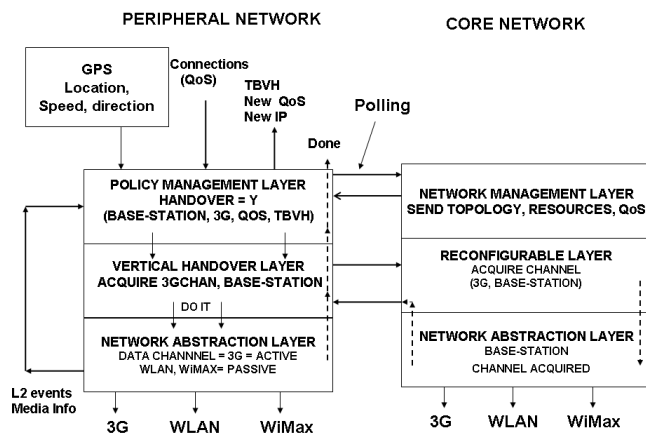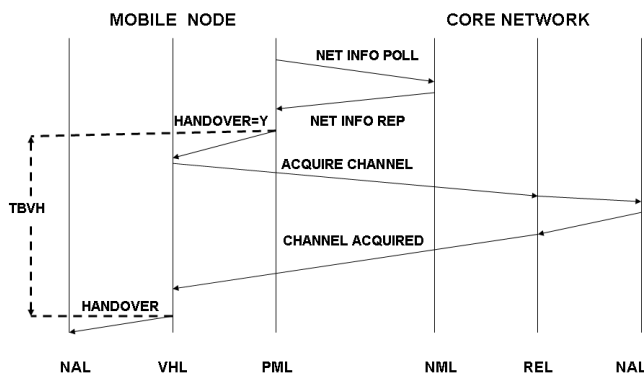
Fig. 8: Proactive Handover

Fig. 9: Proactive Handover Sequence

VIII. CONCLUSIONS AND FUTURE WORK

This paper has detailed the mechanisms to support efficient vertical handover using the Y-Comm Framework. We believe that the adoption of proposed mechanisms will enhance seamless connectivity. Work is proceeding to build a testbed to test the working and performance of these algorithms in real environments.

ACKNOWLEDGMENTS

REFERENCES

[1] J. McNair and F. Zhu, "Vertical Handoffs in Fourth-Generation Multinetwork Environments," in *IEEE Wireless Communications*, vol. 11, 2004.
[2] Networking Research Group, EIS, Middelsex University, "The Y-Comm Research WebPage." [Online]. Available: http://www.mdx.ac.uk/eis/research/groups/YComm.asp
[3] L. Patanapongpibul and G. Mapp, "A Client-Based Handoff Mechanism for Mobile IPv6 Networks," in *Proceedings of 8th IEEE Symposium on Computers and Communications(ISCC)*. IEEE Society Press, June 2003.
[4] S. Bansal, L. Lasebae, and R. Comley, "Freedom to Choose along with Freedom of Mobility," in *Proceedings of ICCTA'04*, 2004.
[5] L. Patanapongpibul, G. Mapp, and A. Hopper, "An End-System Approach to Mobility Management for 4G Networks and its Application to Thin-Client Computing," *ACM Mobile Computing and Communications Review (ACM SIGMOBILE)*, July 2006.
[6] D. Cottingham, I. Wassell, and R. Harle, "Performance of IEEE 802.11a in Vehicular Contexts," in *IEEE Vehicular Technology Conference*, July 2007, pp. 854–858.
[7] F. Shaikh, A. Lasebae, and G. Mapp, "Proctive Policy Management for Heterogeneous Networks," in *ICTTA'08 3rd IEEE International Conference on Information and Communication Technologies: From Theory to Applications*, April 2008.
[8] D. Johnson, C. Perkins, and J. Arkko, *RFC 3775 - Mobility Support in IPv6*, IETF, June 2004.
[9] P. McCann, *RFC 4260 - Mobile IPv6 Fast Handovers for 802.11 Networks*, IETF, November 2005.
[10] P. Vidales, L. Patanapongpibul, G. Mapp, and A. Hopper, "Experiences with Heterogeneous Wireless Networks - Unveiling the Challenges," in *Proceedings of Second International Working Conference on Performance Modeling and Evaluation of Heterogeneous Networks*, July 2004.
[11] P. Vidales, R. Chakravorty, and C. Policroniades, "PROTON: A Policy-based Mobile Client System for the Fourth Generation Mobile Environments," in *Proceedings of POLICY 2004: IEEE Fifth International Workshop on Policies for Distributed Systems and Networks*, June 2004.
[12] G. Mapp, D. Cottingham, F. Shaikh, P. Vidales, L. Patanapongpibul, J. Baliosian, and J. Crowcroft, "An Architectural Framework for Heterogeneous Networking," in *Proceedings of the International Conference on Wireless Information Networks and Systems (WINSYS 2006)*, August 2006, pp. 5–10.
[13] G. Mapp, F. Shaikh, J. Crowcroft, D. Cottingham, and J. Baliosian, "Y-Comm: A Global Architecture for Heterogeneous Networking (Invited Paper)," in *3rd Annual International Wireless Internet Conference (WICON)*, October 2007.
[14] J. der Merwe and I. Leslie, "Switchlets and Dynamic Virtual ATM Networks," in *Proceedings of Integrated Network Management V*, 1997.
[15] P. Stuckman, *The GSM Revolution - Mobile Packet Data Service*. Wiley, 2003.
[16] F. Shaikh, G. Mapp, and A. Lasebae, "Proactive Policy Management using TBVH Mechanism in Heterogeneous Networks," in *Proceedings of the International Conference and Exhibition on NEXT GENERATION MOBILE APPLICATIONS,SERVICES and TECHNOLOGIES (NGMAST'07)*, September 2007.

Proactive Policy Management for Heterogeneous Networks

Fatema Shaikh   Aboubaker Lasebae   Glenford Mapp

School of Computing Science

Middlesex University

London, UK

{f.shaikh, a.lasebae, <u>g.mapp</u>}@mdx.ac.uk

Abstract—**Context-awareness is a vital requirement of heterogeneous devices which allows them to predict future network conditions with sufficient accuracy. In this paper we present a proactive modelling-based approach for policy management which allows the mobile node to calculate Time Before Vertical Handover for open and closed environments. The paper explains how the knowledge of this component can improve the manner in which multi-class traffic streams are allocated to available network channels. Simulation results confirm the feasibility of the concept.**

Keywords-policy management; QoS management; vertical handover;

I. INTRODUCTION

With the full fledged deployment of fourth generation (4G) networks just around the corner, the past few years have displayed a rapid growth in multi-interfaced devices which promise simultaneous connectivity to different networks like UMTS, WLAN, WiMAX and UWB. One of the main challenges faced in heterogeneous networking is the effective delivery of multi-class traffic across diverse channels offering different levels of Quality of Service (QoS). It must be done in a way which minimises the forced termination of ongoing connections during vertical handovers. This requirement has pushed forward demands for increased context and resource awareness among heterogeneous devices. Context gathering tasks are performed by policy management mechanisms which consist of a set of rules that evaluate the circumstances under which a handover should occur. With the rapid adaptation of the heterogeneity paradigm, policy management will play an increasingly important role in improving the stability of network connections.

Heterogeneous networking has also increased complexity of network components. Components of the 4G protocol stack will exhibit more complex functionality than components of the normal OSI protocol stack due to the additional tasks they will need to support in order to achieve seamless interoperability. In this paper, we briefly introduce our proposed architectural framework similar to the OSI framework which encapsulates the key challenges of heterogeneous networking. We then introduce a novel proactive policy management mechanism that calculates a sufficiently accurate estimate of the Time Before Vertical Handover (TBVH) for both indoor and outdoor environments. Our simulation results validate the feasibility of the concept and demonstrate the flexibility of the model which can be plugged into both simulations and real-time systems with ease.

The rest of the paper is organized as follows: Section 2 describes the heterogeneous framework, Section 3 discusses policy management, Section 4 introduces the TBVH mechanism, Section 5 discusses simulation results, Section 6 demonstrates the application of TBVH, and finally, Section 7 concludes the paper.

II. THE HETEROGENEOUS FRAMEWORK

Due to the increase in the complexity and number of tasks in heterogeneous networking, successful implementation of seamless interoperability requires the introduction of a new level of intelligence to components at the network, device and application levels. Some of these new features include network component reconfigurability, policy management during vertical handovers and QoS management. There is also the pressing need for a reference model similar to the OSI model, which will clearly define the functions of all layers and provide a framework for the exchange of information between network components. We therefore propose the new heterogeneous framework [1] which consists of seven layers as follows:

- Hardware Platform Layer: This layer's function is the definition of the hardware components and technologies required to support a wireless network. It defines characteristics like electromagnetic spectrum, modulation schemes and Media Access Control (MAC) algorithms.

- Network Abstraction Layer: This layer provides a common interface for supporting the different network technologies present at the lower layer. It is responsible for controlling and maintaining networks on the MN.
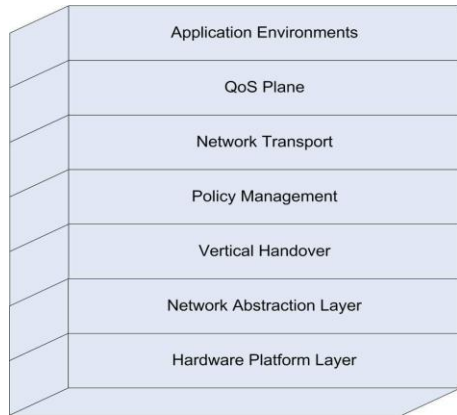

Figure 1.   The heterogeneous framework

- Vertical Handover Layer: This layer is mainly responsible for the specification of mechanisms including state engines and triggers for vertical handovers. It supports both network-controlled and client-controlled handovers.

- Policy Management Layer: This layer evaluates the circumstances when a handover should occur. It consists of a set of rules which evaluate the relevant parameters and their values to make a decision about a handover.

- Network Transport Layer: This layer examines the addressing, routing and transport issues in peripheral networks.

- Quality-of-Service (QoS) Layer: This layer supports both upward and downward QoS. Its task is to ensure that the QoS offered to applications can be maintained at an acceptable level during the lifetime of a connection.

- Application Environments Layer: This layer specifies mechanisms and routines that assist in building applications which can use all the layers of the framework.

This paper mainly explains the peripheral (client) side of the heterogeneous framework. Detailed information on the complete framework can be found in [1]. As the development of each layer involves extensive research, with each layer evolving into a separate study, the paper focuses mainly on the practical implementation of the Policy Management Layer.

III. POLICY MANAGEMENT AND HETEROGENEOUS NETWORKS

As described earlier, the main function of the Policy Management layer is the evaluation of available context information like changes in signal strength, available channel resources and the state of active TCP connections to decide when to perform a vertical handover. This layer resides in the mobile node (MN) and contributes to the client-controlled vertical handover approach [4] which we have adopted in this study. In a client-controlled approach, the MN plays an active role in deciding when to perform a vertical handover. Being directly in touch with the different networks, the MN is more aware of the latest medium access, network and transport conditions that exist at each physical interface. Thus it is in a more superior position to decide when a vertical handover should take place.

Policy management mechanisms can be classified in to two categories:

- Reactive: In this category, the MN reacts according to explicit triggers received from lower layers which inform it of changes in network conditions.

- Proactive: The MN in this category attempts to predict existing and future conditions through the evaluation of measurable network parameters like the ones mentioned earlier. Proactive mechanisms can be further classified into knowledge-based and modelling-based approaches.

Several studies in literature have proposed different types of policy management schemes. Soh et al. in [5] proposed a scheme which relied on knowledge of road topology and MN position to predict future conditions. This approach was knowledge-based and mainly relied on large volumes of data on road maps stored in prediction databases inside every BS. Hence it was not possible to predict the path for an MN that strayed away from road topology. Cottingham et al. [7] applied the knowledge-based approach in the form of data coverage maps to predict the availability of network coverage in a particular location. This scheme however was mainly for outdoor environments and did not consider indoor coverage. Ebersman et al. [6] proposed calculating time before horizontal handovers based on the change in received signal strength (RSS). However, the study failed to capture the accuracy of the MN's movement and temporary fluctuations in RSS could falsely triggers handovers.

IV. PROACTIVE POLICY MANAGEMENT USING TBVH

In this paper, we propose a novel modeling-based proactive policy management mechanism which aims to predict vertical handovers based on mathematical calculations. The mechanism targets both indoor and outdoor environments. Along with the usual parameters, our policy management mechanism depends largely on a new dynamically derived parameter call Time Before Vertical Handover (TBVH) which is derived from available information, namely, distance from BS, MN velocity, and its direction of motion.

A. TBVH – why is it important?

A significant observation made in [8] was that TCP-connection adaptation latency after a vertical handover can

actually be longer than the total handover latency. It is therefore crucial to broaden the scope and look beyond simply reducing delays and packet error rates during vertical handovers. The mere presence of another network offering increased network resources is no longer a sufficient reason for performing a vertical handover, it is vital to ensure that the new network coverage will be available long enough to allow the connection to recover from the handover and transmit for at least a certain minimum duration. Thus, in a heterogeneous environment, the knowledge of the duration for which a network channel may be available can significantly change the manner in which multi-class traffic streams are assigned to different available channels. This knowledge can also assist in minimising packet loss and latency due to handovers. For instance, consider a MN with several types of active multimedia connections. If this MN which is under the coverage of WLAN is aware that it may lose this coverage in the next minute, it can avoid allocating an interactive video stream to it. By choosing the next best available network, it can avoid the overhead associated with an upward vertical handover. Similarly, a user's PDA connected to UMTS may pick up the coverage of a WLAN for a short period when the user walks near a hotspot. The awareness that this coverage is only for a short period can help the MN in deciding not to perform a complete downward vertical handover to WLAN. Additionally, the knowledge that a high-bandwidth connection may be lost soon could actually allow the allocation of more resources to active data transfer connections and based on file size, allow the completion of transfer before the MN moves out of the current coverage. TBVH can therefore, play a crucial role in increasing the efficiency of channel allocation and resource reservation mechanisms for an MN and assist in the prevention of unnecessary vertical handovers.

One of the key requirements in the calculation of TBVH is knowledge of some aspects of network topology, in particular the knowledge of network boundaries. We propose some topological changes to networks by introducing additional specification to BSs at network boundaries, calling them Boundary Base Stations (BBS). These BBS inform the MN of imminent network boundaries. For example in outdoor scenarios, the BBS informs the MN of the vertical handover threshold and for indoor environments the dimensions of the enclosed space and the position of various exits. The BBS can also inform the MN of other networks that may be in its vicinity to which the MN is likely to perform a vertical handover but which it is yet to discover.

B. TBVH for outdoor environments

This scenario considers the case of an MN in an outdoor setting and under WLAN coverage, moving towards the boundary with velocity v (Figure 2). The networks considered are UMTS and WLAN, however, the model can be extended to other types of networks as well. For the sake of simplicity in explanation, we consider a circular coverage cell of radius R although a circular cell is not a requirement. In the above figure, the inner dotted circle of radius r represents the handover threshold where the MN is expected to perform the vertical handover. Angle x is the angle made

by the MN's movement direction with the BBS and d is the distance of the MN from the BBS. All these parameters can be determined from the location coordinates of the various network components which in turn can be recorded using various available location prediction techniques. In this scenario we
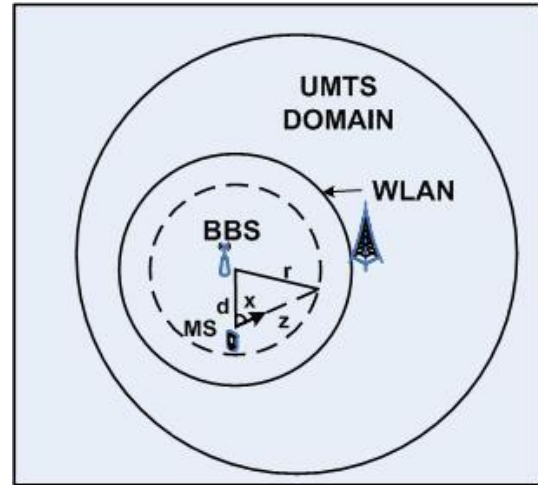

Figure 2.　　　MN under BBS coverage

need to calculate z which is the point on the threshold circle where the MN is expected to vertically handover. As

$$r^2 = d^2 + z^2 - 2dz\cos x \quad\quad)$$

Due to geometric considerations, we only consider one root of the quadratic equation as the formula below (2) will always give positive solution. So the value of z is

$$z = (d\cos x) + \sqrt{r^2 - d^2\sin^2 x} \quad (2)$$

Thus the estimated TBVH for this scenario is:

$$\frac{(d\cos x) + \sqrt{r^2 - d^2\sin^2 x}}{v} \quad (3)$$

Different cases in TBVH arise based on the movement of MN either towards or away from a network boundary. These have been discussed in detail in [2]. In all these cases, the formula for TBVH calculation remains essentially the same.

C. TBVH for indoor environments

This scenario considers movement of the MN under indoor WLAN coverage. TBVH can be predicted with greater accuracy for indoor environments due to the precise definition of coverage due to availability of accurate topological information. Indoor scenarios also facilitate ease in testing. During the connection setup phase the MN receives a beacon from the BBS which contains indoor topological information such as room dimensions and points of exit.

Unlike outdoor coverage, TBVH calculation here cannot depend only on handover threshold for several reasons.

Firstly, an MN moving under a small coverage like WLAN is likely to exhibit frequent random movements, characteristic of pedestrian behaviour, causing frequent change in direction. For example the MN moving towards an exit may suddenly undergo a change in direction and move in the exact opposite direction. Secondly, as shown in
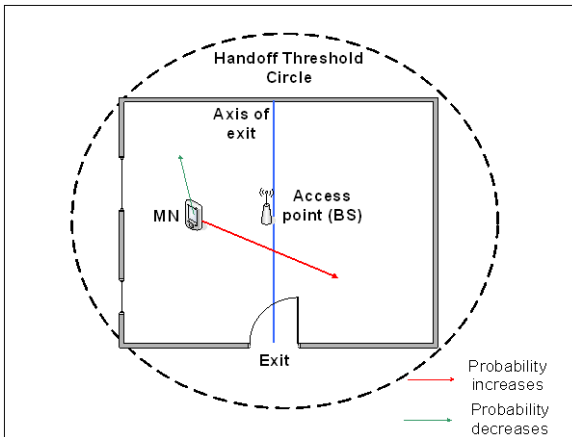
Figure 3.   MN in indoor environment

Figure 3, the MN may appear to move closer to the threshold circle but in the direction of a wall instead of an exit. In this case TBVH value alone is not a sufficient indicator of handover because although the value reduces as the MN approaches the boundary, in reality it cannot leave the WLAN coverage as it will be stopped by the wall. It is thus important to develop a mechanism that will take into consideration these random movements of the MN. To address this issue we propose assigning a weight W1 to TBVH. W1 is the cosine of the MN direction calculated with respect to a particular point of exit. The higher the value of W1, the more likely it is to pass through the exit. TBVH mechanism for indoor environments must also accommodate the presence of multiple exits points. In this case, TBVH and W1are calculated separately for each point of exit. Thus for indoor scenarios, the final probability of when the MN will perform a vertical handover is indicated by both TBVH and W1. Once the MN moves out of the enclosed area, TBVH is calculated as per equation (3). The next section demonstrates the simulations of TBVH mechanisms for both indoor and outdoor environments.

V. SIMULATION AND RESULTS

Based on the ideas proposed earlier in the paper, the experimental proactive TBVH simulation model was developed in OPNET Modeler. The TBVH module's block diagram is shown below. Input parameters employed in TBVH calculation were mainly the location co-ordinates for for the MN and BBS. Figure 5 represents the scenario where the MN moves in open space. Figure 6 displays the graphed results for TBVH calculated for the moving mobile node.

Results agreed with intuition and instantaneous TBVH values closely coincided with the location and behaviour of the MN along its trajectory as shown in table 1. When the
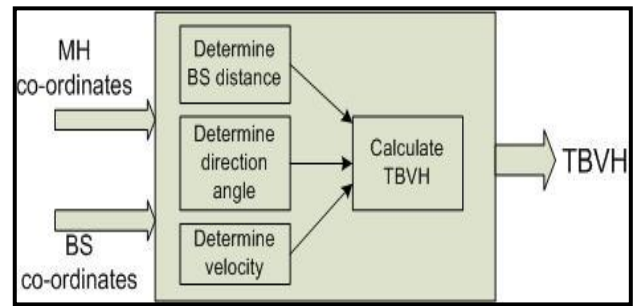
Figure 4.   TBVH node module

TABLE I.   TBVH POINTS IN THE MN TRAJECTORY

MN location	TBVH before direction change	TBVH after direction change
Point 1	35.42	176.50
Point 2	37.10	166.97
Point 3	77.10	158.97
Point 4	97.85	150.24
Point 5	63.40	193.13
Point 6	88.07	185.66
Point 7	122.30	
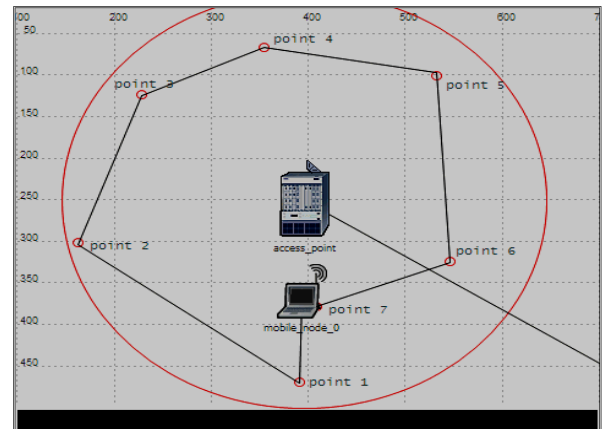
Figure 5.   MN movement in open environment

MN moved towards another BS in the WLAN cell but not the network boundary, TBVH represented time before handover to next WLAN cell instead of a vertical handover. In the TBVH graph, each physical point is represented twice. This corresponds to the value of TBVH before and after the MN changes its direction at a particular position. For

Figure 6.   TBVH graph for open environment

Figure 7.   MN movement in a closed enviornment

Figure 8.   TBVH weight correstponding to MN direction

example, point1 lower down in the graph represents the TBVH value when the MN's direction is downward while the point 1 higher up represents TB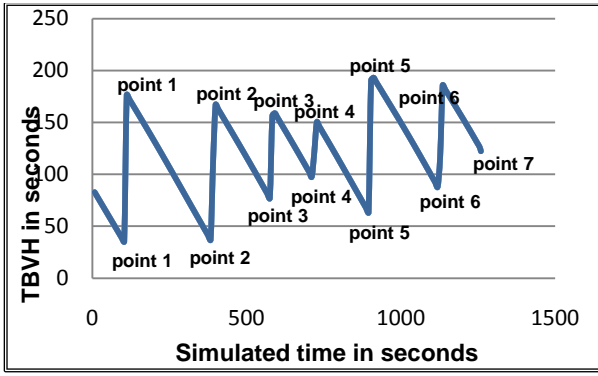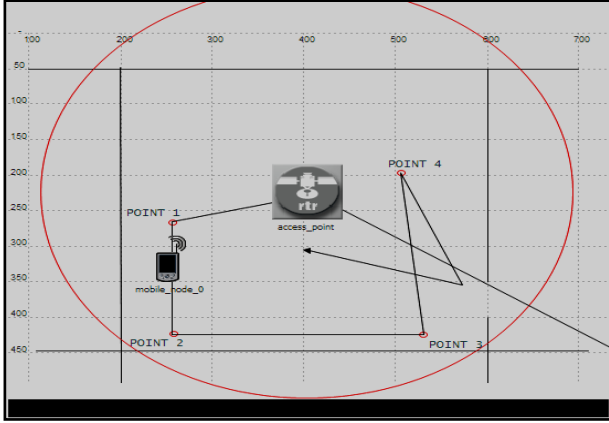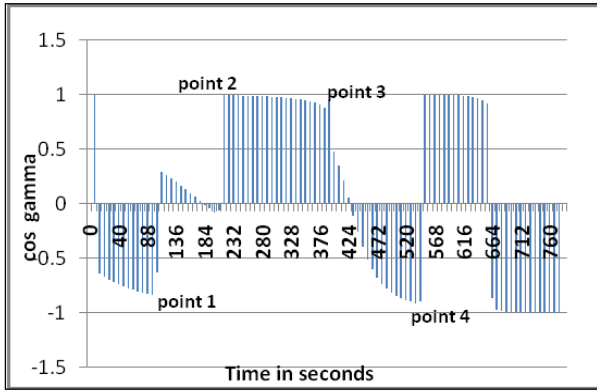VH calculated at the same position but when the MN changes direction and moves upwards. Figures 7 and 8 represent the network model and TBVH weight graph for indoor scenario respectively. We now have the graph for TBVH weight which is mainly the cosine of the direction angle made by the MN's direction with respect to the point of exit. This graph of cosine values captures closely the MN's direction,

displaying values greater than 0 each time the MN approaches towards the exit. For instance, considering the movement between points 3 and 4, it can be observed that the weight value begins to decrease as the MN moves away from the exit but it remains above zero while the MN roams in the vicinity of the exit and goes negative only when the MN moves further away, approaching minus 1 eventually at point 4. A similar behaviour can be observed between points 1 and 2. Thus experimental results clearly demonstrate the successful implementation of our proposed mechanim for calculating the time before vertical handover for an MN.

VI.   TBVH FOR DOWNWARD QOS MANAGEMENT

In this section, we briefly demonstrate how TBVH can be applied for the management of downward QoS. More detailed explanation is found in [3]. For a multi-interfaced MN, the simultaneous presence of different network channels offering different levels of QoS causes an increase in the complexity of multi-class traffic management issues such as resources management, traffic scheduling and flow control. Downward QoS management can be defined as the task of mapping application stream requirements down to the appropriate available network channel. Downward QoS at the MN requires answers to several key issues including:

- The QoS requirements of application streams.

- ☐Most suitable networks among currently available ones for allocating a particular call.

- The current and likely future conditions of these networks.

- How long are these networks likely to remain available.

In a multi-interfaced client, the context parameters for each physical interface are stored in a two-dimensional matrix called the Network Descriptor Matrix (NDM).

$$
\begin{array}{llllll}
NWid_1 & status_1 & avbw_1 & RSS_1 & TBVH_1 & RTT_1 \\
NWid_2 & status_2 & avbw_2 & RSS_2 & TBVH_2 & RTT_2 \\
NWid_3 & status_3 & avbw_3 & RSS_3 & TBVH_3 & RTT_3 \\
NWid_4 & status_4 & avbw_4 & RSS_4 & TBVH_4 & RTT_4 \\
NWid_5 & status_5 & avbw_5 & RSS_5 & TBVH_5 & RTT_5
\end{array}
$$

Parameters of each row in the NDM represent network ID (NWid), network status (status) with on/off values, available bandwidth (avbw), received signal strength (RSS), time before vertical handover (TBVH), and round trip time (RTT) between BS and MN respectively.

Application traffic streams can be of largely varying behavioural characteristics e.g. interactive video, streaming video, audio and data. Improved context awareness of available networks is necessary for an MN before it can bundle these traffic streams more efficiently over them. In such situations, the TBVH parameter can play an important role in deciding the choice of a network and the amount of

resources allocated to a particular traffic stream. Figure 9 shows the choice of a network for video/ftp traffic based on conditions mentioned above. As these applications are more likely have high resource requirements, WLAN is designated as the first network choice for these types of applications. The algorithm checks if the MN speed is less than a specific

threshold required for WLAN and then checks for other network parameter conditions. If both conditions are satisfied, the stream is allocated to WLAN else resource availability is checked for UMTS. If both resource availability checks fail, the stream request is queued and an urgency value is incremented. The urgency value is assigned
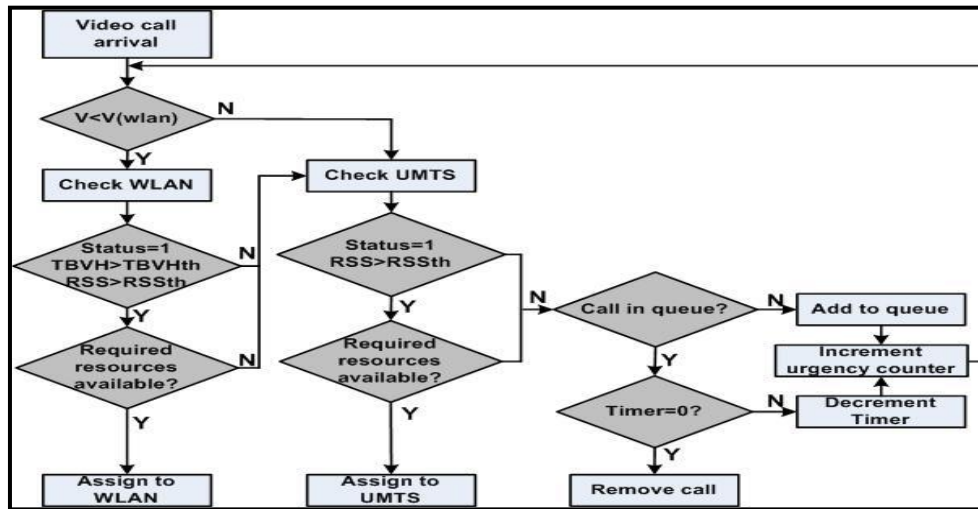

Figure 9.   Application of TBVH in choice of networks

in order to avoid the starvation of low priority traffic. This value increases the longer the application request remains in the queue. In future, the application leaves the queue when a channel becomes available or when its waiting timer expires. The amount of resources e.g. available bandwidth allocated to a stream is decided with the help of the Weighted Resource Allocation (WRA) equation

$$(TBVH \, x \, W1) \; \square\square \, (UV \, x \, W2) \; \square\square \, (V \, x \, W3) \qquad (4)$$

where UV is the urgency value for the stream, V the velocity of MN and (W1+ W2 + W3 = 1).

VII.   CONCLUSION AND FUTURE WORK

In this paper we dealt with one of the main concerns with heterogeneous networking – QoS issues in multi-class traffic management. We highlighted the importance of policy management mechanisms in improving the context awareness of mobile nodes and proposed a client-based proactive policy management scheme for the prediction of the time before vertical handovers in mobile nodes. This scheme was developed for both open and closed environments and successfully captured the random movement behaviour of devices. The proposed mechanism was practically implemented in OPNET Modeler and results demostrated that the scheme worked correctly for both environments. The The paper also explained how the knowledge of TBVH helped to improve the management of multi-class traffic streams in a heterogeneous client. Future work in this area will include the performance study of the TBVH model after its real-lfe implementation in a proposed extended test-bed [9].

VIII.   REFERENCES

[1] G. Mapp, D. Cottingham, F. Shaikh, P. Vidales, L. Patanapongpibul, J. Baliosian and J. Crowcroft, "An architectural framework for heterogeneous networking." Proc. WINSYS'06. August 2006, pp. 5-10

[2] F. Shaikh, A. Lasebae , G. Mapp, "Client-based SBM layer for predictive management of traffic flows in heterogeneous networks." Proc. IEEE ICTTA'06, April 2006.

[3] F. Shaikh, G.Mapp, A.Lasebae, "Proactive policy management using TBVH mechanism in heterogeneous networks." Proc. NGMAST'07, September 2007.

[4] L. Patanapongpibul, G. Mapp, A. Hopper, "An end-system approach to mobility management for 4G networks and its application to thin-client computing." ACM SIGMOBILE. July 2006.

[5] W-S Soh and H.S. Kim, "Dynamic bandwidth reservation in cellular networks using road topology based mobility predictions", Proc. IEEE INFOCOM'23, Honkong, March 2004.

[6] H.G. Ebersman, O.K. Tonguz, "Handoff ordering using signal prediction priority queuing in personal communication systems," IEEE Trans. Veh. Techno., Vol. 48. No. 1 1999, pp. 22-35.

[7] D Cottingham and J Davies. "A vision for wireless access on the road network." 3rd International Workshop on Intelligent Transportation, March 2007.

[8] D Cottingham and P Vidales. "Is latency the real enemy in next generation networks." Proc. First International Workshop on Convergence of Hetergeneous Wireless Networks, July 2005.

[9] J. Crowcroft, D. Cottingham, G. Mapp and F. Shaikh. "Y-Comm: A global architechture for heterogeneous networking." Proc. WICON'07, October 2007.

Proactive Policy Management using TBVH Mechanism in Heterogeneous Networks

Fatema Shaikh   Glenford Mapp   Aboubaker Lasebae

Middlesex University, London. UK.
{f.shaikh, g.mapp, a.lasebae}@mdx.ac.uk

Abstract

In order to achieve seamless interoperability in heterogeneous networking, it is vital to improve the context-awareness of the mobile node (MN) so that it is able to predict future network conditions with sufficient accuracy. In this paper, we introduce a predictive mathematical model for calculating the estimated Time Before Vertical Handover (TBVH) component from available network parameters. The model is practically implemented in OPNET and our simulation results confirm the validity of the concept. We then demonstrate how the knowledge of TBVH along with other network parameters can be applied by downward Quality of Service management policies which bundle multi-class traffic streams on to available network channels based on application QoS, device mobility patterns and prevailing channel conditions.

1. Introduction

In 4G networks, the introduction of heterogeneity among existing wireless standards adds a new level of complexity to traffic and resource management issues, particularly during vertical handovers. Multi-interfaced heterogeneous clients supporting networks such as 3G, WLAN, WiMAX and UWB are expected to roam freely among these different networks without experiencing disruption to services, and benefit from the best available location-based network facilities. These devices are offered the choice of transmission on multiple but changing wireless channels exhibiting varying levels of QoS. The aim is to effectively deliver multi-class traffic across these diverse channels, by minimising the forced termination of ongoing connections during vertical handovers.

The advent of heterogeneous networking has increased the complexity of network components. In order to achieve seamless interoperability, components of the 4G protocol stack will exhibit more complex functionality than components of the normal OSI protocol stack due to the additional tasks they will need to support. In this paper, we briefly introduce our proposed architectural framework similar to the OSI model which encapsulates the key challenges of heterogeneous networking. The paper then proposes a

proactive analytical model for calculating a sufficiently accurate estimate of the Time Before Vertical Handover (TBVH) for a roaming MN. Simulation results confirm the feasibility of the concept and demonstrate the flexibility of the model which can be plugged into both simulations and real-time systems with ease. Next the paper demonstrates how TBVH is utilised by the proposed Stream Bundle Management Layer (SBM) which is an intelligent, client-based layer for downward QoS management. It consists of a set of policies that combine knowledge of TBVH along with the knowledge of networks' conditions for bundling multi-class traffic streams onto the most appropriate wireless channel.

The rest of the paper is organised as follows: Section 2 introduces the heterogeneous framework, Section 3 explains policy management and the TBVH model with simulation results. Section 4 introduces the SBM layer and demonstrates the application of TBVH; and finally the paper concludes with Section 5 with a brief discussion on future work.

2. Heterogeneous Networking Framework

The development of 4G heterogeneous networks introduced the attractive paradigm of "seamless QoS provisioning". However, it also brought along with it a plethora of new challenges at the network, device and application levels. What is needed is a new framework to encapsulate mechanisms that address these challenges in heterogeneous environments. Some of the key requirements of this framework include reconfigurability of network components, QoS management, and policy management for vertical handovers.

In order to address the issues mentioned above, we proposed the architectural framework for heterogeneous networking [2]. It acts as a reference model similar to the OSI model and clearly defines the functions of all layers and provides a framework for exchanging data between network applications. The seven layers of the framework are as follows:

- **Hardware Platform Layer:** This layer defines the hardware components and technologies required to support a wireless network. It defines characteristics like electromagnetic spectrum, modulation schemes and Media Access Control (MAC) algorithms.
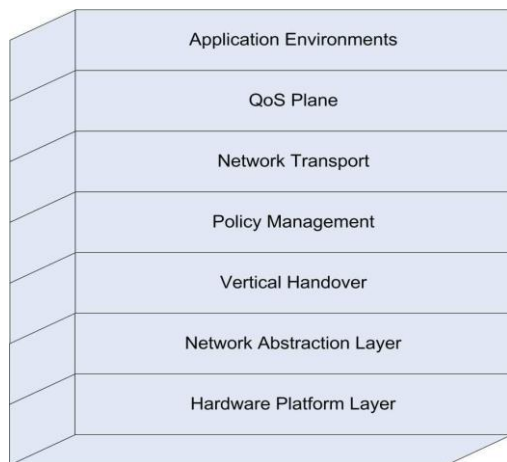
Figure 1. The heterogeneous framework

- **Network Abstraction Layer:** This layer provides a common interface for supporting the different network technologies. It is responsible for controlling and maintaining networks on the MN.

- **Vertical Handover Layer:** The layer supports both network-controlled and client-controlled handovers. It is mainly responsible for the specification of mechanisms including state engines and triggers for vertical handovers.

- **Policy Management Layer:** The function of this layer is to evaluate the circumstances when a handover should occur. It is implemented through the definition of a set of rules with regard to all the relevant parameters and their values which are evaluated with respect to handover.

- **Network Transport Layer:** This layer examines the addressing, routing and transport issues in peripheral networks.

- **Quality-of-Service (QoS) Layer:** This layer is designed to support both upward and downward QoS in a heterogeneous device. It ensures that the QoS offered to applications can be maintained at an acceptable level during the lifetime of a connection.

- **Application Environments Layer:** The function of this layer is to specify mechanisms and routines that assist in building applications which can use all the layers of the framework.

3. Policy Management and TBVH

A successful implementation of the heterogeneous framework involves a lot of intensive research, and each layer's functioning could evolve into a separate research study. In this paper we focus on the Policy management and QoS layers and propose mechanisms for their practical implementation.

3.1. Client-controlled approach for vertical handovers

Vertical handovers can be classified into two main categories - network-controlled and client-controlled. Before the advent of heterogeneous networking handover management was mainly network-controlled where the base station (BS) maintained up-to-date context information for the MN and decided when and how it should perform a handover. However, this approach is not scalable in heterogeneous networking where a MN is simultaneously connected to several different BSs. Furthermore, in a loosely coupled heterogeneous scenario where networks may belong to different service providers, a network-controlled approach will require the detailed sharing of network and customer information, something service providers may be unwilling to share.

In a client-controlled approach, the MN is assigned a more active role in handover management. As the MN is directly connected to different networks, it possesses up-to-date knowledge of the medium access, network and transport conditions for each active network interface. It is thus in a superior position to decide when a vertical handover should take place. In this study we adopt the client-based approach as explored in the Cambridge wireless test bed [3] and the IEEE 802.21 working group [4].

3.2. Time Before Vertical Handover (TBVH)

As described earlier in the reference model, the main function of the Policy Management layer is to decide when to execute a vertical handover based on several key parameters like changes in signal strength, available network bandwidth, state of active transport connections on the MN and the time the MN has before it performs a vertical handover. While most of these parameters can be sensed directly from network interfaces, TBVH is derived from other available information, namely, distance from BS, MN velocity, and its direction of motion.

A few attempts were made earlier for determining the time before handover. The study in [5] proposed a history-based mobility prediction technique based on positioning knowledge and road topology. However, the approach mainly relied on large volumes of data on road maps stored in prediction databases inside every BS so it was not possible to predict the path for an MN that strayed away from road topology. Ebersman et al. [6] proposed calculating time before horizontal handovers based on the change in received signal strength (RSS). However, by relying only on change in RSS, the study failed to capture the accuracy of the MN's movement and temporary fluctuations in RSS could falsely triggers handovers.

In a heterogeneous environment, the knowledge of the time for which an MN will have access to a particular network channel offers it the opportunity to minimise packet loss and latency due to handover mechanisms. For instance, an MN that is aware that it may loose WLAN coverage in the next minute will avoid allocating an interactive video stream to it. By choosing the next best available network, it avoids the overhead associated with an upward vertical handover. Similarly, a user's PDA connected to UMTS may pick up the coverage of a WLAN for a short period when the user walks near a hotspot. The awareness that this coverage is only for a short period can help the MN in deciding not to perform a complete downward vertical handover to WLAN. TBVH therefore, plays a key role in increasing the efficiency of channel allocation and resource reservation mechanisms for an MN and assists in the prevention of unnecessary vertical handovers.

A key requirement in the calculation of TBVH is the knowledge of some aspects of network topology, in particular the knowledge of network boundaries. To address this issue we proposed some topological changes to networks by introducing additional specification to BSs at network boundaries, calling them Boundary Base Stations (BBS). An MN approaching a BBS becomes aware of the fact that it may have to perform a vertical handover. The main function of the BBS is to inform the MN of its location parameters and other information it may need for the vertical handover, e.g. a record of other networks that may be in its vicinity to which the MN is likely to perform a vertical handover but which it is yet to discover.

3.3. Determination of TBVH

Here we explain the mathematical derivation of TBVH and the different scenarios that arise based on the MN's location in the network and its direction of motion. For the sake of simplicity, only the UMTS-WLAN network combination is considered, although the model can be easily applied to other network combinations. The explanation here is restricted to upward vertical handovers, although the concept can also be applied in a similar manner to avoid downward vertical handovers. TBVH does not need to be calculated when an MN moves among normal BSs. The two scenarios for TBVH calculation are as follows:

Case I: Outward movement of MN in BBS towards boundary

This scenario considers the case of an MN that is roaming under the coverage of a BBS and is moving towards the boundary with velocity v (Fig. 2). Here we consider a circular cell of radius R. The inner dotted concentric circle represents the handover threshold of radius r which is the distance from the BSS where a MN is expected to perform a handover. Angle x is the angle
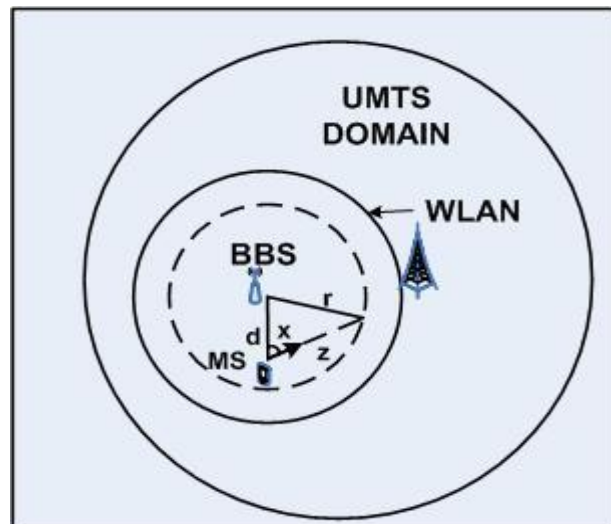

Figure 2. Movement of MN in BBS

made by the line joining the BBS and the direction of MN movement, and d is the distance of the MN from the BBS which can be determined either from the received signal strength (RSS) or from the location co-ordinates of the MN and BBS. In order to determine the TBVH in this scenario we need to calculate the distance z which is the point on the threshold circle where the MN is expected to vertically handover. As

$$r^2 = d^2 + z^2 - 2dz \cos x \qquad (1)$$

Due to geometric considerations, we only consider one root of the quadratic equation as the formula below (2) will always give positive solution. So the value of z is

$$z = (d \cos x) + \sqrt{r^2 - d^2 \sin^2 x} \qquad (2)$$

Thus the estimated TBVH for this scenario is:

$$TBVH = \frac{(d \cos x) + \sqrt{r^2 - d^2 \sin^2 x}}{v} \qquad (3)$$

As WLANs may have specific points of exits such as doors in a building, the prediction accuracy of TBVH can be improved if the co-ordinates of these exit points are stored in the BBS and passed on the MN when required.

CASE II: Movement of MN from normal BS to BBS

In this scenario (Fig. 3), a MN (point C) is under the coverage of a normal BS (NBS) (point A), but is moving towards a BBS (point B) with velocity v. The goal here is to improve the prediction capability of the model by making it able to predict the TBVH of MN while it still moves in the coverage of A. In this case, the concept of threshold distance TD [8] is used in the normal BS. This is a distance smaller than the cell's radius and defines a smaller   concentric   circle
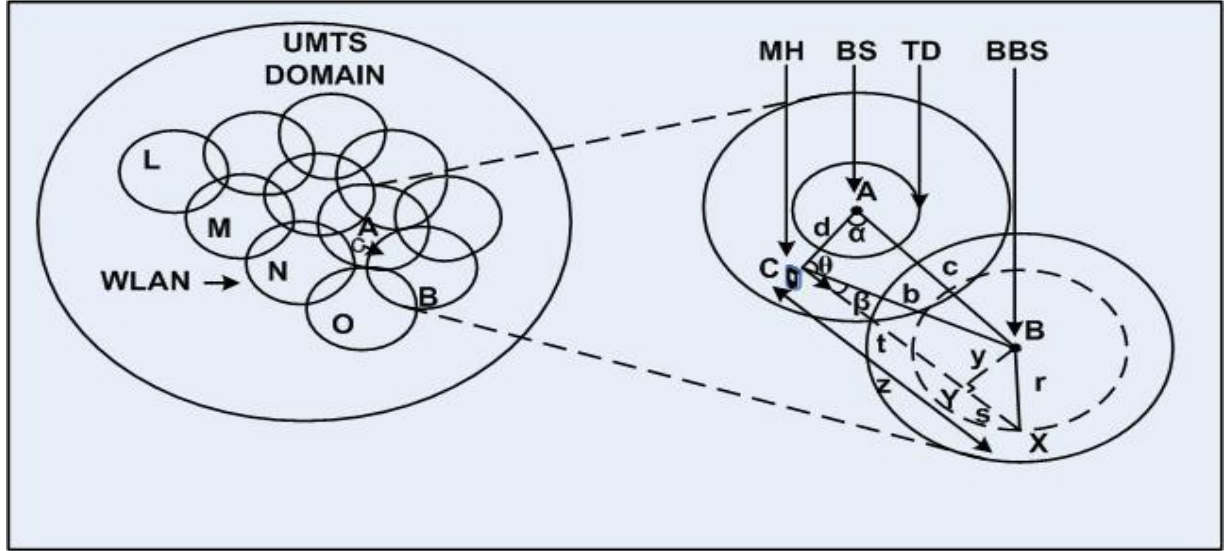
Figure 3. Movement of MN towards BBS

located within the cell. A MN moving inside the TD circle is more likely to change direction, however on moving out of this circle, it is less likely to undergo a sudden change in its direction, thus enabling a correct prediction of the next cell the MN is moving towards. As the MN is too far from the BBS to get a reasonably accurate value of b, it is necessary first to find this distance and the angle β in order to calculate distance z. In Fig. 3,

$$c^2 = d^2 + b^2 - 2db\cos\theta \qquad (4)$$

Therefore,

$$\theta = \cos^{-1}\left(\left(b^2 + d^2 - c^2\right)/2db\right) \qquad (5)$$

Depending on which side of line AB point X lies,

$$\text{angle } \beta = |x - \theta| \qquad (6)$$

Considering triangle BYC, we have

$$t = b\cos\beta \qquad (7)$$
$$y = b\sin\beta \qquad (8)$$

Therefore, in triangle BYX,

$$s = \sqrt{r^2 - b^2 \sin^2\beta} \qquad (9)$$

As

$$z = t + s,$$

From (7), (8) and (9) we have

$$z = b\cos\beta + \sqrt{r^2 - b^2 \sin^2\beta} \qquad (10)$$

Thus the TBVH component for this scenario is,

$$TBVH = \frac{b\cos\beta + \sqrt{r^2 - b^2 \sin^2\beta}}{v} \qquad (11)$$

This is similar to the equation obtained in (3).

It is important to note that while TBVH may vary due to random movements of the MN, the calculated TBVH when considered in the weighted resource allocation mechanism (discussed in next section) along with other parameters will greatly assist in improving the context-awareness of the MN during an imminent vertical handover.

3.4. TBVH Simulation and Results

Based on the ideas proposed earlier in the paper, the proactive TBVH simulation model was developed in OPNET Modeler. The TBVH module's block diagram is shown below.
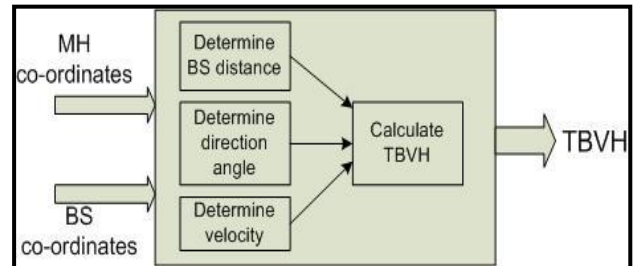
Figure 4. TBVH node module

Input parameters employed in TBVH calculation were mainly the location co-ordinates for the MN and BBS available using GPS or network-based positioning techniques. These input parameters were used to calculate the MN-BBS distance, angle of direction and MN velocity.

The TBVH module was embedded in the node model of a wireless LAN workstation. The scenario simulated (Fig. 6) was a WLAN of cell radius 300m with a handoff threshold of 290m. Positioning delay was set to 3 seconds and positioning interval to 5 seconds. The

Figure 5. TBVH choice flow chart

mobile node was assigned a random trajectory and a speed of 2m/s.

Based on the values of changing RSS for current BS (CBS) and next BS (NXBS), the different cases for TBVH calculation were applied as shown in Fig. 5.

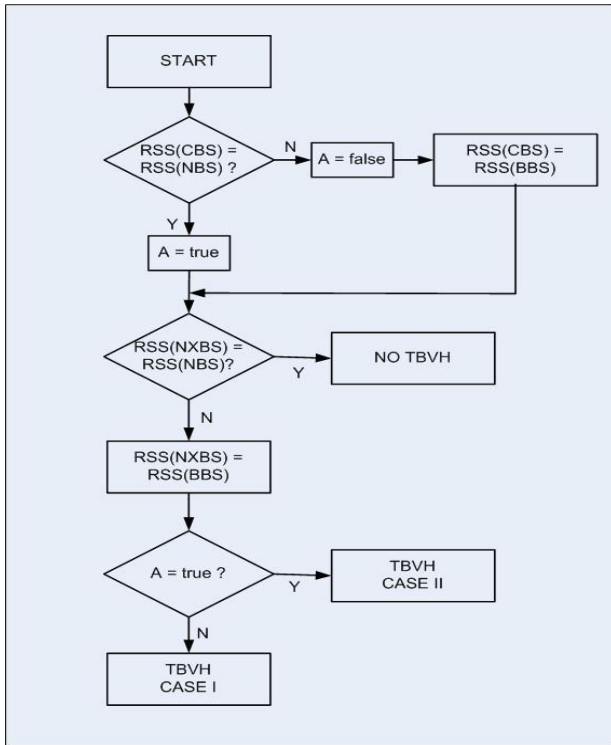Fig. 7 depicts the simulated results for TBVH calculated for the moving mobile node. The horizontal axis represented the simulated time in minutes while the vertical axis represented TBVH in seconds. Results agreed with intuition and instantaneous TBVH values closely coincided with the location and behaviour of the MN along its trajectory as shown in table 1. When the MN moved towards another BS in the WLAN cell but not the network boundary, TBVH represented time before handover to next WLAN cell instead of a vertical handover. For example points 6 and 7 in table 1 showed the time before handover to the next WLAN cell. This was decided based on the change in RSS values of current BS and neighbouring BSs.

4. Application of TBVH for Downward QoS Management

After explaining how TBVH can be calculated at the policy management level, we demonstrate how it can be applied to support downward QoS in heterogeneous devices using the Stream Bundle Management (SBM) Layer. This section mainly highlights the role of TBVH in multi-class traffic management. Readers can refer to [1] for a detailed explanation on the SBM layer.
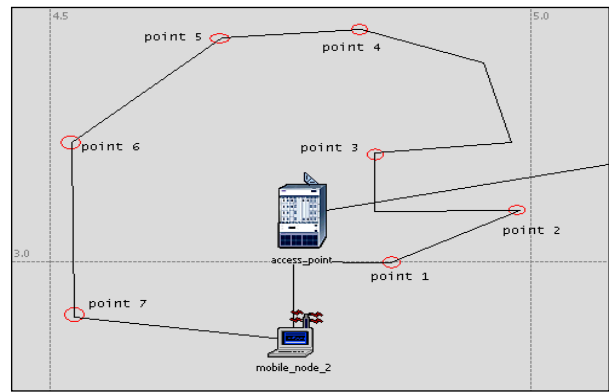
Figure 6. Simulated path of MN in WLAN BBS
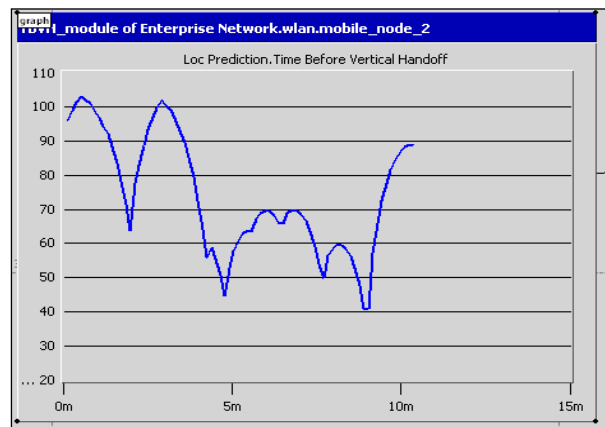
Figure 7. Generated TBVH for MN

Table 1. TBVH in MN trajectory

POINT	TBVH (SECONDS)
Point 1	96.026
Point 2	65.81
Point 3	95.61
Point 4	63.74
Point 5	65.81
Point 6	51.32
Point 7	40.97

The presence of multiple network channels at the MN in heterogeneous networks, offering different levels of QoS increases the complexity of multi-class traffic management issues such as resource management, traffic scheduling and flow control. Downward QoS management at the MN requires answers to several key issues including:

- The QoS requirements of the application streams.
- Most suitable networks among currently available ones for allocating a particular call.
- The current and likely future conditions of these networks.
- How long are these networks likely to remain available.

Thus it is crucial for the MN to improve its knowledge of available networks' contexts before bundling traffic streams over them.

The SBM layer is a specialised proactive layer residing in the QoS Plane in a heterogeneous MN and is responsible for handling downward QoS. It consists of a set of intelligent policies for allocation and scheduling of multi-class traffic streams onto different available channels based on application priority and behavioural patterns, device mobility patterns as well as network and transport conditions of available wireless channels. Network context information for each network interface is stored in the two-dimensional matrix called the Network Descriptor Matrix (NDM). Parameters of each row in the NDM represent network ID (NWid), network status (status) with on/off values, available bandwidth (avbw), received signal strength (RSS), time before vertical handover (TBVH), and round trip time (RTT) between BS and MN respectively.

$$NWid_1 \quad status_1 \quad avbw_1 \quad RSS_1 \quad TBVH_1 \quad RTT_1$$
$$NWid_2 \quad status_2 \quad avbw_2 \quad RSS_2 \quad TBVH_2 \quad RTT_2$$
$$NWid_3 \quad status_3 \quad avbw_3 \quad RSS_3 \quad TBVH_3 \quad RTT_3$$
$$NWid_4 \quad status_4 \quad avbw_4 \quad RSS_4 \quad TBVH_4 \quad RTT_4$$
$$NWid_5 \quad status_5 \quad avbw_5 \quad RSS_5 \quad TBVH_5 \quad RTT_5$$

Traffic streams considered are interactive video, one-way streaming video, audio and data. The layer maintains a prioritised list of compatible networks for each traffic type. In the absence of a suitable channel the application is assigned to a dynamic priority waiting queue. The application leaves the queue when a channel becomes available or when its waiting timer expires. In order to avoid starvation of low priority traffic, application streams are assigned an urgency value which increases the longer the application request remains in the waiting queue.

The SBM layer's Resource Allocation and Traffic Scheduling mechanism (RATS) continuously monitors the amount of resources allocated to a traffic stream. Here, the TBVH parameter plays an important role in deciding the choice of a network and amount of resources allocated to the traffic stream. Fig. 8 demonstrates the choice of network for video/ file transfer traffic. Possible high resource requirements for these types of traffic make WLAN the first network choice. The algorithm checks if the MN speed is less than a specific threshold required for WLAN and then checks for other network parameter conditions. If both conditions are satisfied, the stream is allocated to WLAN else resource availability is checked for UMTS. If both resource availability checks fail, the stream request is queued and urgency value is incremented.

The amount of resources (available bandwidth) allocated to a stream is decided with the help of the Weighted Resource Allocation (WRA) equation.

$$(TBVH \times W1) + (UV \times W2) + (V \times W3)$$

where UV is the urgency value for the stream, V the velocity of MN and (W1+ W2 + W3 = 1). For example for a file transfer stream, the WRA mechanism can allocate more bandwidth to the flow and allow it to complete transfer before the MN performs the upward vertical handover. The concept of WRA is similar to the reservation ordering scheme in [7]. However, instead of assigning a single priority to all streams on a MN, our WRA mechanism adapts a more refined approach and bandwidth allocation is considered on a per-flow basis.

5. Conclusion and Future Work

QoS management in heterogeneous networking requires the development of improved mechanisms capable of adapting to continuous changes in network conditions. In this paper we proposed the predictive mathematical model for calculating the estimated TBVH. Different cases in TBVH calculation that could arise based on the MN's trajectory were discussed. The model was practically implemented in OPNET and simulation results demonstrated the validity of the concept. We then addressed the problem of QoS management and introduced the Stream Bundle Management Layer which consisted of a set of policies that applied knowledge of TBVH along with other channel conditions for bundling multi-class traffic onto different available channels. Future work in this area will include the performance study of the TBVH model and SBM layer after their implementation in a proposed extended test-bed.

6. Acknowledgements

7. References

[1] F. Shaikh, A. Lasebae and G. Mapp, "Client-based SBM Layer for predictive management of traffic flows in heterogeneous networks." *Proc. IEEE ICTTA'06*, April 2006.

[2] G. Mapp, D. Cottingham, F. Shaikh, P. Vidales, L. Patanapongpibul, J. Baliosian and J. Crowcroft, "An Architectural Framework for Heterogeneous Networking." *Proc. WINSYS'06*. August 2006, pp. 5-10.

[3] P.Vidales, L. Patanapongpibul, G. Mapp, A. Hopper, "Experiences with Heterogeneous Wireless Networks, Unveiling the Challenges", *Proc. HET-NET'04*, July 2004, pp. WP10/1-WP10/8.
[4] A.Dutta, S.Das, D. Famolari, Y.Ohba, K.Taniuchi, T.Kodama and H.Schulzrinne, "Seamless Handover across Heterogeneous Networks – An IEEE 802.21 Centric Approach", *Proc. IEEE WPMC*, Sept 2005.

[5] W-S Soh and H.S. Kim, "Dynamic Bandwidth Reservation in Cellular Networks Using Road Topology Based Mobility Predictions", *Proc. IEEE INFOCOM'23*, Honkong, March 2004.

[6] H.G. Ebersman, O.K. Tonguz, "Handoff ordering using signal prediction priority queuing in personal communication systems," IEEE Trans. Veh. Techno., Vol. 48. No. 1 1999, pp. 22-35.

[7] F. Hu and N.K. Sharma, "Multimedia Call Admission Control in Mobile Networks: A Dynamic Pool Reservation Approach," *Computer Networks*, vol. 43, pp. 263-288, 2003.
[8] M-H Chiu, M. Bassiouni, "Predictive Schemes for Handoff Prioritization in Cellular Networks Based on Mobile Positioning", *IEEE J. Select Areas Commun.*, vol. 18, no. 3, Mar. 2000, pp. 510-522.
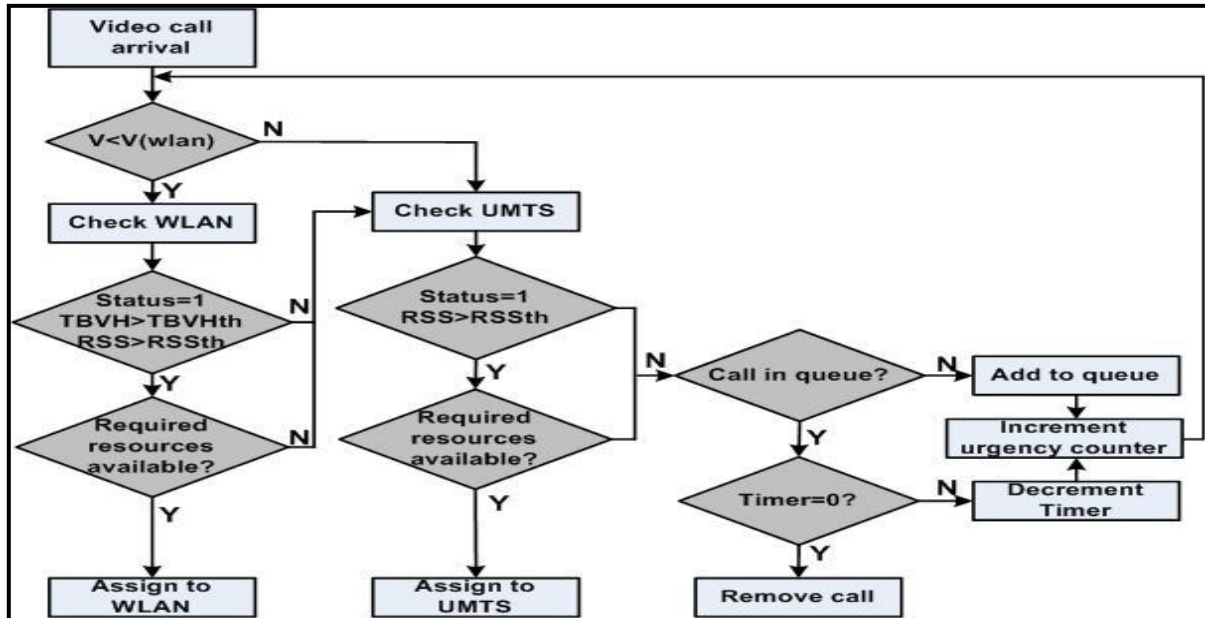
Figure 8. Network choice for video/data calls

BIBLIOGRAPHY

AGUIAR, R.L., SARGENTO, S., BANCHS, A., BERNARDO, C.J., CALDERON, M., SOTO, I., LIEBSCH, M., MELIA, T. AND PACYNA, P. Scalable QoS-aware mobility for future mobile operators. *IEEE Communications Magazine*, Vol. 44, 95-102. 2006.

AHMED, T., KYAMAKYA, K. AND AND LUDWIG, M. A context-aware vertical handover decision algorithm for multimode mobile terminals and its performance. *Proceedings of IEEE/ACM Euro American Conference on Telematics and Information Systems* 19-28. February 2006.

AKYILDIZ, I.F. AND WANG, W. 2002. A dynamic location management scheme for next-generation multitier PCS systems. *IEEE Transactions on Wireless Communications,* Vol. 1, Issue 1, 178-189. 2002.

ALGEIER, V., DEMISSIE, B., KOCH, W. AND THOMA, R. 2008. State Space Initiation for Blind Mobile Terminal Position Tracking. *EURASIP Journal on Advances in Signal Processing.* 2008, Hindawi Publishing Corporation.

ARSLAN, H., CHEN, Z.N. AND BENEDETTO, M. Ultra Wideband Wireless Communication. WileyBlackwell, New Jersey, USA. 2006.

DUTTA, A., DAS S., FAMOLARI D., OHBA Y., TANIUCHI K., SCHULZRINNE H. Seamless Handover across Heterogeneous Networks - An IEEE 802.21 Centric Approach. *Proceedings of the I.E.E.E. WPMC Conference*, September 2005.

AUST, S., PROETEL, D., FIKOURAS, N.A., PAMPU, C. AND GORG, C. Policy based Mobile IP handoff decision (POLIMAND) using generic link layer information. *Proceedings of 5th IEEE International Conference on Mobile and Wireless Communication Networks*, October 2003,

BALASUBRAMANIAM, S. AND INDULSKA, J. Vertical handover supporting pervasive computing in future wireless networks. *Computer Communications* Vol. 27, Issue 8, 708-719. 2004.

BANERJEE, N., WU, W., BASU, K. AND DAS, S.K. Analysis of SIP-based mobility management in 4G wireless networks. *Computer Communications Vol.* 27, 697-707. 2004.

BAHL, P. AND PADMANABHAN, V.N. 2000. RADAR: an in-building RF-based user location and tracking system. *Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings.* Vol 2, 775-784. 2000.

BAKMAZ, B., BOJKOVIC, Z. AND BAKMAZ, M. Network Selection Algorithm for Heterogeneous Wireless Environment. *IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, 1-4. 2007.

BARI, F. AND LEUNG, V. Service Delivery over Heterogeneous Wireless Systems: Networks Selection Aspects. *Proceedings of the International Conference on Communications and Mobile Computing*, 251-256, July 3-6 2006.

BARI, F. AND LEUNG, V. Network Selection with Imprecise Information in Heterogeneous All-IP Wireless Systems. *Proceedings of the 3rd International Conference on Wireless Internet.* 2007.

BERNASCHI, M., CACACE, F., IANNELLO, G., ZA, S. AND PESCAPE, A. Seamless Internetworking of WLANs and Cellular Networks: Architecture and Performance Issues in a Mobile IPv6 Scenario. *IEEE Wireless Communications,* Vol. 12, Issue 3, 73-80. 2005.

BING, H., LINGGE, J., Performance Analysis of Vertical Handover in a UMTS-WLAN Integrated Network. *Proceedings of the 14th IEEE International Symposium on Personal, Indoor and Mobile Radio Communication.* 2003.

BING, B., Emerging Technologies in Wireless LANs - Theory, Design and Deployment. Cambridge University Press, UK. 2008.

BOURKE, P. Intersection point of two lines (2 dimensions). Available Online: *http:// astonomy.swin.edu.ac.uk/pbourke/geometry/lineline/2d/.* 1989.

BRIGNONE, C., CONNORS, T., LYON, G. AND PRADHAN, S. SmartLOCUS: An autonomous, self-assembling sensor network for indoor asset ad systems management. HPL-2003-41, HP Labs, 2003.

CAMPBELL, A.T. AND GOMEZ-CASTELLANOS, J. October. IP micro-mobility protocols. *ACM SIGMOBILE Mobile Computing and Communications Review* Vol. 4, Issue 4, 45-53, October 2000.

CHAKRAVORTY, R., VIDALES, P., SUBRAMANIAN, K., PRATT, I. AND CROWCROFT, J. Performance issues with vertical handovers - experiences from GPRS cellular and WLAN hot-spots integration. *Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications,* 155-164, 2004.

CHAN, P.M.L., HU, Y.F. AND SHERIFF, R.E. Implementation of fuzzy multiple objective decision making algorithm in a heterogeneous mobile environment. *Proceedings of the Wireless Comunications and Networking Conference,* Vol. 1, 332-336, 2002.

CHAN, P.M.L., SHERIFF, R.E., HU, Y.F., CONFORTO, P. AND TOCCI, C. Mobility management incorporating fuzzy logic for a Heterogeneous IP Environment. *IEEE Communications Magazine,* Vol. 39, Issue 12, 42-51. 2001.
CHEN, W., LIU, J. AND HUANG, H. An Adaptive Scheme for Vertical Handoff in Wireless Overlay Networks. *Proceedings of the 10th International Conference on Parallel and Distributed Systems.* 2004.

CHEN, W. AND SHU, Y. Active Application Oriented Vertical Handoff in Next Generation Wireless Networks. *Proceedings of IEEE WCNC,* 1383-1388. March 2005.

CIANCA, E., MOLINARO, A., SANCTIS, M.D., IERA, A., ARANITI, G. AND RUGGIERI, M. Location/Situation-Aware Architecture for Mobility Management over Heterogeneous Networks. *Proceedings of the 2nd international conference on Mobile multimedia communications.* 2006.

COTTINGHAM, D. AND VIDALES, P. Is Latency the Real Enemy in Next Generation Networks? *Proceedings of First International Workshop on Convergence of Heterogeneous Wireless Networks*, July 2005.

COTTINGHAM, D. Vehicular Wireless Communication. UCAM-CL-TR-741, January 2009.

CROWCROFT, J., COTTINGHAM, D., MAPP, G. AND SHAIKH, F. Y-Comm: A global architechture for heterogeneous networking. *Proceedings of WICON'07.* October 2007.

DUONG, H.H., DADEJ, A. AND GORDON, S. Proactive Context Transfer in WLAN-based Access Networks. *Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots,* 61-70. 2004.

EBERSMAN, H.G. AND TONGUZ, O.K. Handoff Ordering using Signal Prediction and Priority Queuing in Personal Communication Systems. *IEEE Transactions on Vehicular Technology* Vol. 48, 20-35. 1999.

EEFOCUS WEBSITE. Action frame format details - Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. http://www.eefocus.com/article/08-06/43512s.html. Last accessed 23/02/09.

ESTRIN, D., CULLER, D., PISTER, K. AND SUKHATME, G. Connecting the physical world with pervasive networks. *Pervasive Computing* Vol. 1, 59-69. 2002.

FUKUJU, Y., MINAMI, M., MORIKAWA, H. AND AOYAMA, T. DOLPHIN: an autonomous indoor positioning system in ubiquitous computing environment. *IEEE Workshop on Software Technologies for Future Embedded Systems.* 53-56. 2003.

FULOP, P., SZALKA, T., SZABO, S., LENDVAI, K. AND IMRE, S. Accurate mobility modelling and location prediction based on pattern analysis of handover series in mobile networks. *Proceedings of MoMM2008*, 219-226. 24-26 November 2008.

GANZ, A.,GANZ, Z. AND WONGTHAVARAWAT, K. Multimedia Wireless Networks - Technologies, Standards and QoS. Prentice Hall, New Jersey, USA, 2004.

GLISIC, S.G. Advanced Wireless Networks: 4G Technologies. John Wiley and Sons Ltd., England. 2006.

Global Positioning System. http://www.gps.gov/systems/gps/index.html. Last Accessed 13/07/2009.

GUO, Q., ZHU, J. AND XU, X. 2005. An adaptive multi-criteria vertical handoff decision algorithm for radio heterogeneous network. *Proceedings of the IEEE International Conference on Communications,* 2769-2872. 2005.

GUSTAFSSON, E. AND JONSSON, A. Always Best Connected. *IEEE Wireless Communications* Vol. 10, Issue 1, 49-55. 2003.

HADJIANTONIS, A.M., CHARALAMBIDES, M. AND PAVLOU, G.. A Policy-based Approach for Managing Ubiquitous Networks in Urban Spaces. *Proceedings of IEEE International Conference on Communications*, 2089-2096, Glasgow, UK, June 2007.

HADJIEFTHYMIADES, S. AND MERAKOS, L. Proxies + Path Prediction: Improving Web Service Provision in Wireless-Mobile Communications. *Mobile Networks and Applications,* Vol. 8, 389-399. Kluwer Academic Publishers. 2003.

HARTER, A., HOPPER, A., STEGGLES, P., WARD, A. AND WEBSTER, P. The Anatomy of a Context-aware Application. *Wireless Networks* Vol. 8, Issue 2-3, 187-197. March 2002.

HATIMI, A., KRISHNAMURTHY, P., PAHLAVAN, K., YLIANTTILA, M., MAKELA, J. AND PICHNA, R. 1999. Analytical Framework for Handoff in Non-Homogeneous Mobile Data Networks. *Proceedings of PIMRC'99,* September 1999.

HELLEBRANDT, M., MATHAR, R. AND SCHEIBENBOGEN, M. Estimating position and velocity of mobiles in a cellular radio network. *IEEE Transactions on Vehicular Technology* Vol. 46, Issue 1, 65-71. February 1997.

HINCH, S.W. Outdoor Navigation with GPS. Wilderness Press, 2nd Edition, 2004.

HONGYANG BING, CHEN HE AND LINGGE JIANG. Performance analysis of vertical handover in a UMTS-WLAN integrated network. *Proceedings of 14th IEEE Conference on Personal, Indoor and Mobile Radio Communications,* Vol 1, 187-191, 2003.

iPhone Website. iPhone 3G S Features. http://www.apple.com/uk/channel/iphone/orange_choose/features/. Last Accessed 10/10/2009.

JOHNSON, D., PERKINS, C. AND ARKKO, J. Mobility support in IPv6. IETF RFC 3775, June 2004.

JUNYANG ZHOU, YEUNG, W.M.-. AND NG, J.K.- Y. Enhancing Indoor Positioning Accuracy by Utilizing Signals from Both the Mobile Phone Network and the Wireless Local Area Network. *22nd International Conference on Advanced Information Networking and Applications,*138-145. 2008.

KASSAR, M., KERVELLA, B. AND PUJOLLE, G. An Intelligent Handover Management System for Future Generation Wireless Networks. *EURASIP Journal on Wireless Communications and Networking*. 2008.

KATZ, R.H. AND GIESE, J. Policy-enabled handoffs across heterogeneous wireless networks. *Proceedings of IEEE Workshop on Mobile Computing Systems and Applications*, 51-60. 1999.

KHALAF-ALLAH, M. A Novel GPS-free method for mobile unit global positioning in outdoor wireless environments. *Wireless Personal Communications* Vol. 44, Issue 3, 2008.

KIKILIS, A. AND ROUSKAS, A. Formulation of Optimization Problems for Access Selection in Next Generation Wireless Networks. *Proceedings of 3rd International conference on Mobile Multimedia Communication*, August 27-29 2007.

KUPPER, A. Location-Based Services - Fundamentals and Operation. John Wiley and Sons Ltd. 2005.

LAN/MAN STANDARDS COMMITTEE, IEEE 802.11 Specification document - Wireless LAN Medium Access Control and Physical Layer Specifications. 1999-Release.

LEWIS, L. Implementing Policy in Enterprise Networks. *IEEE Communications Magazine* Vol. 34, Issue 1, 50-55. January 1996.

LI, M., CHEN, S. AND XIE, D. An Adaptive Multi-criteria Vertical Handoff Decision Algorithm, *Proceedings of the 2nd ACM workshop on Performance monitoring and measurement of heterogeneous wireless and wired networks,* 119-123, 2007.

LIANG, B. AND HAAS, Z.J. Predictive distance-based mobility management for multi-dimensional PCS networks. *IEEE/ACM Transactions on Networking* Vol. 11, Issue 5, 718-732, October 2003.

LIAO, Y. AND GAO, L. Practical Schemes for Smooth MAC Layer Handoff in 802.11 Wireless Networks. *Proceedings of the 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks*, 181-190. 2006.

LIU, G. AND MAGUIRE, G.J. A Class of mobile motion prediction algorithms for wireless mobile computing and communications. *Mobile Networks and Applications* Vol. 1, 113-121, 1996.

LIU, M., LI, Z., GUO, X. AND DUTKIEWICZ, E. Performance Analysis and Optimization of Handoff Algorithms in Heterogeneous Wireless Networks. *IEEE Transactions on Mobile Computing* Vol. 7, Issue 7, 846-857, July 2008.

MAKELA, J., YLIANTTILA, M. AND PAHLAVAN, K. Handoff decision in multi-service networks. *The 11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*. Vol. 1, 655-659, 2000.

MAPP, G., COTTINGHAM, D., SHAIKH, F., VIDALES, P., PATANAPONGPIBUL, L., BALIOSIAN, J. AND CROWCROFT, J. An Architectural Framework for Heterogeneous Networking. *Proceedings of the 3rd WINSYS International Conference*, August, 2006.

MAPP, G., SHAIKH, F., AIASH, M., VANNI, R.P., AUGUSTO, M. AND MOREIRA, E. Exploring Efficient Imperative Handover Mechanisms for Heterogeneous Wireless Networks. *IEEE International Symposium on Emerging Ubiquitous and Pervasive Systems*, 19-21 August 2009.

MARTI, A.B. AND GUERRERO, J.A. A network selection mechanism for next generation mobile communications. *Third IET International Conference on Intelligent Environments*, 22-29. 2007.

MCNAIR, J. AND FANG ZHU. Vertical handoffs in fourth-generation multi-network environments. *IEEE Wireless Communications* Vol. 11, Issue 3, 8-15. 2004.

MICHAELIS, S., LEWANDOWSKI, A., DANIEL, K., YOUSAF, F.Z. AND WIETFELD, C. A Comprehensive Mobility Management Solution for Handling Peak Load in Cellular Network Scenarios. *Proceedings of the 6th ACM International Symposium on Mobility Management and Wireless Access* , 9-18. 2008.

MING-HSING CHIU AND BASSIOUNI, M.A. Predictive schemes for handoff prioritization in cellular networks based on mobile positioning. *IEEE Journal on Selected Areas in Communications* Vol. 18, Issue 3. 510-522. 2000.

MISHRA, A., SHIN, M. AND ARBAUGH, W. An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process. *ACM SIGCOMM Computer Communications Review* Vol. 33, Issue 2. 93-102. 2003.

MOFFET, J.D. AND SLOMAN, M.S. Policy Hierarchies for Distributed Systems Management. *IEEE Journal on Selected Areas in Communications* Vol. 11, Issue 9, 1404-1414.1993.

MURRAY, K., MATHUR, R. AND PESCH, D. Intelligent Access and Mobility Management in Heterogeneous Wireless Networks using Policy. *Proceedings of the 1st international Symposium on information and Communication Technologies*. 181-186. September 24 – 26, 2003.

MUTHUKRISHNAN, K., LIJDING, M. AND HAVINGA, P. Towards Smart Surroundings: Enabling Techniques and Technologies for Localization. *Location and Context-Awareness,* Springer / Heidelberg, Berlin, 350-362. 2005.

NAVIDI, W. AND CAMP, T. Predicting node location in a PCS network. *Proceedings of the International Conference on Performance, Computing and Communications* 165-170. 2004.

OPNET Simulation Manual Documentation. 1990.

Orange Price Plans. Website: http://shop.orange.co.uk/mobile-phones/price-plans/pay-monthly?WT.svl=205. Last Accessed: 11/11/2009.

PAREEK, D. WiMax: Taking Wireless to the MAX. First Edition. Auerbach Publications, Florida, USA. 2006.

PERKINS, C.E. Mobile IP. *IEEE Communications Magazine* Vol. 40, Issue 5. 66-82. 2002.

PERSSON, H. AND KARLSSON, J.M. An Approach to Structure the Handover Terminology. *Proceedings of SNCNW'05.* 2005.

RADUNOVIC B., GKANTSIDIS C., GUNAWARDENA D. KEY P. Balancing TCP over multiple paths in wireless mesh networks. *Proceedings of MSN'08.* 2008.

RAICIU C. AND HANDLEY M. Designing a Multipath Transport Protocol. *Proceedings of MSN'08.* 2008.
RANDELL, C. AND MULLER, H. Low Cost Indoor Positioning System. *Proceedings of UBICOMP'01* 42-48. 2001.

RODRIGUEZ, P., CHAKRAVORTY, R., CHESTERFIELD, J., PRATT, I. AND BANERJEE, S. MAR: A Commuter Router Infrastructure for the Mobile Internet. *Proceedings of the ACM MOBISYS conference*, Boston, Massachusetts, USA, June, 2004.

SALEH, A. A Location-aided Decision Algorithm for Handoff Across Heterogeneous Wireless Overlay Networks. Masters Thesis, Virginia Polytechnic Institute and State University. July 2004.

SAATY, T.L. The Analytical Hierarchy Process. Wiley, New York. 1980.

SCHILLER, J. Mobile Communications. Pearson Education, England. 2003.

SHAIKH, F., LASEBAE, A. AND MAPP, G. Client-based SBM Layer for Predictive Management of Traffic Flows in Heterogeneous Networks. *Proceedings of ICTTA'06,* April 2006.

SHAIKH, F., MAPP, G. AND LASEBAE, A. Proactive policy management using TBVH mechanism in heterogeneous networks. *Proceedings of NGMAST'07* September 10-13 2007.

SHEN, X., MARK, J.W. AND YE, J. User Mobility Profile Prediction: An Adaptive Fuzzy Inference Approach. *Wireless Networks* Vol. 6, 363-374, 2000.

SHENOY, N. A Framework for Seamless Roaming Across Heterogeneous Next Generation Wireless Networks. *Wireless Networks* Vol. 11, 757-774. 2005.

SHENOY, N. AND MISHRA, S. Vertical Handoff and Mobility Management for Seamless Integration of Heterogeneous Wireless Access Technologies. *Heterogeneous Wireless Access Networks, S*pringer, New York, 181-215. 2008.

SMITH, C. AND COLLINS, D. 3G Wireless Networks. Tata McGraw-Hill, Second Edition, New Delhi. 2007.

SOH, W. AND KIM, H., S. Dynamic Bandwidth Reservation in Cellular Networks using Road Topology Based Mobility Predictions. *Proceedings of the 23rd IEEE INFOCOM Conference,* Vol. 4, 2766-2777. March, 2004.

SONG, Q. AND JAMALIPOUR, A. An adaptive quality-of-service network selection mechanism for heterogeneous mobile networks. *Wireless Communications and Mobile Computing* Vol. 5, 697-708. 2005.

SONG Q. AND JAMALIPOUR, A. Network selection in an integrated wireless LAN and UMTS environment using mathematical modeling and computing techniques. *IEEE Wireless Communications* Vol. 12, 42-48. 2005.

SONG Q. AND JAMALIPOUR, A. A Negotiation-Based Network Selection Scheme for Next-Generation Mobile Systems. *Proceedings of IEEE GLOBECOM'06,* 1-5. 2006.

SONG, W., JIANG, H., ZHUANG, W. AND SHEN, X. Resource Management for QoS Support in Cellular/WLAN Interworking. *IEEE Network* 12-18. September/October 2005.

STEVENS-NAVARRO, E., YUXIA LIN AND WONG, V.W.S. An MDP-Based Vertical Handoff Decision Algorithm for Heterogeneous Wireless Networks. *IEEE Transactions on Vehicular Technology* Vol. 57, Issue 2, 1243-1254. 2008.

STUBER, G.L. Principles of Mobile Communication. Second Edition, Kluwer Academic Publishers, Norwell, MA. 2001.

SUNGHOON SEO, SEUNGCHAN LEE AND JOOSEOK SONG. Policy Based Intelligent Vertical Handover Algorithm in Heterogeneous Wireless Networks. *Proceedings of International Conference on Convergence Information Technology*, 1900-1905. 2007.

TAFAZOLLI, R., Ed. Technologies for the Wireless Future. John Wiley and Sons Ltd., England. 2005.

TECHNOLOGIES, Aglient. LTE and the Evolution to 4G Wireless. WileyBlackwell, 2009.

TRAN, P.N. AND BOUKHATEM, N. The Distance to the Ideal Alternative (DiA) Algorithm for Interface Selection in Heterogeneous Wireless Networks. *Proceedings of MobiWac'08.* 61-68. October 30-31 2008.

VBOX Accuracy. *http://www.racelogic.co.uk/?show=VBOX-Accuracy*. Last Accessed 07/04/2009.

VERMA, D.C. Simplifying network administration using policy-based management. *IEEE Network* Vol. 16, Issue 2, 2002.

VIDALES, P. Seamless Mobility in 4G systems. Technical Report. UCAM-CL-TR-656, November 2005.

WANG, J., PRASAD, R.V. AND NIEMEGEERS, I.G.M.M. Solving Incertitude of Vertical Handovers in Heterogenoeus Mobile Wireless Network. *Proceedings of WICON'07,* October 22-24, 2007.

WANG, H.J., KATZ, R.H. AND GIESE, J. Policy-enabled handoffs across heterogeneous wireless networks. *Proceedings of Second IEEE Workshop on Mobile Computing Systems and Applications,* 51-60.1999.

WEN-HONG ZHU AND LAMARCHE, T. Velocity estimation by using imperfect accelerometer and encoder for rigid contact modeling and control. *Proceedings of IEEE International Conference on Robotics and Automation,* 4234-4239. 2006.

WEN-TSUEN CHEN, JEN-CHU LIU AND HSIEH-KUAN HUANG. An adaptive scheme for vertical handoff in wireless overlay networks. *Proceedings of Tenth International Conference on Parallel and Distributed Systems,* 541-548. 2004.

WEN-TSUEN CHEN AND YEN-YUAN SHU. Active application oriented vertical handoff in next-generation wireless networks. *Proceedings of IEEE Wireless Communications and Networking Conference,* Vol. 3. 1383-1388, 2005.

WISELY, D. IP for 4G. WileyBlackwell, England. 2009.

WU, G., MITSUHIKO, M. AND HANVINGA, P.J.M. MIRAI Architecture for heterogeneous network. *IEEE Communications Magazine* Vol. 40, 126-134. February 2002.

XIAOHUAN YAN, MANI, N. AND CEKERCIOGLU, Y.A. A Traveling Distance Prediction Based Method to Minimize Unnecessary Handovers from Cellular Networks to WLANs. *IEEE Communications Letters* Vol. 12, 14-16. 2008a.

XIAOHUAN YAN, SEKERCIOGLU, Y.A. AND MANI, N. A method for minimizing unnecessary handovers in heterogeneous wireless networks. *Proceedings of International Symposium on World of Wireless, Mobile and Multimedia Networks,* 1-5. 2008b.

YANG, L. Access Network Selection in 4G Network Environment. Masters Thesis. Unversity of Waterloo. 2007.

YANG, X., YAP, J.H., GHAHERI-NIRI, S. AND TAFAZOLLI, R. Enhanced position assisted soft handover algorithm for UTRA. *Proceedings of IEEE Vehicular Technology Conference,* Vol. 2, 567-571. 2001.
YOON, K.P. AND HWANG, C.-. Multiple Attribute Decision Making: An Introduction. *Sage University Paper Series on Quantitative Applications in the Social Sciences,* Thousand Oaks: CA. 07-104. 1995.

YLIANTTILA, M., PANDE, M., MAKELA, J. AND MAHONEN, P. Optimization scheme for mobile users performing vertical handoffs between IEEE 802.11 and GPRS/EDGE networks. *Proceedings of IEEE GLOBECOM'01,* Vol. 6, 3439-

3443, 2001.

ZAHARIADIS, B.T. Home Networking Technologies and Standards. Artech House, Norwood, MA. USA. 2003.