LOCATION-AWARE DECISION ALGORITHM FOR HANDOVER ACROSS HETEROGENEOUS WIRELESS NETWORK.

BY

STANLEY O.OKOYE

This thesis is submitted in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering with Emphasis on Telecommunication.

Department of Telecommunication

Blekinge Institute of Technology

Karlskrona, Sweden

August 2010
DEDICATION:

To my wife Gloria Okoye, and my child Kamsi Okoye for their love, patience while doing this work; to my brothers Stephen E. Okoye and Augustine I. Okoye for their wonderful support and encouragement.
ABSTRACT

Vertical handover is the processes that switch a mobile node from one technology to another in order to maintain a communication in a network. Heterogeneous Networks are two Networks whose entities support different technologies. Because of the benefits brought about by 3G networks such UMTS, it is increasingly desirable to integrate 3G networks with WLAN. WLAN is a low coverage, high speed network compared to 3G networks. Consequently, WLAN is used to extend 3G networks at certain locations in order to provide improved services and address QoS issues. To achieve a beneficial vertical handover in a network, an algorithm that departs from the conventional RF based algorithm is necessary. An attempt is made in this study to provide such algorithm which aims to utilize location information stored in a WLAN coverage database, and the location service entities of UTRAN as defined by 3GPP to determine a valuable/beneficial vertical handover between UMTS and WLAN. RF based conventional downward vertical handovers can be inefficient and wastes resources. This study aims to correct the lapses associated with conventional RF based vertical handover across heterogeneous network.

Keywords:

Wireless LANs, Location Services, Vertical Handover, Location-Aided
ACKNOWLEDGEMENT

I am grateful to the following people for assisting me in many ways during this work: My wife Gloria Nkechi Okoye and my child Kamsi R. Okoye, for their patience; my brothers Stephen and Augustine Okoye for supporting me in many ways.

My special thanks goes to my supervisor, Professor Adrian Popescu for his guidance; my program manager Mikael Åsman and the Student Administrator Lena Magnusson for giving me invaluable pieces of advice.

Thanks to many others whose names I could not mention here as a result of space constraints.
# TABLE OF CONTENTS

## Chapter 1-INTRODUCTION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Stating The Problem</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Thesis Objective</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Solution In View</td>
<td>4</td>
</tr>
<tr>
<td>1.4 Analysis Overview</td>
<td>4</td>
</tr>
<tr>
<td>1.5 Thesis Outline</td>
<td>5</td>
</tr>
</tbody>
</table>

## Chapter 2–UMTS NETWORK

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0 Introduction</td>
<td>6</td>
</tr>
<tr>
<td>2.1 Overview Of Third Generation Technology</td>
<td>6</td>
</tr>
<tr>
<td>2.2 Standardization Of UMTS</td>
<td>7</td>
</tr>
<tr>
<td>2.3 UMTS Architecture</td>
<td>8</td>
</tr>
<tr>
<td>2.4 UMTS Radio Interface</td>
<td>10</td>
</tr>
<tr>
<td>2.5 UMTS Access Network(UTRAN)</td>
<td>10</td>
</tr>
<tr>
<td>2.6 UMTS Core Network</td>
<td>11</td>
</tr>
<tr>
<td>2.7 UTRAN Protocol Stack</td>
<td>13</td>
</tr>
<tr>
<td>2.8 UMTS/WLAN LAN Integration</td>
<td>15</td>
</tr>
<tr>
<td>2.8.1 Requirement For Integration Between UMTS And WLAN</td>
<td>15</td>
</tr>
<tr>
<td>2.8.2 Integration benefits</td>
<td>15</td>
</tr>
</tbody>
</table>
Chapter 3- Overview of Local Area Networks

3.0 Introduction

3.1 IEEE 802.11 Standards And Versions

3.1.1 IEEE 802.11b

3.1.2 IEEE 802.11a

3.2 Hardware Entities Of Local Area Network

3.3 Transmission Techniques in WLANs

3.4 WLAN Networks and System Architecture

3.4.1 WLAN Infrastructure Networks and System Architecture

3.4.2 WLAN ADHOC Networks and System Architecture

3.5 WLAN protocol architecture

3.5.1 Physical Layer

3.5.2 The Mac Layer

3.5.3 Mac Management

3.5.3.1 Roaming

3.6 Chapter Summary

Chapter 4-HandOver In UMTS and WLAN

4.0 Introduction

4.1 Handover, A Brief History

4.1.1 Types Of Handovers

4.2 Mobile IP

4.3 Positioning Overview

4.3.1 Location Services (LCS)

4.3.2 Positioning Functions
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3.3 UTRAN LCS architecture</td>
<td>34</td>
</tr>
<tr>
<td>4.3.4 Positioning Method</td>
<td>36</td>
</tr>
<tr>
<td>4.3.4.1 Assisted GPS(AGPS)</td>
<td>36</td>
</tr>
<tr>
<td>4.3.4.2 UTDOA</td>
<td>37</td>
</tr>
<tr>
<td>4.4 Chapter Summary</td>
<td>38</td>
</tr>
</tbody>
</table>

**Chapter 5-Motivation**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0 Introduction</td>
<td>39</td>
</tr>
<tr>
<td>5.1 Handover Procedures</td>
<td>39</td>
</tr>
<tr>
<td>5.1.1 Downward Vertical Handover</td>
<td>41</td>
</tr>
<tr>
<td>5.1.1.1 Discovery</td>
<td>42</td>
</tr>
<tr>
<td>5.1.1.2 Address Configuration</td>
<td>43</td>
</tr>
<tr>
<td>5.1.1.3 AAA</td>
<td>43</td>
</tr>
<tr>
<td>5.1.1.4 Stabilization</td>
<td>44</td>
</tr>
<tr>
<td>5.1.2 Upward Vertical Handover</td>
<td>45</td>
</tr>
<tr>
<td>5.2 Aggregate Handover Effect</td>
<td>45</td>
</tr>
<tr>
<td>5.3 Problem Description</td>
<td>46</td>
</tr>
<tr>
<td>5.4 Possible Solution</td>
<td>47</td>
</tr>
<tr>
<td>5.5 Chapter Summary</td>
<td>48</td>
</tr>
</tbody>
</table>

**Chapter 6-Design Consideration**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.0 Introduction</td>
<td>49</td>
</tr>
<tr>
<td>6.1 Conventional Approach (Discovery And Handover)</td>
<td>50</td>
</tr>
<tr>
<td>6.2 Proposed Location Aided Handover Decision Algorithm</td>
<td>51</td>
</tr>
<tr>
<td>6.2.1 Downward Vertical Handover Decision Algorithm</td>
<td>52</td>
</tr>
<tr>
<td>6.2.2 Horizontal Handover Decision Within WLAN</td>
<td>55</td>
</tr>
<tr>
<td>6.2.3 Upward Handover Decision</td>
<td>57</td>
</tr>
</tbody>
</table>
Chapter 6-Analysis

6.3 WLAN Coverage Database
6.3.1 Database Structure
6.3.2 WLAN Footprint Accuracy
6.3.3 Granularity

6.4 Mathematical Model Algorithm’s Performance
6.4.1 Assumptions
6.4.2 Required Visit Duration
6.4.3 MN’s Travel Speed
6.4.4 Predicted Path Length
6.4.5 User Travel Direction
6.4.6 Coverage Query

6.5 The Proposed Architecture
6.6 Communication Protocol
6.7 Chapter Summary

Chapter 7-Analysis

7.0 Introduction
7.1 Qualitative Analysis Of The Algorithm Performance
7.1.1 Ideal Algorithm Performance
7.1.1.1 Scenario 1 -Downward Vertical Handover
7.1.1.2 Scenario 2 - Layer 3 Horizontal Handover
7.1.2 Proposed Algorithm Limitations
7.1.2.1 Physical Boundary Limitation
7.2 Summary

Chapter 8-Conclusion
8.0 Thesis Summary 79
8.1 Conclusion 81
8.2 Future Work 82
REFERENCES 84
LIST OF FIGURES AND TABLES

LIST OF FIGURES

Figure 2.0 IMT -2000 Frequencies as used by different countries 7
Figure 2.1 UMTS main components and reference architecture 8
Figure 2.2 UMTS domain and interfaces 9
Figure 2.3 Basic architecture of the UTRA Network 11
Figure 2.4 UMTS core Network comprising 3G RNS and 2G BSS 12
Figure 2.5 UMTS user plane 13
Figure 2.6 UMTS control plane 14
Figure 2.7 Interaction between RRC and Lower Layers in C-Plane 14
Figure 3.1 A WLAN Infrastructure Network 20
Figure 3.2 WLAN Infrastructure System Architecture 20
Figure 3.3 WLAN Adhoc Network and Architecture 21
Figure 3.4 IEEE 802.11 Protocol Architecture and Bridging 22
Figure 3.5 IEEE 802.11 Protocol Architecture with Management 22
Figure 4.1A, A Large Single Cell 26
Figure 4.1B The same cell with frequency reuse 26
Figure 4.2 Hard Handover, Soft Handover and Softer Handover 29
Figure 4.3 Layer 2 And Layer 3 handover 30
Figure 4.4 Mobile IP Example Network 32
Figure 4.5 General arrangement of LCS in UMTS 35
Figure 4.6 Clearer View Of Entities that support LCS On UTRAN
Figure 4.7 A-GPS Positioning Method
Figure 5.1 Downward Vertical Handover Procedures

Figure 6.1 Conventional RF Based Upward and Downward Handover
Figure 6.2 The Proposed Location-Aided Decision Algorithm
Figure 6.3 The Proposed Location-Aided Decision Algorithm: User Connected To WLAN Network
Figure 6.4 An Interface Of Site Survey Software
Figure 6.5 Visual Representation Of Three Velocity Vectors From 4 Position Fixe
Figure 6.6 Obtaining the TEP Coordinates From The Users’ Position, Direction And PPL
Figure 6.7 Visual Representation Of The Coverage Query Mechanism Based On TEP Coordinates
Figure 6.8 Architecture For The Proposed Solution
Figure 6.9 Protocol For Message Exchange Between UMTS And MN

Figure 7.1 Scenario 1, Limitation Of The Conventional Approach
Figure 7.2 Scenario 1, Location-Aided Algorithm Applied
Figure 7.3 Scenario 1, Limitation Of The Conventional Approach
Figure 7.4 Scenario 2, The Location-Aided Algorithm In Operation
Figure 7.5 Performance Of The Algorithm, User in Hotspot 2
Figure 7.6 Scenario Demonstrating The Physical Boundary Limitation

LIST OF TABLES

Table 2.1 UMTS Data Rated, Cell Size and Coverage Area
Table 3.1 ISM frequency Spectrum
Table 6.1 Acronym Used in Equation and Their Description
LIST OF ABREVIATIONS

1G        First Generation
2G        Second Generation
3G        Third Generation
4G        Fourth Generation
AMPS      Advance Mobile Phone Service
AP        Access Point
ASST      Application Signal Strength Threshold
BMC       Broadcast and Multicast Sub layer
BS        Base Station
BSS       Basic Service Set
BTS       Base Transceiver Station
CA        Collision Avoidance
CDMA      Code Division Multiple Access
CPCH      Common Packet Channel
CRC       Cyclic Redundancy Check
CS        Circuit Switch
CSMA      Carrier Sense Multiple Access Technique
DCF       Distributed Coordination Function
DCH       Dedicated Channel
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPDCH</td>
<td>Dedicated Physical Control Channel</td>
</tr>
<tr>
<td>DS</td>
<td>Distributed Service</td>
</tr>
<tr>
<td>DS-CDMA</td>
<td>Direct-sequence Code Division Multiple Access</td>
</tr>
<tr>
<td>DT</td>
<td>Discovery Time</td>
</tr>
<tr>
<td>ESS</td>
<td>Extended Service Set</td>
</tr>
<tr>
<td>ESSID</td>
<td>Extended Service Set ID</td>
</tr>
<tr>
<td>ETSI</td>
<td>European Telecommunication Standard Institute</td>
</tr>
<tr>
<td>FDD</td>
<td>Frequency Division Duplex</td>
</tr>
<tr>
<td>FDMA</td>
<td>Frequency Division Multiple Access</td>
</tr>
<tr>
<td>GPRS</td>
<td>General Packet Radio Service</td>
</tr>
<tr>
<td>GSM</td>
<td>Global System for Mobile Communication</td>
</tr>
<tr>
<td>HCC</td>
<td>Handover Control Center</td>
</tr>
<tr>
<td>HE</td>
<td>Handover Execute</td>
</tr>
<tr>
<td>HiperLAN</td>
<td>High Performance Radio Local Area Network</td>
</tr>
<tr>
<td>HSDPA</td>
<td>High Speed Downlink Packet Access</td>
</tr>
<tr>
<td>HSPA</td>
<td>High Speed Packet Access</td>
</tr>
<tr>
<td>IBSS</td>
<td>Independent Basic Service Set</td>
</tr>
<tr>
<td>IR</td>
<td>Infra red</td>
</tr>
<tr>
<td>ITU</td>
<td>International Telecommunication Union</td>
</tr>
<tr>
<td>MAC</td>
<td>Medium Access Control</td>
</tr>
<tr>
<td>MD</td>
<td>Mobile Download</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.0 INTRODUCTION

In the early days of wireless communication, radio channels were used mainly to transmit voices. However, as telecommunication systems evolve from 1st generation to the present 4th generation (LTE, all-IP core network is under test in a number of countries), improvements have been made over channel bandwidth for transmission of voices and data. Consequently, radio channels are being used to transmit voice, data, and video and multimedia services in 3rd generation telecommunication networks. These services and other emerging multimedia services demand higher data rates to achieve an improved quality of services (QoS).

UMTS evolved from GSM GPRS. It is a third generation system aimed at providing a common worldwide communication system, in addition to providing improved bandwidth for multimedia services. Also, UMTS was intended to allow for user and terminal mobility supporting the idea of universal personal communication (UPT). However, the technology is not exactly suitable for use in areas such as densely populated areas or for indoor. Therefore, researchers and standardization bodies have considered other types of technologies that could be utilized to extend 3G/UMTS network services in such areas. The characteristics of Wireless LAN (WLAN) technology allow it to provide high data rates in indoor, small, and densely populated areas. This made it the primary candidate for extending 3G/UMTS in a complementing manner. Since the IEEE 802.11 standard family is the most common type of WLAN technology in many countries, the majority of the research efforts for integrating 3G/UMTS with WLAN technology have been geared towards using this family of standard.

In this study, UMTS and IEEE 802.11 were used as the main representative technologies of 3G and WLAN, respectively. Note that “WLAN” and “802.11” are used interchangeably in this work and “3G/UMTS” is used to refer to UMTS or 3G. As expected, integrating two very different technologies, such as 3G/UMTS and IEEE 802.11, introduced a number of technical issues that must be resolved in order to maximize the benefits envisaged from such integration. As mentioned in this study, 3G/UMTS and IEEE 802.11 are heterogeneous technologies that have different attributes (i.e., data rates, structure, etc.), mechanisms (i.e., coverage discovery, handover decision, etc.), and administrative domains. Handover between access points or base stations has long been an issue within the wireless telecommunication field. However, a higher
level of handover complexity is introduced due to the differences between inter-networked heterogeneous wireless networks. This added complexity causes additional delays to the handover processes and utilizes more resources, from the mobile node’s (MN/MS) perspective and the network’s perspective. These handover-related delays typically result in a period of instability caused by the interruption in data transfer as the handover procedures take place and the MN’s traffic gets re-routed to its new location (i.e., over the new connection). The longer the handover-related delays the more severe the performance degradation, particularly with respect to the application-level throughput and response time. The handover from 3G/UMTS to WLAN is not done out of necessity (i.e., the need for coverage) but rather for performance optimization purposes. Therefore, ideally such handovers should only be triggered if it will result in enhanced performance. Typically, the data rates offered by WLANs are significantly higher than those offered by 3G/UMTS networks. Therefore, the degradation in performance that occurs as a result of the handover-related delays is compensated for through the higher data rates, and thus does not affect the overall level of performance observed by the user. This is only possible if the user’s visit to the WLAN coverage area is long enough to:

(a) Complete the reconfiguration and mobility management procedures, the security procedures, and the accounting procedures.

(b) Allow the recovery of the upper layer protocols and the application(s) after the handover, in terms of getting used to the new data rates and returning to stable performance level (i.e., take appropriate advantage of the higher data rates)

(c) Transfer sufficient user data packets to compensate for the interruption or throttling in data transfer, which took place during the handover procedures.

1.1 STATING THE PROBLEM

The increase in demand for wireless communication has become explosive from the past few years till date. This is evident in the number of licenses issued to operators of wireless communication in various countries. The demand for services provided by mobile service operators are increasing also. To meet these demands, therefore, measures must be in place to support transmission of high data rate and maintenance of quality of service (Qos).

One of the steps already taken to improve capacity and QoS in a 3G network is the integration of 3G network, such as UMTS with WLAN network. WLAN network (IEEE 802.11) has been used to extend UMTS in such areas as airport, subway stations, campuses, parks etc. This extension also meant that a subscriber to a UMTS network is going to traverse, within small time period, a lot more WLAN networks whose coverage areas are small. Due to the benefits achieved from transferring users from 3G/UMTS to WLAN, the default procedure is expected to trigger a handover from 3G/UMTS to (the higher data rate) WLANs upon discovering coverage availability. Therefore, the rate of handoff will increase significantly due to the user’s short visit duration to the discovered WLAN(s) coverage areas.
In most technologies, the conventional approach for coverage discovery and handover decisions is based entirely on RF measurements (i.e., received signal strength, and signal to noise ratio, etc.). A few technologies such as UMTS, take into consideration other factors in addition to the RF measurements. For instance, the base stations in UMTS are aware of the status of their neighboring base stations and can transfer users to neighboring cells merely for load balancing purposes. However, the IEEE 802.11 standard does not have such capabilities, and thus relies primarily on physical layer (PHY) measurements and the MAC layer evaluation of such measurements to determine the appropriate time for triggering the handover. According to the current state of technology, a dual-mode (UMTS-WLAN) mobile node (MN) would rely entirely on RF measurements for discovering WLAN coverage availability and for deciding on the suitability of the handover to the discovered WLAN. Therefore, in environments that have smaller WLAN cells and high user mobility, the conventional RF-based handover decision method could trigger unnecessary handovers to WLANs. These handovers are unnecessary and inefficient since the user does not remain in the WLAN’s coverage area long enough to benefit from its higher data rates. Handover procedures, and particularly those needed for handovers across heterogeneous networks, have significant delay. Therefore, by the time the handover procedures have completed and the stabilization period was over, the user could already be headed out of such coverage area and requiring a handover back to UMTS or to another WLAN. As with any handover, this degrades performance in terms of interrupting data transfer and increasing application response time. Moreover, due to the short visit duration the application does not have the chance to benefit from the higher data rates which compensate for the handover-related delays. Therefore, the performance degradation will have a noticeable effect from the user’s perspective. This defeats the purpose and counters the objective of inter-networking 3G/UMTS and IEEE 802.11 and performing handovers to WLANs. The problem described here is a result of the limitations of the conventional handover decision method, which lacks awareness of the user’s position, speed, and direction. This knowledge is necessary to prevent inefficient handovers to the discovered WLANs, which only result in performance degradation and unnecessary costs. Therefore, a different approach that combines the RF measurements with the location and spatial parameters of the user and the WLAN coverage area is needed to mitigate this issue and prevent its performance-degrading effects.

1.2 THESIS OBJECTIVE

In this research study, effort will be concentrated in studying the feasibility and benefits of aiding conventional radio frequency based handover decision mechanism to execute location-aware handover procedures. The scope is narrowed down to downward vertical handover decision across heterogeneous networks. The underline objective is to eliminate or at least reduce the triggering of handovers that result in performance degradation and wasting of resources due to the user’s short visit to the discovered network coverage area.
1.3 SOLUTION IN VIEW

Location based information is supported by location services (LCS) architecture which has been defined by 3GPP. The proposed solution is based on location based information which has continued to attract the attention of researchers. The emergence of “emergency call services”, to locate users calling the emergency line from a wireless node, as well as the emerging commercial applications utilizing location-aware information, have prompted further research to support such services. The relevant standardization bodies have developed architecture and protocol plans to support such services in technologies such as 3G/UMTS. In addition to utilizing location-based information in user applications or emergency situations, the standardization efforts have researched the use of location based services for network optimizations purposes. Therefore, these new developments were utilized in this research effort to study the feasibility of a location-aided solution, which consists of combining the RF-based conventional handover decision mechanism with a location-based evaluation mechanism. A location-aided handover decision algorithm was developed to run a location-based evaluation at a node in the 3G/UMTS control network. The RF measurements at the MN’s WLAN adapter were used to discover WLAN coverage and to determine the suitability of the handover from an RF-perspective (i.e., based on Received Signal Strength, Signal to Noise Ratio, etc). Upon discovery of WLAN coverage, the MN triggers the location-based handover evaluation at the UTRAN and waits for the response before proceeding with the Layer-3 handover procedures to the discovered WLAN. The different aspects of Layer-1 to Layer-3 handovers are described further in this study. The location-based evaluation relies on a positioning method for obtaining user location fixes. Once the user’s position, speed, and direction are computed, the evaluation takes place to predict the user’s trajectory with respect to the WLAN’s spatial characteristics. The evaluation node will accordingly predict the user’s trajectory and whether the user will remain in the same WLAN coverage area long enough to result in handover benefit (i.e., allow the application to transfer sufficient data over the higher data rate connection). If so, then the MN receives a notification from the UMTS network recommending the handover. Otherwise, the evaluation cycle will periodically trigger to re-assess the user’s trajectory. To avoid infinite triggering of the evaluation cycle, the MN notifies the UMTS network if the user exits the WLAN coverage area, which is determined by the PHY layer measurements obtained through the WLAN adapter. The underlying mechanism for estimating the user’s trajectory and for predicting the user’s visit duration, are described later, along with the design requirements.

1.4 ANALYSIS OVERVIEW

The performance of the algorithm was demonstrated and analyzed by comparing it to the performance of the conventional handover decision method for making decisions in heterogeneous wireless overlay networks. Two primary scenarios were used to demonstrate the weaknesses of the current conventional handoff decision method and demonstrate the capabilities of the proposed algorithm, which was designed to tackle the source of such weaknesses. The limitations of the algorithm were also addressed. Moreover, the algorithm’s
benefits were demonstrated by comparing the outcome of the conventional handover decision method to that of the algorithm’s outcome, in a quantitative manner. The metrics for comparing the outcome of the two methods were the application-level throughput and application response time. Finally, based on such evaluation and analysis of the proposed algorithm’s capabilities and limitations, specifications were established to point out the type of environments that would be most suitable for deploying the algorithm.

1.5 THESIS OUTLINE.

In chapter 2, detailed background information is given on UMTS network. UMTS is the technology studied for the purpose of this research work. Chapter 3 focused on WLAN technology, its architecture and networks. WLAN network is the choice network studied with UMTS to achieve the objective of this research study. In chapter 4, handover procedures across heterogeneous network(UMTS/WLAN) was discussed generally while in chapter 5 the problem motivating the research work was discussed. Chapter 6 and 7 treated “Design Consideration” and “Analysis” of observation respectively. Chapter 8 give conclusion arrived at after the research study and also recommended future work.
CHAPTER 2
UMTS NETWORK

2.0 INTRODUCTION

In this Chapter, the UMTS network is brought to the reader. The aim is to provide more information on the radio and access network of UMTS network in order to help the reader understand the concept of location-aware handover procedure between heterogeneous networks. This thesis work focuses on handover scenarios between wireless overlay networks assumed to be complementary and integrated to an extent. Together in the treatment of UMTS network, the chapter will also discuss UMTS integration with WLAN and its benefit. The chapter also discussed the reason behind the integration of UMTS with WLAN. WLAN was discussed in Chapter 3.

2.1 OVERVIEW OF THIRD GENERATION TECHNOLOGY.

The International Telecommunication Union (ITU) envisaged a telecommunication system which could allow user and terminal mobility to support the concept of a universal personal communication system. Consequently, ITU made request for proposals for a radio transmission technology (RTT) for the international mobile telecommunication (IMT) 2000. ITU also made recommendations which included the architecture of the envisaged radio technology. The environment where IMT-2000 was expected to be used includes indoor, vehicles, satellites and pedestrian. Frequencies were allocated for use. However, most of the frequency bands allocated for IMT-2000 were already in use in many countries. Consequently, this gave rise to different technologies which undermined the original idea of common global system. To resolve these differences in technology, a family of 3G standards was adopted at the expense of the initial idea of common global communication system.

Third Generation technology refers to a generation of wireless systems with increased data rates and enhanced services. Both services and data rates are better and higher, respectively, than the previous 2G technology. 3G systems support email, web-access and multimedia applications. There three primary standards that comprise 3G technology. They include: W-CDMA, CDMA2000, and TD-CDMA.

The Universal Mobile Telecommunications System is the European proposal for IMT-2000 prepared by ETSI. The system supports the W-CDMA standard and is currently deployed in Europe. UMTS is based on the Global System for Mobile Communications (GSM). GSM has a very large market share of the global wireless market. And because of this reason, UMTS was chosen to represent the 3G technology in the context of heterogeneous wireless overlay networks
throughout this thesis work. The difference between UMTS and other 3G types notwithstanding, the observations and findings of this study can be applied to other 3G standard with minimal modifications. It is expected that the study would apply to a variety of heterogeneous overlay networks that supports handover across their boundaries and with other types of networks. In the sections that follow, UMTS will be presented in perspectives that support this study.

2.2 STANDARDIZATION OF UMTS.
As mentioned in section 2.1, 3G standard was adopted as a result of the differences in the technology implemented, by different countries, within the frequency band allocated for IMT-2000. Figure 2.0 illustrates the allocated frequencies and regions where they are used for the implementation of other telecommunication technology. 3G technology standardization was adopted by International Telecommunication Union so as to homogenize all the variants of 3G technologies that were being developed. The ITU defined a list of requirements for the 3G system.

Figure 2.0 IMT-2000 frequencies as used by different countries.

These requirements include the requirement to support:

- High data rates.
• Packet Switched and Circuit Switched data transfers.
• IP services.
• Efficient use of the allocated spectrum.
• High speech quality.
• Backward compatibility for smooth transition from 2G to 3G.

Accordingly, proposals were submitted to the ITU for standardization of 3G. The group, Third Generation Partnership Project (3GPP) is the group largely responsible for the standardization of UMTS based on evolution of 2G (GSM). On the other hand, Third Generation Partnership Project2 (3GPP2) focuses on the standardization of the sets of 3G standard based on earlier 2G CDMA technology-CDMA2000. The concept of single 3G standard gave rise to family of five 3G wireless standard. Of this five, the most widely accepted are CDMA2000 and W-CDMA (UMTS). The standardization process of UMTS is continuously updated by 3GPP. Release 10 is in progress as at the time of doing this study.

2.3 UMTS ARCHITECTURE

Consequent upon the failure to achieve a common global frequency band as proposed by ITU, a group of five radio access technology was standarized by the same body. Of these five groups, two became widely accepted. One of the widely accepted two is W-CDMA (UMTS) which has been selected for study in this work.

Five regions of the world had started using parts of the frequency allocated to IMT-2000 in 2G. For instance, Europe uses part of the 1885-2025 MHz in GSM and DECT. What was left of the remaining frequencies has been split into bands for Universal Terrestrial Radio Access-Frequency Division Duplex (UTRA-FDD, uplink: 1920-1980 MHz, downlink: 2110-2170 MHz) and UTRA-TDD (1900-1920 MHz and 2010-2025MHz). The frequency band, 2110-2200MHz was also allocated to IMT-2000. The technology behind UTRA-FDD and UTRA-TDD forms the basis of the UMTS.

UMTS architecture is based mainly on 2G technology. GSM has the greatest market share in cellular technology world over. Principally, the technology is digital and circuit switched. It is a narrow band system most suited for voice and limited data communication. Figure 2.1 shows a simplified, basic UMTS architecture consisting of three sections: The User Equipment (UE), the Radio Network Subsystem (RNS) and the Core Network (CN).

![Figure 2.1 UMTS Main components and reference architecture.](image-url)
The **UTRA** network (UTRAN) is responsible for cell level mobility and consist of the several Radio Network Subsystems (RNS). UTRAN is, in fact, another name for RNS. UMTS architecture has interfaces similar to GSM. For instance, the **Uu** interface is comparable to GSM Um interface. Similarly, the **Iu** interface is comparable to the **A** interface in GSM. In UMTS, the user equipment is connected via the Uu radio interface to the UTRAN. The UTRAN communicates with the core network (CN) via the **Iu** interface. UMTS is the proposal submitted by ETSI (European Telecommunications Standard Institutes) for IMT-2000. The radio access technology is also known as UMTS but is preferably called **Universal Terrestrial Radio Access (UTRA)**. This technology was conceived to evolve from 2G technology so as to save players in the GSM industry costs of designing and implementing new architectures. As already shown, UMTS network is a type of network know as Public Land Mobile Network (PLMN) with specific purpose of providing land mobile telecommunication services to the public. To do this, each section of UMTS is associated with a number of defined responsibilities. The core network contains functions for inter-system handover, gateways to other networks and data bases to store user profile, mobility management information, user access authorization, cell location, billing and accounting information. The core network transports user data, carries out signaling and switching functions as well.

In UMTS, the system architecture is subdivided into domains. These domains include user equipment domain which contains the **USIM** domain. The **USIM** domain in turn contains the SIM for UMTS. The end device belongs to the mobile equipment domain while the infrastructure domain consists of the access network domain. The access network domain contains the radio access network (RAN), and the core network domain. The user equipment domain is assigned to a single user. Its function is to grant access to UMTS services to users. The USIM domain contains the subscriber identification module (SIM) for UMTS. The SIM performs encryption and authentication of users, and stores all the necessary user-related data for UMTS. Mobile equipment domain is where the end device is categorized. Functions for radio transmission as well as user interfaces are found in the mobile equipment domain.

All users share the infrastructure domain. The domain offers UMTS services to all accepted users. The core network is further divided into three domains with specific tasks. This includes the serving network domain, the home network domain and the transit network domain. Figure 2.2 shows the UMTS domains and interfaces.

![Figure 2.2: UMTS domains and interfaces.](image-url)
2.4 UMTS RADIO INTERFACE.

The radio interface for UMTS is the Uu interface through which the user equipment connects to the UTRA network. UMTS radio interface operates in two modes namely, UTRA-FDD (W-CDMA) and UTRA-TDD (TD-CDMA). Both FDD and TDD are existing technology in 2G. Frequency Division Duplex (FDD) and Time Division Duplex (TDD) are technologies used to send data in two directions; usually between a mobile station and the base station. Uplink is the connection that allows a mobile station to send data to base station while downlink is the connection that allows a base station to send data to a mobile station.

Direct Sequence Code Division Multiple Access (DS-CDMA) technology is also used in UMTS. This technology is used to spread users signal by multiplying a stream of bits with a chipping sequence. A unique chipping sequence separates different users even when users transmit signals in the same frequency. This means that orthogonal codes (codes whose cross correlation is zero) must be used in order to separate different users.

A constant chipping rate of 3.84Mchips/s is used in UMTS. To achieve different user data rates, different spreading factors are used to multiply user data. A spreading factor is the number of chips used to multiply a bit of user’s data. Chips are simply short bursts of digital signal represented by 1(ones) and 0(zeros).

A sender’s signal is first spread using orthogonal spreading codes. This means that different user data stream is separated using orthogonal codes. In UMTS orthogonal variable spreading factor (OVSF) is used to separate user data. Orthogonal codes are generated by doubling a chipping sequence, say X. The sequence can have positive or negative sign. The effect of doubling a chipping sequence is that a bit is spread twice as much as before. To maintain orthogonality, one code is not allowed to be generated from another code.

2.5 UMTS ACCESS NETWORK (UTRAN).

The main UMTS access network is known as Universal Terrestrial Radio Access Network (UTRAN). It consists of Radio Network Subsystems (RNS). The nodes in the core networks connect to UTRAN via the Iu interface. UTRAN also provides the interface Uu, through which the mobile stations connect to it. It provides management function and related functionality.

Figure 2.3 shows the basic architecture of the UTRA network. Comparable to BSC control to BTS in GSM, RNC controls each radio network subsystems (RNS) in UMTS. Components in the RNS include node B, which are comparable to BTS in GSM. One or more antennas are
controlled by each Node B. Similarly, one mobile station can connect to one or more antennas which make up a radio cell. Each RNC is connected with the core network over the interface $I_u$ and also connects with the node B over the interface $I_{ub}$. The interface $I_{ur}$ has no counterpart in GSM. This interface connects two RNCs and is part of handover mechanism in UMTS.

The basic architecture shown in figure 2.3 is subject to changes as 3GPP proposes the 4G technology which is expected to be ALL-IP UMTS core network.

![Figure 2.3: Basic architecture of the UTRA network.](image)

2.6 UMTS CORE NETWORK

The UMTS core network consists of the circuit switched domain (CSD) and the packet switched domain (PSD). These two domains consist of GSM and GPRS components respectively. The
The circuit switched domain is concerned with the classical circuit switched services and signaling. In a circuit switched network, resources are reserved at connection setup. The GSM components used are the MSC, GMSC, and VLR. As can be seen from Figure 2.4, the interfaces connecting various domain or components are Iu, IuCS. The circuit switched domain connects to the radio network subsystem via a part of the Iu interface called IuCS.

GPRS components include the gateway GPRS support node (GGSN) which is the interworking unit between the GPRS network and external packet data network, the serving GPRS support node (SGSN) which supports the mobile station through the Gb interface. The SGSN and the GGSN connects to the radio network subsystem through the IuPS part of the Iu interface. As can be seen from the figure, the two domains connect to Equipment Identification Register (EIR) and the Home Location Register (HLR). Both databases are used for equipment identification and location management respectively.

Figure 2.4: UMTS core Network comprising 3G RNS and 2G BSS (Source: Network Communication)
2.7 UTRAN PROTOCOL STACK.

Figure 2.5 and 2.6 show the protocol stacks of the UMTS network. As shown above in Figure 2.4, the UMTS core network (CN) consists of two domains. One of the domains is the circuit switched domain (CSD) and the other is the packet switched domain (PSD). The circuit switched domain is concerned with resource reservation before information is exchanged between the network entities, whereas the packet switched domain is responsible for the transmission of user data. UMTS release 99 distinguished these two domains in UTRAN protocol stacks. The CSD is associated with 3G MSC (mobile switching center) whereas the PSD is associated with 3G GGSN (GPRS gateway support node).

According to information from 3GPP, the protocol stack between MN and the CN are divided into “Control plane” and “User plane” protocols. The “Control plane” protocols are responsible for exchange of control information between network entities. This is done before communication is established between them. In addition, resources are reserved. The “User plane” protocols are responsible for the transmission of user data. The two planes of the protocol are shown below.

![Diagram of UMTS User plane](source: 3GPP)
MS – SGSN for UMTS

Figure 2.6. UMTS Control plane: source 3GPP

The radio resource control protocol (RRC) has many functions. These functions handle control plane signaling of layer 3 between the MN and the radio access network. Because of it role in establishment of RRC connections, broadcasting of system information, mobility functions etc, it is an important protocol relevant in this research effort. RRC evaluates measurements relayed to it by lower layers and execute functions to support mobility procedures such as paging and handovers. Figure 2.7 shows the interactions between RRC with lower layers.

Figure 2.7. Interaction between RRC and lower layers in the C plane: source 3GPP.
2.8 UMTS/ WLAN INTEGRATION.

Research has been going on in industry and in the academia on how to improve communication networks and network services. Consequently, there has been tremendous improvements in the speed of networks and also the Qos experienced. The concept of UMTS and WLAN integration is aimed at improving the overall network performance and providing enhanced level of services. This is achieved by extending the relatively low data rate offered by UMTS network by relatively high data rate offered by WLAN network.

2.8.1 REQUIREMENT FOR INTEGRATION BETWEEN UMTS AND WLAN.

WLAN and UMTS Networks are different in many ways. For instance, WLAN operates in the license free frequency spectrum and offers higher data rate than UMTS. UMTS, on the other hand, has wider coverage area and, depending on speed and size of cell, support mobility more than WLAN. What this means is that the technology behind the two networks is different.

For UMTS and WLAN to cooperate in the process of serving communicating users, there has to be a common protocol binding the two technologies. IP is the common protocol between WLAN and UMTS. IP separates the upper layer (transport and application) from the heterogeneous lower layers where the two technologies are different. Consequently, IP is the defacto protocol for the integration of UMTS and WLAN. IP facilitates the use of Mobile IP, a protocol proposed by IETF to support transparency in the upper layer (application) during mobility. While carrying out feasibility study on interworking WLAN and 3G networks, 3GPP considered IP and mobile IP in the planning stages of the interworking process. These protocols are relevant in the process of handover between UMTS and WLAN networks.

2.8.2 INTEGRATION BENEFITS

Travelers and holiday makers are enjoying outdoor network connectivity in places such as arrival lounge of airports, relaxation spots in hotels and various open locations in campuses. This is made possible by WLAN hotspots in these locations. As mentioned earlier, WLAN provides higher data rate than UMTS which, however, has a wider coverage than WLAN. Because of the high data rate offered by WLAN, the network is used to extend UMTS network.

UMTS can, theoretically, offer data rates up to 2Mbps. However, in practice, the data rate is significantly below that. WLAN, on the other hand can, theoretically, offer data rate ranging from 11Mbps to 54Mbps. But, like WLAN, figures realized in practice are below the theoretically rated values. Table 2.1 shows UMTS data rates which depends on other factors such as mobility and coverage area.
Table 2.1, UMTS data rates, cell size and coverage area.

It has been demonstrated, eg 1st ROMANTIK project funded by European Union, that UMTS capacity is enhanced when it is extended by WLAN. Extending UMTS with WLAN increases, therefore, the number of connected users to a network. WLAN operates in free license spectrum. This is one of the benefits which contribute to the capacity enhancement of 3G networks such as UMTS.

Finally, WLAN is cheap to set up. It costs a lot less to set up a WLAN hotspot than to set up a single 3G base station.

2.9 CHAPTER SUMMARY.

UMTS is a 3G network that makes use of the combined architectures of GPRS and GSM. UMTS has been standardized by third generation partnership project. The standards are published in form of releases. The core network of UMTS consists of those network entities from GPRS and GSM responsible for switching, while the access network, UTRAN is responsible for radio access and all radio functionalities.
CHAPTER 3

OVERVIEW OF WIRELESS LOCAL AREA NETWORKS

3.0 INTRODUCTION

WLANs are networks consisting of Nodes and APs that exchange data through air interface. Operations mostly limited to unlicensed frequency spectrum commonly referred to as the ISM frequency band. There are various flavors in the class of 802.11 standards, as they are popularly known.

There are some factors responsible for the growing popularity of WLANs. Its ease of deployment, license free operations and low cost are chief amongst them. Again, network users find the mobility and Internet connectivity in WLANs quite attractive. Wire clusters in offices are rapidly disappearing as a result of progressive replacement of fixed-wire LANs by WLANs. However, WLANs are, at the moment, more commonly overlaid on fixed wired networks and there are security issues to contend with. WLAN networks can be deployed as ad hoc or infrastructure networks as discussed in this chapter.

3.1 IEEE 802.11 STANDARD AND VERSIONS

The Institute of Electrical and Electronics Engineering (IEEE) created the 802.11 standards in 1997 for WLAN. WLAN operated in the 5 Ghz and 2.4 GHz frequency spectrum. The band of frequencies thus defined for WLANs are sometimes referred to ISM (Industrial, Scientific and Medical) band of frequencies. IEEE 802.11 standard is the fundamental standard specified for WLAN. The standard defined some functions and technologies for WLAN. Amongst the functions and technologies defined are WLAN architecture, MAC layer functions and services, Basic Service Sets (BSS) etc. The 802.11 standards have advantage of high data rates, although the range of transmission is far less than the range obtained in cellular systems. However, WLAN is an important network in which mobility is highly supported. The IEEE 802.11 standards to be discussed briefly are IEEE 802.11b, IEEE 802.11a, IEEE802.11g, and IEEE802.11n.

3.1.1 IEEE 802.11b

IEEE 802.11b was, at the time of its introduction to the market, intended to solve various implementation of the original IEEE 802.11 standard. Then, various companies had implemented proprietary solutions with speeds up to 11 Mbits/s. Consequently, IEEE 802.11b was introduced, in the same 2.4GHz, as a common standard to avoid market segmentation. The standard described new PHY layer and, as it turned out, became the most successful version of the IEEE 802.11.

The standard described a new PHY layer. The MAC schemes and management procedures in the original IEEE 802.11 were used. Various speeds, eg 11, 5.5,2, 1 Mbits/s are possible depending
on the distance of separation between sender and receiver; and the interference experienced at the time of transmission. Different coding schemes were used for different speed.

3.1.2 IEEE 802.11a

Despite the fact that the standards are named in the other they were established, IEEE 802.11b was more popular than IEEE 802.11a. IEEE 802.11b hit the market before IEEE 802.11a. 802.11b gained more popularity because it was more cost effective and do not require license unlike the 802.11a counterpart. Furthermore, close to the Ethernet network in bandwidth, IEEE 802.11b supports bandwidth up to 11 Mbits/s.

However, 802.11a supports bandwidth up to 54 Mbits/s and operates in the licensed frequency of 5 GHz. Unlike the 802.11b which is more popular to home users, 802.11a is used more in business networks. To reduce interference, 802.11a uses the orthogonal frequency division multiplexing (OFDM). While 802.11b operates at 2.4 MHz, 802.11a operates at 5GHz which. Consequently, 802.11a covers shorter range than 802.11b given the difference in the operational frequency. The difference in operational frequency is the reason why the two variants of 802.11 are not compatible.

3.2 HARDWARE ENTITIES OF WIRELESS LOCAL AREA NETWORKS.

Wireless Local Area Networks consists of the communication technology standardized by IEEE. Amongst these standards are the 802.11a, 802.11b, 802.11g etc standards. To set WLAN networks up, certain hardware must be assembled and connected together. These hardware include the AP, the computer nodes, the network interface cards( NIC.). Figure 3.1 shows some of these components.

![A LAPTOP (NODE) PCMCIA CARD WIRELESS ACCESS POINT](image)

**Figure 3.1: SOME HARDWARE COMPONENTS IN A WLAN**

3.3 TRANSMISSION TECHNIQUES IN WLANs.

Wireless LAN operations are in various transmission techniques. These techniques include the following:

- **Infra Red LANs**: The nodes in Infra Red LANs communicate with one another via diffuse light reflected off walls, ceiling etc. A transmission technique that requires the transmitter to be in direct line of sight of the receiver is known as LOS transmission. Infra Red LANs are license free LAN networks and are used mostly indoors
- **Narrowband Microwave**: Narrowband Microwave LANs operate at Microwave frequencies in the region of 0.3 GHz to 300GHz. Federal Communication Commission or any communication regulation body in any country, issue licenses for some of the frequency range within the Narrowband Microwave LANs. Others operate at the Industrial, Scientific and Medical bands which is license free.

- **Spread Spectrum**: The mode of operations of spread spectrum LANs involves Frequency Hopping and Direct Sequence Modulation techniques. Both Frequency Hopping technique and Direct Sequence Modulation techniques enhance the security issues in WLAN communication.

Many of the transmission technologies in WLAN fall within the ISM frequency spectrum. ISM (Industrial, Scientific and Medical Band) frequency spectrum is a license free band of frequencies within which Infra Red and Spread Spectrum wireless technologies operate. Below is a table showing the frequency allocation in the ISM band.

<table>
<thead>
<tr>
<th>INDUSTRIAL, SCIENTIFIC AND MEDICAL BAND FREQUENCY SPECTRUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Frequency Limit</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>902 MHz</td>
</tr>
<tr>
<td>2400MHz</td>
</tr>
<tr>
<td>5.725MHz</td>
</tr>
</tbody>
</table>

**Table 3.1 ISM frequency spectrum**

3.4 WLAN NETWORKS AND SYSTEM ARCHITECTURE

The IEEE 802.11 standard specifies sets of standards that must be met by various WLANs products. Medium access and physical layers are specified to meet the demands of WLANs technology and architecture.

The 802.x standards have two basic system architectures and Networks. They are the INFRASTRUCTURE based and AD-HOC based system architecture. Infrastructure and ad-hoc networks are based on these two basic system architectures.
3.4.1 WLAN INFRASTRUCTURE NETWORKS AND SYSTEM ARCHITECTURE.

Infrastructure networks are probably the more commonly deployed WLAN networks. They provide access to other networks, forward calls to other WLANs, control medium access and provide certain management function via the AP (Access Point). In infrastructure Networks, communication is between the AP and the wireless node.

The presence of AP as a central communication point in WLANs, and the presence of other network hardware which may provide links to other networks (wireless or wired) confer on these networks the "Infrastructure" identity. These Networks, in fact, consist of two parts: Infrastructure part and wireless part. A simple diagram illustrates below, a WLAN infrastructure Networks.

Fig 3.1 A WLAN INFRASTRUCTURE NETWORK.(SOURCE: EUSSO)

Fig 3.2 WLAN Infrastructure System Architecture
Underlying the Infrastructure network is the INFRASTRUCTURE SYSTEM ARCHITECTURE. A basic WLAN network consists of an AP, one or two wireless nodes, of all which must be within the same radio coverage. This set is known as basic service set (BSS). WLAN coverage could be extended (and usually is) by linking one or more BSS. The result is what is known as extended service set (ESS). An extended WLAN coverage gives rise to new identifier with which different ESS are known. These ESSIDs, as they are called, are simply names given to different extended service sets. Without the different identities, participation in a WLAN by different ESS could be difficult. Fig 3.2 illustrates infrastructure architecture. As shown in figure3.2, the wireless network components (the node) in BSS1 can communicate with the AP because they are within the same radio coverage. However, the node in BSS2 is outside the radio range of the AP in BSS1 and therefore could not communicate with it. Communication is therefore by linking the two BSSs through a distribution system. The distribution system connects all the APs in different BSS forming a larger single WLAN network.

Within an extended service set, communication between a mobile station (node) and an AP is possible through association of the AP with the mobile station. Because ESS is made up of a number of BSSs, each mobile station from one BSS is allowed to associate with an AP that belongs to another BSS, a functionality known as roaming. Many functions within an infrastructure WLAN are performed by the AP.

3.4.2 WLAN AD-HOC NETWORKS AND SYSTEM ARCHITECTURE

IEEE 802.11 WLAN networks are not only infrastructure based, ad-hoc networks are also built. Ad-hoc networks allow direct communication between stations. Because there is no central point such as AP through which communication between nodes is facilitated, this network can be arranged to meet a particular network need, hence the name ad-hoc.

The nodes are not dependent on the AP. Consequently, the group of nodes operating in the same radio frequency forms an independent basic service set (IBSS). Figure 3.3 illustrates an ad-hoc network and the corresponding architecture.

![Figure 3.3 WLAN ad-hoc network and architecture.](image-url)
3.5 WLAN PROTOCOL ARCHITECTURE.

This thesis discusses handover between WLAN and UMTS. A brief discussion of WLAN protocol architecture will no doubt help the reader to understand the mechanisms of media access in WLAN. Before any handover occurs between WLAN and UMTS, the mechanism to ensure that connectivity exists in the WLAN must be in place.

WLAN is somewhat slow version of wired LAN. Consequently, applications in WLAN connected to switched LAN only notices the difference in connection speed and lower bandwidth. Higher layer protocols such TCP, IP for wired LAN appear the same for wireless nodes. However, the logical link layer takes care of the differences in the media access control layer needed for the different media. Figure 3.4 shows the layers in WLAN. As can be seen, these layers: APPLICATION, TCP, IP are found too in the 802.x Ethernet LAN.

The physical layer is subdivided, as would be seen in figure 3.5, into the physical layer convergence protocol (PLCP) and the physical medium dependent sub layer (PMD). The MAC layer, primarily, is responsible for medium access, fragmentation of user data, and encryption. On the other hands, the PLCP sub-layer provides carrier sense signal known as Clear Channel Assessment (CCA). It also provides a common physical service access point (SAP) independent of the transmission technology. The PMD is responsible for modulation, encoding and decoding of signals.

![Figure 3.4 IEEE 802.11 protocol architecture and bridging.](image1)

![Figure 3.5 IEEE 802.11 protocol architecture with management.](image2)
IEEE 802.11 specifies management layers. As can be seen from figure 3.5, the MAC management is above the PHY management layer and supports the association and re-association of a station to an access point. In addition, it supports roaming between different access points. Roaming is the movement of a mobile user from one access point to the other without interrupted services. This is what "handoff" or "handover" seeks to achieve, be it horizontal handover or vertical handover.

The PHY management include Channel tuning and PHY management information base (MIB) maintenance. The Station management works with both management layers and is responsible for additional higher layer functions such as control of bridging and interaction with the distribution system in the case of access point.

3.5.1 PHYSICAL LAYER.

Three different physical layers are supported by IEEE 802.11. However, only two were based on radio transmission while the other is based on Infra Red. As mentioned earlier, a clear channel assessment signal is provided by the three types of PHY layers. The signal is important for the media controlling role of the MAC layer as it is used to determine if the medium is idle. Again, the PHY layer provides service access point, offering up to 2 Mbits/s transfer rate to the MAC layer. The service access point is useful to transmission technologies such as Frequency Hopping Spread Spectrum (FHSS) and Direct Sequence Spread Spectrum (DSSS).

Frequency Hopping Spread Spectrum transmission technique allows multiple networks to coexist in the same area. To achieve this, different hopping sequences are used by different networks in the same 2.4 GHz ISM band. Specific channel is accessed by using a pseudo-random hopping pattern.

DSSS, a radio transmission technology, uses spread spectrum technology to allow many networks to coexist in the same area. In this case, however, codes, rather than frequencies, are used to separate the networks. Usually, for IEEE 802.11, an 11-chip Barker sequence is used to achieve the spreading. “Spreading” is the spreading of the bandwidth needed to transmit user data. Spreading reduces the effect of narrow band interference as well as the power in the user signal. To achieve the spread, generally, a narrow band signal is converted into broadband signal such that the power carried by each frequency in narrow band signal is distributed (spread) over a broad band signal. This has the consequence of reducing the power density of the original narrow band signal in the spread signal without losing the data in the original narrow band signal. It is possible to achieve a low power signal, as low as the background noise, in such a way that detection of the original signal in the broadband signal is difficult. In fact, this has been one important advantage of the spread spectrum technique.

3.5.2 THE MAC LAYER.

MAC layer is important in many ways. The basic function of the MAC layer is medium access control, although support for roaming, authentication and power conservation is also offered. In terms of services provided by the MAC layer, the asynchronous data service is mandatory while the time bounded service is optional. Both services, with access point coordinating medium
access, can be offered using infrastructure-based network. However, the asynchronous service, as defined in 802.11, is offered in ad-hoc network mode only.

Furthermore, asynchronous service is offered by Distributed coordination function (DCF) while point coordination function (PCF) offers both asynchronous and time-bound service with an Access Point coordinating medium access as already mentioned. DCF is a name referring to the first two basic medium access mechanism defined in IEEE 802.11. These basic medium access mechanisms are (a) Carrier sense multiple access with collision avoidance (CSMA/CA), (b) an optional method that avoids the hidden terminal problem, (c) and the contention-free polling method for time bounded service. The contention-free polling method is generally referred to as the point coordination function (PCF) as mentioned above. It requires an access point in order to provide both asynchronous and time bounded service. The three basic MAC mechanisms are generally known as the distributed foundation wireless medium access control.

3.5.3 MAC MANAGEMENT

MAC management concerns itself with the management of the mobile stations in terms of controlling all functions relating to the integration of the wireless station into Basic Service Set, etc. Some of the functional groups identified in the MAC management are Synchronization, Power management, Roaming, Management Information Base. Because roaming cannot be discussed without mentioning mobility and changing of Access Point, much like handover (vertical or horizontal) is discussed in conjunction with mobility and changing of cells, it is going to be discussed in this subsection.

3.5.3.1 ROAMING.

Access to WLANs is often limited by the transmission range. Access Points are distributed in various locations in a building requiring access to WLAN to overcome the limitation posed by a single Access Point's inability to cover large area in the building. A mobile station moving from one Access Point to another, therefore, needs to retain connectivity to the WLAN without service interruption. To do this, roaming comes in handy to facilitate the association of one mobile station from one Access Point to the other.

During roaming, a station can determine that the current link quality between it and the Access Point is poor. Consequently, the station initiates Scanning for another Access Point. When scanning, the station actively searches for another Basic Service Set (BSS) in which resides another Access Point. In infrastructure networks, scanning detects a stronger signal from another Access Point. However, for ad-hoc network, new BSS can be set up.

Scanning can be performed in a single or multiple channels as specified in IEEE 802.11. Furthermore, scanning can be passive or active. When the mobile station listens in the medium in order to find other networks, the station is performing passive scanning. The listening station receives a Beacon when an Access Point is synchronizing with the station in the network. On the other hand, when the station sends a probe on each channel and waits for a response, the station is said to perform active scanning. The information contained in beacon and probe responses are necessary for joining with the BSS.
Signal strength is usually the criteria used by station in selecting another AP. Therefore, the station selects, from amongst the available APs, the AP with the strongest signal. Having selected the AP, the station sends an association request to the selected AP. The selected AP responds with association response. A successful response means that the station is accepted, or linked to the new AP and, as a result, has roamed from one Access Point to the other; otherwise, the station has to continue scanning for new Access Point.

Once accepted in the BSS, the Access point responsible for accepting the new station informs the distribution system (DS) of the new station. The Distribution System, having been notified of the new station, then updates its database. The database contains information about the current location of the mobile station. It should be recalled that a number of BSSs together forms ESS. Roaming occurs in an ESS. The Distribution System database contains information needed for forwarding frames between different BSSs. Furthermore, the DS informs the old AP in charge of the roaming station that the station is no longer under its control. IEEE 802.11f (Inter Access Point Protocol, IAPP) is meant to be all vendor protocol used for roaming. It informs the old AP of the location of the new AP, load balances between Access Points, and generate key for security algorithms based on IEEE 802.1x.

### 3.6 SUMMARY

This chapter concentrated on WLAN technology. Types of WLAN networks are discussed. WLAN adhoc networks are deployed without any central AP serving the MNs while infrastructure WLAN networks have a serving central AP through which MN’s get access to the network.

The transmission techniques in WLAN include infra red, Narrow Band Microwave and Spread Spectrum.
CHAPTER 4

HANDOVER IN UMTS AND WLAN

4.0 INTRODUCTION

Handover (handoff is used in some literature) in any network is the processes involved in disconnecting a mobile node from its current communication link and re-establishing the link again in the same network, or different networks. This process usually involves handing the MN over to other network entities, within or outside the current network, to continue with the ongoing communication. After reading this chapter, the reader will appreciate the process of handover generally. But in particular, the reader will learn the process of handover in the same network and between two different networks. Handover in the same network is known as horizontal handover. Similarly, handover between two different networks is known as vertical handover. 3GPP specifications describe mobility management issues and location of MN within network. This chapter is going to briefly discuss mobility as it relates to network and the concept of locating MN in a network. However, more treatment will be given on vertical handover in later chapters since this research work focuses more on vertical handover. The term MN, Mobile Node, Mobile Station refers to the same network element.

4.1 HANDOVER, A BRIEF HISTORY.

Mobile communication has come a long way from first generation up to fourth generation currently on trial deployment in some parts of Europe, Canada and America. The first generation mobile radio systems were deployed to cover large distances (up to 150km radius). Each cell was supported by a single transmitter. Then, this was a limitation because only few channels were available. Consequently, a limited number of users were supported. In addition to channel limitation, transmitters and mobile stations during communication, transmit at high power due to the large area of coverage by each cell. The large area of coverage means that a long distance of separation exists between a mobile station and a transmitter.

To overcome these limitations, the concept of frequency reuse was introduced. The overall effect of frequency reuse was to enhance cell capacity and reduce transmission power of both the
mobile station and transmitter at the base station. Within relatively small distance of each other, the same frequency can be used in two cells covering small distance. Figure 4.1 shows a cell without frequency reuse and the same cell with frequency reuse.

Figure 4.1 B shows cells with different and similar colors. Same colored cells use the same frequency.

It can be seen from the two diagrams that both single big cell and multiple cell clusters cover the same area. The hexagon is used to denote a cell in figure 4.1B. Within a particular coverage area, the number of frequencies available for use depends on the number of cell clusters in that coverage area.

The concept of frequency reuse created the need for a mechanism to switch frequencies used by mobile station within a particular area covered by a number of cells. The switching simply means handing over a communicating mobile station from one particular cell to the other. However, the process of handover involves other technical issues, part of which this research work intends to address. Subsection that follows discusses common types of handovers between UMTS and WLAN networks

4.1.1 TYPES OF HANDOVERS

As telecommunication system is increasing in numbers, so is the number of networks that make use of different telecommunication technologies. In the early days of cellular networks, network entities were of the same technology. However, this is not the same in today’s network. Consequently, handovers in networks are categorized depending on a number of factors. Handover occurs within a network using the same technology. Similarly, handover occurs between networks interworked with different technologies, etc. Below is some of the categorization.

CATEGORY A: TECHNOLOGY INVOLVED IN THE HANDOVER

a. **Horizontal Handover**: When a handover occurs in a cell that supports the same technology, it is referred to as a horizontal handover.

b. **Vertical Handover**: On the other hand, when a handover occurs between base stations supporting different technologies, the term Vertical handover is used. A handover between 3G UMTS base stations and WLAN is a typical example vertical handover. Vertical handover is further classified into downward handover and upward handover. A downward handover occurs when a mobile station is handed over from a network with large cell but low data rate to a network with small cell but with a high data rate. The opposite is true for upward handover.
CATEGORY B: ADMINISTRATIVE DOMAINS INVOLVED IN THE HANOVER

A Network operator has administrative authority over the systems used in the network. The network usually is made up of a collection of end systems, intermediate systems, subnets etc. This form of organization of network systems over which an operator exercises authority is known as “Administrative Domain”. Components within an administrative domain interact with mutual trust while they interact with suspicion with components in other administrative domains. Therefore, handovers in Administrative domains are classified as shown below:

a. **Intra-Administrative Domain Handover:** This type of handover involves MN that switches links between base stations that support the same or different technology. The base stations are managed by the same administrative authority or domain.

b. **Inter-Administrative Domain Handover:** In this type of handover, the mobile station switches between base stations supporting the same or different technologies. The base stations are managed by different administrative authority or domain.

CATEGORY C: NUMBER OF CONNECTIONS INVOLVED IN THE HANOVER.

The handover within cellular networks are also classified as follows:

a. **Hard Handover:** Within a cellular network, a mobile station connects with a base station. However, during handover, the mobile station is disconnected from that base station and switched to another base station. This process is called “hard handover” because the mobile station lost its connection to the current base station before establishing new connection with another base station. Another term for this kind of handover is “break before make”. Hard handover may be seamless or non-seamless. During seamless handover, the user of the mobile station does not notice any interruption of service in form of dropped calls, terminated session or any form of delay.

b. **Soft Handover:** A mobile station moving within a cellular network can be switched to another base station without first terminating the connection with the current base station. This type of handover is called “soft handover”. The mobile station is connected to at least one base station as it moves within the cellular network. Another term used to refer to this kind of handover is “make before break”.

c. **Softer Handover:** This type of handover occurs when a mobile station connected to a base station with multiple radio links switches from one link to the other. A base station in a cellular network may serve many sectors belonging to the same cell. Figure 4.2 shows hard handover, soft handover and softer handover.
Fig 4.2, Hard handover, Soft handover and Softer handover

**CATEGORY D: FREQUENCY INVOLVED IN THE HANDOVER.**

a. Intra-Frequency Handover: This type of handover occurs when mobile station switches between base stations that operate at the same frequency. This means that the mobile station is not connected to a different frequency after the handover. In 3G systems, CDMA (code division multiple access) scheme allows users to transmit at the same frequency and time. Code space is used to separate different users.

b. Inter-Frequency Handover: Inter frequency handover is used to describe a handover that involves mobile station switching between base stations that operate at different frequencies. This type of handover type is more common in 2G systems where access schemes are TDMA (time division multiple access) and FDMA (frequency division multiple access). However, in 3G systems that use CDMA media access schemes with TDD (Time Division Duplex), this type of handover is also used. Hard handover is a common feature of inter-frequency handover.

**CATEGORY F: LAYER INVOLVED IN THE HANDOVER.**

Handover is also described according to the OSI (Open System Interconnects) layer in which the handover takes place. For instance, handover can occur at layer two or layer 3 of the OSI model.

a. Layer 2(L2) Handover: In the case of WLAN, a layer two handover will involve two Access Points between which a mobile station switches connectivity. The two Access
Points must belong to the same ESS (Extended Service Set). L2 handover occurs between layer two network devices.

b. Layer3 (L3) Handover: This is used to describe a handover in which a mobile station switches between two Access Points belonging to different ESS in the case of WLAN. In order words, the handover is between two different networks. For this type of handover to occur, handover signaling is required to manage the routing data to the mobile station’s new location. Figure 4.3 illustrates the two types of handovers.

![Layer 2 and Layer 3 handover](image)

**Figure 4.3, Layer 2 and Layer 3 handover**

**CATEGORY E: NETWORK ELEMENTS INVOLVED IN THE HANDOVER.**

Handover in UMTS network can depend on the hierarchy of the network elements involved in the handover. Some of the handover that occur in UMTS network is as shown below.

a. **Intra-NodeB Handover:** UMTS network has several network elements such Node Bs, radio network controllers (RNCs), Mobile Switching Centers (MSCs). An intra Node B handover occurs when a mobile station moves between two antennas controlled by the same Node B. This type of handover is a typical example of soft handover mentioned earlier.

b. **Inter-NodeB Handover:** This handover type is used to describe a handover that switches mobile station from one Node B to another Node B. This type of handover is
also referred to as Intra-RNC handover if the two different NodeBs are controlled by the same RNC.

c. **Inter-RNC Handover:** Handover can occur also between two Node Bs controlled by two different RNC. In this case, the mobile station moves from one Node B controlled by an RNC to another Node B controlled by different RNC.

### 4.2 MOBILE IP.

Solicitation Handover process more or less involves mobility. Consequently, handover cannot be discussed without presenting basic knowledge about mobility as it relates to network. This section attempts to present to the reader mobility protocol proposed as a solution to some problems encountered as mobile station moves from one point to the other in a network.

Internet protocol addressing, by its nature, is static. What this means is that a mobile station with a particular IP address retains its address, and can only be reached with this address when it moves from one network to the other. This is a problem when a mobile station moves from its home network to a foreign network.

Mobile IP was proposed as solution to the problem associated with static IP of a mobile station that moves from one network to the other. Interestingly, mobile IP is based on the IP protocol. It is meant to mask the mobility from upper layer of the protocol stack in order to offer seamless mobility. To be able to carry out this function, the protocol establishes a tunnel between the mobile station’s home network and the current foreign network visited by the mobile station. Actually, the tunnel is nothing but the process by which a node known as the home agent (HA) communicates with another node known as the foreign agent (FA). The two nodes form the two end points of the tunnel. The HA is responsible for maintaining the knowledge of the current location of the MN (mobile node, mobile station) and forwarding of incoming packets to it. The FA, in turn, is responsible for communicating with the HA to ensure the delivery of packets to the mobile station during the time it stays in a foreign network.

The protocol executes in three steps namely: agent discovery, registration, and tunneling. As soon as a mobile station visits a foreign network, the agent discovery process runs to inform the mobile station of existing foreign agent serving the visited network. FA can be obtained agent advertisement or by by the mobile station. After discovering the FA, the HA of the mobile station is notified of the new care of address (CoA) assigned to the mobile station in the visited network. This information is important for binding the IP address of the mobile station by the newly acquired CoA. The process of notification and binding of IP address is known as Registration. Finally, tunneling involves the encapsulation by HA, the packets received from the correspondence node (CN) at the user’s home network. In addition, the packets are forwarded to the CoA of the mobile station. The forwarded packets are extracted at the CoA. CN can be one
or more nodes that communicate with the mobile station. It is important to note that CN may be a mobile or fixed node. Also, CoA may be the IP address of the FA, or it may be a new address specifically assigned to the mobile station in the foreign network. Figure 4.4 illustrate the idea behind mobile IP.

Figure 4.4, Mobile IP example network.

4.3 POSITIONING OVERVIEW.

This research effort focuses on location or position aware handover across heterogeneous network. As probably would have been obvious based on previous explanation, UMTS and WLAN have been adopted as the heterogeneous network of interest. Within this network, therefore, the handover approach proposed is based on availability of location based information and the positioning of mobile station in the network.

A number of positioning methods exists for 2G, 3G and other wireless technologies. However, this section will discuss the positioning method chosen by 3GPP to be integrated into UTRAN. 3GPP has been standardizing positioning method in an effort to define the concept of location services (LCS) and develop an architecture and configuration to meet LCS needs. This includes the role of positioning functions within the UTRAN.
4.3.1 LOCATION SERVICES (LCS)

It is well known that services provided by network add flavor to the network user’s experience. This is also true, in particular, for location based services. 3GPP was prompted to standardize location based services for UTRAN as a result of the increased demand for such services. Some LCS application requires the subscription to a provider while others do. An example of LCS that does not require subscription to the provider is Emergency Positioning situations. Four types of location service clients were defined by 3GPP standard. They include:

a. **Emergency Service Client:** This service is used when the user places an emergency call.

b. **Lawful Intercept Client:** This is used for legal intercepts, depending on local regulations.

c. **PLMN Operator Client:** Used to improve network performance. It is also used to collect statistics for network optimization and enhancements.

d. **Value added Services Clients:** These are application running in mobile node. Good examples are weather application, navigation application, traffic update application etc.

The vertical handover approach considered in this research effort falls under PLMN Operator Client LCS client type. Consequently, the overview that follows focuses on the standard that supports this type of client.

3GPP standard distinguished four attributes that impacts on LCS. These four attributes include:

a. **Accuracy**

b. **Coverage**

c. **Transaction –Rate**

d. **Privacy.**

The definitions of these attributes are given here with respect to LCS. Accuracy is the closeness of the estimated location to the actual location. Coverage is the area where sufficient Qos is observed at the mobile station. Privacy is the confidentiality level of the mobile station’s location information. Transaction –Rate is the frequency of exchanging control message.

4.3.2 POSITIONING FUNCTIONS

According to the standard, LCS functions are categorized into groups. These include Internal Client group, UTRAN System Handling group, and Position group. Position group include four
functions designed to operate in the UTRAN or in some cases the mobile station. The functions are contained in Release9 of 3GPP specification document and are summarized as follows:

1. **U-PRCF**: This is the UTRAN Position Radio Coordination Function. The function manages the positioning task of a mobile station by coordinating resources (e.g. radio or GPS), communicating with other positioning functions, and handling positioning failure and recovery.

2. **U-PSMF**: UTRAN Position Signal Measurement Function is responsible for collecting radio and, at other times, satellite measurement to be used in positioning calculation functions. The positioning method used will determine the type of measurements collected. U-PSMF may be located at the mobile station, within Node B, or in a stand-alone location measuring unit node.

3. **U-PRRM**: The UTRAN Position Radio Resource Management monitors the effect of LCS related operation on performance of the network’s radio resources. It also controls the frequency of measurement collection, signaling, and the acceptance or rejection of LCS requests according to their potential impact on overall performance.

4. **U-PCF**: The UTRAN Position Calculation Function is responsible for the execution of the algorithm that took the previously obtained measurement as parameters and calculates the position of a target mobile station along with the achieved accuracy level.

**4.3.3 UTRAN LCS ARCHITECTURE**

Figure 4.5 shows a general arrangement of LCS in UMTS. LCS specific entities are required to be added in PLMN in order to support LCS services. PLMN is Public Land Mobile Network which is a term used to refer to infrastructure for mobile networks. Universal Terrestrial Radio Access Network (UTRAN) supporting LCS has Serving Mobile Location Center (SMLC) which can be implemented in the UTRAN within the Serving RNC (SRNC) see figure 4.6. It can also be implemented as a Stand Alone SMLC (SAS). The SAS can provide assistance data to a positioning unit eg GPS unit to enhance the performance of the positioning functions. In addition, the SAS can actually act as a location calculation server if the SMLC functionality in the SRNC is not configured to calculate the final location estimate. The Location Measurement Unit (LMU) can either be implemented as a function of Node B or as Stand Alone Entity. LMU is responsible for taking measurements and providing them to the RNC periodically or upon request. These measurements can be used to provide assistance to a positioning function to the UTRAN.
Figure 4.5, General arrangement of LCS in UMTS: source 3GPP

Figure 4.6, Clearer view of entities that support LCS in UTRAN
4.3.4 POSITIONING METHOD

In earlier releases of 3GPP as well as recent ones such as Release 9, a number of positioning methods have been considered. These methods provide location estimates to LCS clients. In addition, they use different approaches to get the information. Therefore, they exhibit certain attributes that make them suitable for specific environments or specific positioning tasks. It has been suggested in some publication that a hybrid of these method will enhance positioning accuracy in different environments in addition to compensating for each method’s short coming. Some of these methods include Network Assisted GPS method, Angle of Arrival (AOA), Observed Time of Arrival (OTOA), Observed Time Difference of Arrival (OTDOA) etc.

The two promising method for obtaining the position of mobile stations in 3G systems is discussed below.

4.3.4.1 Assisted GPS (A-GPS)

GPS, global positioning system, is the brain child of U.S department of defense (DoD). Back then-in the early 1970-GPS was used by the military as satellite navigation system. Much later, its use was extended to civilians. Being a global positioning system, it provides information to a target no matter the target’s location in the world. The standalone GPS has been a useful tool in both civilian and military environments. However, to use GPS in mobile stations required certain modifications. The reason is that mobile stations transmit in low power. In addition, due to their compact size, mobile station can only accommodate small modules. Furthermore, a number of emerging location-based applications require enhanced GPS performance. This includes faster start up time, which is normally hindered by long acquisition time. Signal detection capability of GPS also needed enhancement to facilitate detection of GPS signals in an environment where signals are weak due to obstructions.

A-GPS, shown in figure 4.7, was found to be the most accurate of all the standardized positioning method in most outdoor and unobstructed areas. It consists of reference receivers that have clear view of the sky and the constellation of GPS-satellite visible to the mobile station. They can therefore provide the mobile GPS receiver with assistance information for the measurement or calculation phase. The information can be provided through a broadcast or as a direct response to an update request from a mobile station over point to point connection.

A-GPS method is base on the concept of Differential GPS (DGPS), which is claimed to improve the location accuracy of the traditional 20m -3.5m. In some literatures, the accuracy of A-GPS has been claimed to be from 20m to 1m. In addition, a few research efforts on location –aware handover which demand high accuracy have been conducted based on acceptable assumption. This assumption was base on the notion that the increasing public demand for location-based
services and the emerging related-applications provide strong assurances that accurate positioning will become available in the near future.

Figure 4.7, A-GPS positioning method in UTRAN

4.3.4.2 UTDOA

Another method used in locating a Mobile Station is known as the Uplink Time Difference Of Arrival (UTDOA). Its approach consists of locating a Mobile Station by comparing the time it takes a radio signal to reach several LMUs after leaving the Mobile Station. Unlike GPS which needs improvements so as to be able to be integrated with Mobile Station handset, UTDOA can be used with legacy systems without any modification. It does not require additional resources or add any noise to the channels during active positioning since it relies on the power of the signals transmitted by the Mobile Station during any uplink signal.

This method has been acclaimed to be suitable for UMTS based on the wide bandwidth and high data rate associated with UMTS. It is believed that this method will allow a greater number of LMUs to participate in the positioning process and thereby provide better accuracy. Consequently, 3GPP has standardized this method for use in UMTS.
UTDOA has been found to perform well in indoor and urban areas as well as suburban environments. However, its performance is merely adequate in rural areas where the number of LMUs is lower and thus the accuracy level is not as high. Because UTDOA is considered suitable and UMTS and in fact has been standardized as at Release9 of 3GPP, it is therefore considered for use in the implementation of approach proposed in this research effort. Furthermore, its suitable for use in indoor and urban areas works well with WLAN which is also used in these environments as well.

4.4 SUMMARY

In this chapter, the stages involved in handover process are discussed. With UMTS and WLAN in mind, upward and downward handover were discussed. Handovers causes delay of some sorts. In particular, vertical handovers in heterogeneous networks are sources of delay during communication.

Positioning in UMTS network was discussed also. There is a number of positioning methods as described by 3GPP document. However, two positioning methods relevant to this research effort were discussed. The positioning overview section is aimed at acquainting the reader with the components involved in locating Mobile Station in UMTS network.
CHAPTER 5

MOTIVATION

5.0 INTRODUCTION

Demand for wireless access has increased tremendously. The reason is not far-fetched. Wireless access provides flexibility which wired access does not provide. Consequently, demand for enhanced wireless access is on the rise. In addition, emerging developments in a variety of wireless data/multimedia services have prompted current efforts to deliver such services over various types of network and technologies. Each of these technologies has its unique set of characteristics with respect to bandwidth, QoS, capacity etc. In both adjacent and overlapping network, they can provide a user with a range of level of services. Consequently, internetworking these heterogeneous wireless overlay networks to take advantage of the emerging services has received a tremendous progress.

It can be recalled that the primary motivation for integrating a heterogeneous network such as WLAN and UMTS is because WLAN offers higher data rate than 3G/UMTS. And for this reason, UMTS get added capacity when WLAN hotspots are used to extend it. This no doubt makes downward handover a desirable process whenever a WLAN is available. However, in some situations, switching from UMTS to WLAN may not be carried out efficiently and as such, would result to performance degradation. This chapter discusses such situations and provides the motivation for developing a solution that would predict the occurrence of these events and make the handover decision accordingly.

Again, handover between heterogeneous wireless networks occur with some degrees of latencies which cause significant delays that interrupt data transfer. In this chapter also, detailed description of the handover procedure will be provided, and the cause of their associated latencies identified. This is done to provide insight on the significant impact such handover will have on performance level. The chapter will proceed to discuss problems that motivated this research effort, and conclude with brief overview of other research effort.

5.1 HANDOVER PROCEDURES

Inter-working several small cells to provide wireless connectivity has several benefits. Amongst the benefits are low transmission power, aggregate bandwidth, and efficient method for locating mobile node during calls. However, depending on the size of network and the type of user mobility (either vehicular or pedestrian), frequent handover is usually experienced by a mobile node traversing the cells.
Early in the life of cellular networks, handover was limited to base stations that supported the same technology and, more often than not, belonged to the same administrative domain. This approach, however, has changed remarkably in this 21st century, given the emergence of new technologies in large numbers. Consequently, many smaller cells interworked to provide wireless connectivity support different technologies. For this reason, handover between these cells and across different technologies has become more complex. These handover complexities contribute a lot to the delay/latency associated with the handover process.

Delay/latency caused by handover process has prompted lots of research in seamless mobility. Seamless mobility is a concept that describes how the user moving from one point in a network to the other experiences the effect of handover during a call session, or when using real time application. Seamless mobility accommodates mobility across heterogeneous networks. To achieve seamless mobility, some of these requirements are essential as:

- Automatic handover from one base station to the other
- Transparent network transition through a common protocol such as IP
- Application session persistence during handover procedures and temporary lower layer disconnection.
- Minimal handover related delays and packet loss.

Seamless mobility is consequent upon seamless handover from one base station to the other. A seamless handover is considered a “fast handover” and “smooth handover”. The former ensures no delay/latency is observed by the user during handover while the later ensures minimal or no packet loss. Ways to reduce packet loss and delay/latency has received significant attention. Consequently, algorithms exist to address these problems and improve performance during mobility. However, notwithstanding the progress made in this area, seamless handover still present challenges in wireless networks. In wireless heterogeneous network, these challenges are even more pronounced due to the complexity of the handover process across such networks.

Typically, heterogeneous networks belong to different administrative domains. Therefore, certain procedures are necessary to support mobility across UMTS and WLAN. 3GPP addressed some the procedure in the standard. In many literatures addressing some of the procedure, mobile IP and AAA have been consistently included in the literature as the primary protocol for enabling mobility across UMTS and WLAN. Nevertheless, these protocols have latencies that contribute to the aggregate delay associated with handover from one network to the other. In addition, there exist other unavoidable latencies that results from handover which invariably contribute to the overall handover latency as shown in figure 5.1.
5.1.1 DOWNWARD VERTICAL HANDOVER

In figure 5.1 different sub-procedures that comprise the downward vertical handover are shown. Also shown is the vertical handover from a home UMTS to a visited WLAN together with the latency distribution among these procedures. The downward handover was shown in this diagram because the handover from UMTS to WLAN, which is a downward vertical handover, is considered in this research in the demonstration of location/position aware vertical handover. However, the handover procedure is also the same for a user who belongs to a home UMTS networks but is currently connected to a visited WLAN wishes to handover to another WLAN. To be able to execute handover to another WLAN while in a visitor’s network, the home network –UMTS network-need to be contacted to re-route traffic to the new WLAN.

Figure 5.1. Downward Vertical Handover Procedure From UMTS to WLAN
As shown in figure 5.1, the handover process is divided into two main stages, “Coverage Discovery” and “Execution”, which include Address Configuration, AAA procedures, Mobile IP procedures, and the Stabilization period needed recover from the effect of handover/mobility on the upper layer protocols and running applications.

5.1.1.1 DISCOVERY

The current non-optimized method for detecting different types of wireless coverage in areas that have heterogeneous wireless overlay networks requires simultaneously active network adapters on the dual-mode MN. The IEEE 802.11 standard defines the method for detecting the presence of WLAN access points (APs). The MAC layer is responsible for triggering passive or (the optional) active scanning to discover WLAN coverage. Passive scanning consists of listening for beacons sent periodically by access points (APs) on specific channels, whereas active scanning consists of the MN sending probes on each channel and waiting for the AP(s) responses. The optional active scanning method offers a faster but less efficient approach—from a power and bandwidth consumption perspective.

When the WLAN adapter receives a beacon or probe response, the MAC layer checks its associated Received Signal Strength (RSS), then compares it with other received beacons’ RSS values. The contents of the beacon are also checked to determine the supported data rates and other network-dependent information. The MN then sends an “association request” to the AP with the strongest beacon, which then authenticates the user and sends an “association response” to confirm the association. Upon receiving the response, the MN sends a router/agent solicitation to discover the access router (or foreign agent as described in chapter 4) currently serving the AP. The router/agent returns an advertisement to the MN providing its relevant information.

Some literature distinguishes between “AP discovery” and “Router Discovery,” whereas others combine the two and also include the Duplicate Address Detection (DAD) as part of the discovery. In this research effort, the discovery latency was considered to be the time between activating the WLAN adapter and receiving the first router advertisement from the router associated with the selected/target AP. As for DAD, it was considered part of the “address configuration” stage. The discovery latency varies depending on the type of scanning, the scanning intervals, the beacon intervals, and the router advertisement interval as well as the frequency of these procedures.

Overlay networks have overlapping coverage areas that support different technologies. The “upper” network is typically larger and has lower data rates, whereas the “lower” network is smaller and offers higher data rates. Therefore, the discovery step in the downward vertical handoff is not delay-sensitive since the MN is not at risk of exiting the large cell it is currently connected to (e.g., 3G/UMTS). Instead, the handoff is typically initiated for performance optimization reasons. On the other hand, upward vertical handoff takes place when the user exits the small coverage area (lower cell). Therefore, the discovery step in upward handoff is usually
delay-sensitive and can cause disconnection and severe packet loss. This reaction would occur in cases where the user suddenly exits the coverage area, and at a high speed, thus not allowing the MN sufficient time to trigger the discovery stage to search for other 802.11 or UMTS coverage. Moreover, the network adapters’ capabilities and configuration on the MN are also a contributing factor to the amount of delay associated with coverage discovery.

5.1.1.2 ADDRESS CONFIGURATION

When a MN visits a new (or previously visited) network, it usually needs to acquire a temporary Care-of IP address (CoA) in order to support seamless mobility, as described in the next subsection. There are a few alternatives for acquiring such CoA. For instance, in the context of Mobile IP; there is an option for using the router’s IP address as the CoA without the need for allocating a unique IP address to the visiting node. In such cases, the latency could almost be negligible, but in some experiments was found to be as high as 1s. According to the arguments made in the literature, other alternatives for acquiring a CoA are important to consider in terms of avoiding a bottleneck or in terms of supporting protocols such as IPv6, which has a different addressing and mobility scheme. The primary method for obtaining the so-called “co-located CoA” in a foreign network is based on the DHCP protocol. According to some reported experiments, the delay associated with obtaining a CoA via DHCP can involve random wait times that can lead to delays in the order of seconds, even when performing a handover between high speed wireless networks. In addition, when co-located addresses are allocated to users visiting in a foreign network, there is a risk that another user may have already manually assigned the same IP address to his/her MN’s interface. Therefore, a Duplicate Address Detection (DAD) feature associated with DHCP was created to check for such situations, using methods such as ICMP Echo request/reply. The DHCP server checks for duplication prior to allocating an address to the visiting MN. It also recommends another check by the MN after it receives the CoA message (prior to finalizing the assignment to its interface). Therefore these checks contribute to the latency associated with obtaining a CoA.

Note that in some literature, the DAD-related delay was acknowledged but not distinguished as a separate latency value during experimentation. In such experiments it was either assumed to be part of the discovery latency or the address configuration latency for the sake of simplification.

5.1.1.3 AAA

The Authentication, Authorization, and Accounting (AAA) protocol is currently viewed as the most suitable solution for managing and exchanging subscriber information and credentials between different networks, particularly combined with Mobile IP. This protocol also enables the mobility and roaming across different networks, particularly ones that have different administrative domains. DIAMETER is the next-generation of AAA and is being standardized by the IETF. DIAMETER has evolved from its predecessor RADIUS.
The basic AAA protocol, which is the basis for DIAMETER and RADIUS, is based on the communication between a Home AAA (HAAA) server and a Foreign AAA server (FAAA). The HAAA is responsible for authenticating and/or authorizing user access in a particular domain. The FAAA represents the authority in the foreign network and is responsible for negotiating client credentials with the HAAA. Furthermore, an intermediate node (proxy) between the FAAA and HAAA is referred to as the AAA Broker (BAAA). In many cases, a few proxies/agents are necessary to provide intermediate communication. As expected, the higher the number of proxies that must be traversed, the longer the AAA latency.

Another protocol worthy of being mentioned once again is Mobile IP. In chapter 4, this protocol was treated in detail. Recall that Mobile IP is a protocol introduced to enhance transparent communication between MN and its home network when the MN is in a foreign network.

5.1.1.4 STABILIZATION

The stabilization period is a combination of the time required for the application(s) to recover after the handoff and adapt to the new data rate, as well as the residual TCP back-off time caused by the handover. TCP backoff is a mechanism that was originally introduced to handle packet loss due to network congestion in the wired network. However, such mechanism interprets the delays and packet losses, caused by handover in wireless networks, as an indication of network congestion. Therefore, it reacts to such events in the same manner that it would network congestion.

Its reaction consists of first dropping the transmission window size to significantly reduce the amount of transmitted data to cope with the presumed congestion. It then activates what is known as the “slow-start” algorithm to slowly increase the size of the transmission window and allow the network to recover from the congestion. Finally, it resets the retransmission timer to an interval that would double with each timeout. Note that timeouts occur when acknowledgements, for sent data, have not been received within a specific period. In such cases, TCP would retransmit the packets believing that they had been lost.

Several research efforts investigated and addressed this issue and have quantified the impact of mobility and handoff on TCP and the resulting impact on throughput and delays during and after the IP-level handover. The primary contributor in this case was found to be the timeout (or consecutive timeouts). Recall that the IP-level (i.e., Layer-3) handover procedures, which come after the coverage discovery step, cause a period of non-communication with the MN (i.e., no user data is received or transferred). Therefore, during this period the reliability mechanism in TCP causes it to perform the backoff procedure described in the previous paragraph. The longer the delay associated with the IP-level handover procedures, the longer the timer intervals will become.

According to some the experiments, the long timeout could result in a pause that lasts from 0.8 seconds to a few seconds depending on the number of timeouts. Therefore, despite the
completion of IP-level handover, the data transmission will not resume until the timer expires and the first acknowledgement is received. At this point, the slow-start algorithm would also require some time to recover and allow the window size to return to its maximum value. The experiments indicate that the slow-start could take up to one second. However, for that duration the transmission is merely throttled, thus the impact of the slow-start on throughput is relatively moderate compared to timeout(s).

5.1.2 UPWARD VERTICAL HANDOVER

A MN currently attached to a WLAN AP monitors the current AP’s beacons and evaluates the Signal to Noise Ratio (SNR) value to detect when a handoff is needed (i.e., when a user exits the current coverage area). Upon noticing degradation in the SNR, the WLAN adapter starts scanning for beacons in search of another AP. After a certain period of scanning, if no other APs exist or have a strong beacon, the upward handoff to 3G/UMTS is triggered. The MN then evaluates the RSS of the Broadcast Control Channel (BCCH) for the available 3G/UMTS cells and decides which cell to attach to accordingly. It then communicates with the nodes in the UTRAN to provide the appropriate nodes in the core network (CN) with its current cell position and requests that the Radio Access Bearer (RAB) be established. RAB are the access stratum for transferring user data and are under the control of the RRC protocol, which is the protocol responsible for the L1 & L2 handover management in UMTS.

The IP-level handover steps (i.e., after the discovery step) are the same for upward and downward handover, so long as the user is moving from a home network to a visited network. If the user is returning from a visited network to a home network, then the Mobile IP and AAA procedures would not be necessary and thus the overall latency is significantly lower. Note that if the user is moving from one visited network to another, the latencies associated with Mobile IP and AAA can be longer due to the number of proxies/gateways that must be traversed to update the user’s CoA and obtain the AAA credentials from the user’s home network. Finally, the upper layers (i.e., transport and application) react in the same manner to the handover-related delays, packet loss, and change in data rate, regardless of the handover direction. Therefore, the stabilization stage’s characteristics and impact are the same regardless of the direction of the handover (upward or downward).

5.2 AGGREGATE HANDOVER EFFECT

The previous section provided a detailed description of the handover stages and procedures for downward and upward vertical handover in heterogeneous wireless overlay networks, particularly 3G/UMTS and WLAN (IEEE 802.11). The latencies of each stage of upward and downward handoff have been demonstrated and addressed by several research efforts to have a significant impact on delays and packet loss observed at the user level. The aggregate delays and packet loss of an average handover cause interruptions in the data transfer during the handover execution stage and the few seconds that follow. These interruptions result in throttling the
application level throughput and increasing the application response time. However, such decrease in throughput would typically be mitigated by the higher data rates offered by the WLAN. Therefore, ultimately, the overall application response time would be significantly less, despite the handover delay, which may have lasted a few seconds.

In order to benefit from the higher data rates, the user must remain within the WLAN coverage area for a period that exceeds that of the handover latency. Note that this argument is limited to non-real-time and non-delay-sensitive applications since such applications would suffer from handover-related performance degradation regardless of the time spent in WLAN and the data rates offered.

5.3 PROBLEM DESCRIPTION

The number of wireless data services and users is rapidly increasing, thus prompting an increase in the number of WLAN coverage areas to support customer demand. Furthermore, the internetworking between 3G and WLAN overlay networks has also prompted the deployment of more WLAN hotspots in areas that have a high population density. This requires smaller cell in order to maintain an acceptable level of QoS, from a bandwidth perspective. Moreover, WLAN hotspots are being deployed in areas characterized by relatively high user mobility such as airports, subway stations, convention centers, outdoor urban areas, parks, and university campuses.

As a result of all these factors, the rate of vertical handover is expected to be significantly higher in such areas. Furthermore, the probability and rate of short WLAN visit durations is also expected to increase. As a user enters and enter a WLAN coverage area a downward handover would be triggered followed by an upward handover as the user exits the coverage area, both triggered based on RF measurements (i.e., the conventional method). The delay associated with the downward than upward handover results in a significant period of delay and packet loss. In essence, this situation would have a similar impact to that of the “ping-pong effect,” which results from a MN traveling on the border of two or more cells thus causing the triggering of a series of unnecessary consecutive handovers between such cells. The negative impact of the “ping-pong effect” has been addressed by some researches.

The negative impact of the consecutive downward than upward handover, without sufficient time to transfer data over the new WLAN connection, would result in data interruptions that throttle the application level throughput and reduce the application response time. Therefore, it would counter the purpose for which the handover from 3G/UMTS to WLAN was triggered in the first place, which is to enhance performance. This represents a problem that has not yet been observed in practice, but that is expected to materialize as the internetworking of heterogeneous, wireless overlay, networks takes place and the number of small WLANs increases in areas of relatively high user density and mobility.
5.4 POSSIBLE SOLUTION

In section 5.3, the problem this research study is attempting to provide solution to, has been identified. In the of the identified problem therefore, a mechanism for predicting the user’s visit duration to a network and deciding on the handover accordingly would prevent the occurrence of inefficient or unbeneficial handovers, thus preventing degradation of application-level throughput and the wasting of resources. This applies primarily to downward handover due to the upper cells (e.g. 3G/UMTS) being significantly larger than the lower cells, which means that the user’s visit duration in the lower networks (e.g., WLAN) would likely be shorter. Other research efforts have called attention to this issue and noted its similarity to the “ping-pong effect” in homogeneous networks.

“Prediction” has also been the focus of research efforts attempting to achieve amore seamless handover in homogeneous and heterogeneous networks, particularly based on a combination of the signal strength and the user’s position. However, the majority of the proposed approaches have relied on signal strength without sufficient assessment of the user’s speed and direction. Furthermore, the research efforts that have considered the user’s velocity have done so with regards to the distance between the MN and the AP only. Therefore, their estimates did not take into consideration the obstacles that could prevent the AP from having a full/continuous range (e.g., walls). They also assessed coverage in an AP-centric manner rather than treating the hotspot coverage as a single entity. Treating multi-AP coverage area as a blanket of coverage would allow the prediction-based algorithm to avoid triggering when only a Layer-2 handover is needed, due to its minimal associated latency. Therefore, this research effort proposed augmenting the current conventional handover decision method with a location-based evaluation that takes into consideration the user’s actual position, speed, and direction as well as the spatial attributes of the discovered WLAN in an aggregated manner (blanket effect).

The main focus of this research effort was the downward handover from 3G/UMTS to WLAN and the layer-3 horizontal handover, of a 3G/UMTS user, between two foreign WLANs. While the handover decision mechanism for triggering upward handover could benefit from location-based information to identity the best time to trigger such handover, this research effort did not support/modify it. This is due to the upward handover’s sensitivity to delay, which requires a mechanism that can make a quick decision to avoid lengthy disconnections, which could cause (the more severe) upper layer disconnections. Therefore, according to the literature, an approach based on RF measurements is more suitable. Examples of such approach include the threshold approach, hysteresis approach, dwell timer approach, or a hybrid of these approaches.
5.5 CHAPTER SUMMARY.

This chapter provided a detailed description of the stages that comprise downward and upward handover, particularly for 3G/UMTS and WLAN overlay networks. This description and qualitative analysis was meant to address the source of the handover related delays in order to show the potential delays associated with vertical handovers and their impact. The problem that motivated this study has been defined in Section 5.4 based on the understanding of the impact of handover, which emphasizes the need for a better handover decision mechanism. Finally, potential solutions found in the literature and their weaknesses were briefly addressed to bring to the reader’s attention the weaknesses that had to be addressed in the proposed solution.
CHAPTER 6
DESIGN CONSIDERATION

6.0 INTRODUCTION

In chapter 2 subsection 2.8.2 the benefit of integrating UMTS with WLAN was highlighted. Integrating such heterogeneous technologies presents some issues as mentioned in chapter 2. The issues include the problems of billing, security, internetworking the infrastructure and the ability to move seamlessly between the two networks. In this study, the main objective is to find ways to enhance mobility and the handover process between UMTS and WLAN wireless network. In this kind of hybrid networks, the handover process involves a cost that is incurred by the network and noticed at the user level (such as degraded performance of an application). This cost would typically be overlooked due to the benefit gained from connecting to a network with a significantly higher data rate. However, this handoff cost would only be mitigated if the user remains in the lower network (i.e., WLAN) long enough to benefit from the higher data rate. However, a short visit to WLAN would resemble the negative impact caused by the ping-pong effect and defeat the purpose of the handover. The need for a location-aided approach was suggested in Chapter 5 to make more location-aware handover decisions that result in the intended performance enhancement rather than degradation.

The assumption made in this study is that the environment in which the user’s home 3G networks and the visited WLAN were loosely coupled- inter-networked without the need for the WLAN access network to have an Iu interface. According to the literature, loose coupling is the primary internetworking approach due to its flexibility for allowing a network operator to offer 3G-WLAN subscriptions to the end user regardless of whether both networks are owned by the same operator. It was also assumed that the MN was equipped with the two corresponding adapters and was capable of automatically activating the WLAN adapter and triggering the WLAN discovery process when the user activated a non-real-time application, and triggering the UMTS discovery process when the user exits 802.11 coverage area.

In this chapter, the location-aided handover decision algorithm capable of handling the problem described in chapter 5 is developed. An overview of the basic conventional handover decision for handovers between heterogeneous wireless-overlay networks is presented. The chapter describes location-aided algorithm, which combined the RF-measurement with a location-based evaluation to evaluate the suitability of a handover and make better handover decisions than the pure RF-based conventional method. An important component of the design, the WLAN coverage database is discussed. The algorithms specifications with respect to the assumptions, requirements, underlying mechanism, and calculations are addressed in the relevant section. The reader will also read the description of the proposed architecture for supporting the algorithm in the UMTS network in addition to the description of the simple communication protocol required.
to support the algorithm's functionality. The chapter concludes with summary of the proposed approach.

6.1 CONVENTIONAL APPROACH (DISCOVERY AND HANDOVER)

The literatures have discussed extensively the conventional approach for discovering and making the decision to trigger an upward or downward vertical handover. Figure 6.1 is an example of the conventional approach. It can be noticed that this approach is based on radio frequency (RF) measurement. Basically, a dual-mode MN currently connected to an upper network such UMTS and scanning for coverage using its other network adapter eg, 802.11b. As soon as the WLAN coverage is discovered through beacons sent by APs, and router advertisement is received as well from the associated router, a downward vertical handover is triggered.

Figure 6.1, Conventional RF based upward and downward handover between UMTS and WLAN
The literatures have discussed extensively the conventional approach for discovering and making the decision to trigger an upward or downward vertical handover. Figure 6.1 is an example of the conventional approach. It can be noticed that this approach is based on radio frequency (RF) measurement. Basically, a dual-mode MN currently connected to an upper network such UMTS and scanning for coverage using its other network adapter eg, 802.11b. As soon as the WLAN coverage is discovered through beacons sent by APs, and router advertisement is received as well from the associated router, a downward vertical handover is triggered.

While the MN is in WLAN network, its connection to the current AP is maintained until the RF measurement by the adapter indicates degradation in RF parameters which is usually the Signal to Noise Ratio (SNR).

At this point the MAC layer triggers passive or active scanning to locate other APs. If another AP within the same hotspot is found, a layer 2 handover takes place without notifying the home network. But if the AP belongs to another hotspot, then a layer 3 handover is required to maintain seamless mobility. A horizontal layer 3 handover to a visited WLAN network includes notifying the MN's home UMTS network of the user's new location. However, if the discovery stage fails to locate other APs with strong beacons, an upward handover is triggered when the SNR value drops below the specified threshold. When such triggering occurs, the UMTS adapter uses RF measurements to discover and evaluate UMTS cells, then triggers the layer 3 handover once the discovery stage is successfully completed. Some literatures discussed several schemes for evaluating RF measurements to determine the best time to trigger a handover in homogeneous and heterogeneous networks. And some of the schemes involve the use of threshold, hysteresis, and/or dwell timer mechanism to reduce the errors that cause unnecessary handovers. As already mentioned, this design does not contribute nor hinder the upward vertical handovers due to their sensitivity to delays and their need for fast decision algorithm based on physical layer measurement

**6.2 PROPOSED LOCATION AIDED HANDOVER DECISION ALGORITHM.**

In this proposed handover decision algorithm, the user's position and direction information was incorporated to assist in predicting the likelihood of handover occurring in a visited WLAN. The algorithm will do this by determining how long a visitor to WLAN coverage area will remain in it. The visit duration must be long enough, that will meet certain minimum time duration, to justify the cost of a handover. This prediction is important since, in its absence, “ping-pong like” effect will be introduced during the handover if the user leaves the WLAN coverage area almost immediately. The “ping-pong like” effect caused by switching back to UMTS is undesirable and has attracted the attention of various researchers. The ping-pong effect is experienced in homogeneous networks at the borders of overlapping network coverage.

The idea behind the algorithm is for it to coordinate the use of location-based information with the mechanisms of the conventional handover procedure that decides on whether a layer 3
handover should take place. This section provides a description of this algorithm for the downward handover. It also provides the algorithm developed for coordinating the use of location-based information in horizontal layer 3 handover to visited WLAN networks. The parameters to be used, as mentioned earlier, by the algorithm include user's position, speed, and direction. These parameters are utilized in combination with other parameters, to determine the minimum required visit duration to the WLAN coverage area. It has been considered that this minimum time period should equal the time the handover procedures takes to complete and the MN has received a specified number of user-data packets. The process for discovering WLAN coverage can also be optimized using location-based information to determine the distance between the user and the nearest WLAN hotspot. However, scanning or probing for beacons is assumed to be the mechanism for discovering WLAN coverage in the following proposed algorithm

6.2.1 DOWNWARD VERTICAL HANDOVER DECISION ALGORITHM.

Looking at the conventional downward/upward handover algorithm in figure 6.1, it can be seen that the discovery of available WLAN coverage included receiving a strong beacon from an AP. In the proposed approach, a trigger that notifies the UMTS network when a strong beacon is received from an AP is included. Figure 6.2 shows the basic outline of the proposed algorithm for incorporating the location-based information into the conventional handover model. The coverage discovery stage would be MN dependent i.e the MN's adapter performs the coverage discovery stage.

Upon receiving the strong beacon, the MN triggers the layer 2 and layer 3 handover procedure to the target WLAN, see figure 6.1. However, it is envisaged in the proposed approach that only the router/agent discovery procedure would take place and a notification sent to the UTRAN to trigger the location-based evaluation mechanism. The layer 3 procedure will occur when the location-based decision recommends a handover process. In order to support such functionality at the MN, the "Location-based Evaluation and Decision Stage" was inserted into the conventional handover paradigm at the MN. The following is a list of the proposed steps, including their descriptions that comprise the mechanism for obtaining and utilizing the users' position and velocity, and the hotspot's coverage footprint to advise the MN with regards to whether a handover would be suitable at a given time. Note that the numbers associated with each step are merely for the purpose of labeling, not to indicate the order of occurrence.

Using the number labels in the algorithm, the description of the algorithm’s functionality is provided:

(1a) MN receives router advertisement, and then sends notification to the UTRAN to trigger the location-based evaluation. The evaluation starts at step (1b).

At step (1b), the appropriate node is notified to trigger the location-based evaluation as soon as UTRAN receives notification from MN. The MN's notification signal contains field that
indicates the hotspot's registration code. This parameter is specifically introduced to support the algorithms mechanism and may be incorporated into the beacon signal as well.

Figure 6.2 The proposed location-aided decision algorithm

(2) Here the location Measuring Unit is notified by the algorithm to obtain position fixes for the user's location and calculates MN's (Mobile Node/Mobile Station) speed and direction of travel. Any positioning tools with high degree of accuracy and good response time can be used to provide such fixes. However, a network based approach such as UTDOA can be used to minimize control signaling and acquisition/response time.

(3) Here the MN's speed is checked. If the MN is moving at a very low speed, its speed may be considered negligible and such evaluation of direction is overlooked. Negligible speed indicates
to the algorithm that the MN is going to remain in the visitor's network within acceptable amount of time to trigger handover. Consequently, the MN receives signal that handover can be executed. Otherwise the number 4 condition is evaluated. This step allows the MN to take advantage of the WLAN coverage as soon as possible because the MN is going to remain in the WLAN coverage area long enough, as determined by its speed.

(4) RVD indicates the minimum Required Visit Duration a user must remain in the WLAN coverage. This time duration must be enough to accommodate "Start" and "End" of handover procedure. RVD is determined such that the user is expected to remain in the WLAN coverage area for the application to benefit from the higher data rate and compensate for handover related delays.

(5) When the condition at step 4 is evaluated, the result of the condition could be that the user would not remain in the same WLAN coverage area for the minimum RVD. In this case, a timer is triggered to allow for a "wait" period between location based iterations. This step ensures that resources are conserved by reducing the number of unnecessary iterations. It also increases the probability of obtaining different positioning values for the MN (if the user was moving).

(6) A flag is checked after the "wait" period. This is done to see if the MN has notified the UTRAN of degraded signal reception from the beacon, indicating the MN has exited the coverage area. The "abort" flag is necessary to ensure that the algorithm did not continue to iterate endlessly irrespective of the fact that the mover has moved away from the WLAN coverage area. If a flag was set, step (7) would be invoked, otherwise step (2) is invoked to start a new location-based evaluation.

(7) If step (6) confirmed that the "abort" flag was set, the location-based evaluation running at the UTRAN terminates.

(8) If the result to queries at step (3) or (4) was a "Yes", then a notification would be sent to the MN. This notification advises the MN to trigger the handover "Execution Stage".

(9) The MN waits for a notification from the location-based evaluation at the UTRAN. If one is received, then the handover "Execution Stage" is triggered. This consists of layer 3 handover procedures. If a notification is not received, step (10) is invoked.

(10) This step represents the MN monitoring the received signal strength from the beacon to check if it had degraded below a certain threshold, thus indicating the user moving out of the coverage area discovered. If so, step (11) is invoked, otherwise, the algorithm returns to step (9).

(11) This step is invoked when the WLAN adapter notices degradation to trigger the handover "Execution Stage". Upon noticing the degradation in RSS, the WLAN "Coverage Stage" is invoked again to search for other WLAN coverage, and a notification is sent to the UTRAN to set the "abort" flag described earlier. This causes the location-based evaluation to be terminated-
to avoid infinite loop at the UTRAN- in order to avoid infinite iterations for evaluating the user's position with regards to a WLAN since the user is no longer located there. The algorithm's iteration at the UTRAN will continue until one of the following events takes place.

* The user is no longer traveling at a noticeable speed

* The user is predicted to remain in the same WLAN hotspot for the minimum RVD value.

* The user exits WLAN coverage and the MN notifies the UTRAN to abort the algorithm.

Note that the algorithm at the UTRAN will only send a notification to the MN if the decision is positive- i.e the outcome of step (3) or (4) is a "yes or 1". Therefore if the coverage discovery successfully completes before the MN receives a notification from the UTRAN, then either a prediction has not been reached or the user's predicted visit duration was found to be less than the minimum RVD(i.e a new evaluation iteration begin after the "wait" timer expires.

**6.2.2 HORIZONTAL HANDOVER DECISION WITHIN WLAN**

Horizontal handovers are handovers that occur in networks that support the same technology. The handover could be between two APs belonging to different Extended Service Sets. Or it could be between base stations belonging to two different cells. The IEEE 802.11 standard and various literatures described such handovers between WLANs based on the strength of SNR noticed by the MN. In other words, handover is triggered when the MN notices degradation on the current Signal to Noise Ratio. As shown in figure 6.1, the discovery stage determines the availability of other APs and whether they belong to the same WLAN hotspot or a different hotspot. Layer 2 handover have minimal latency and does not require contacting the user's home network, the proposed design does not trigger the location-based evaluation when the type of handover involved is in layer 2. However if the target AP belongs to a different WLAN, layer 3 procedures are required.

The latency of horizontal layer 3 handover to a visited network could exceed that of downward handover from a home network to a visited network. This is because the home network of the visiting MN must be contacted in such procedures. Therefore, the proposed location-based algorithm would be triggered, under this kind of handover, to predict the user's visit duration to the new WLAN hotspot and asses the suitability of such handover accordingly. As shown in figure 6.3, if the target AP does not belong to user's current hotspot, the MN notifies the UTRAN to trigger the same location-based evaluation already described. Steps 1-11 are the same as those described for the downward vertical handover. However, in this case, while the MN waits for the notification from the UTRAN with regards to the decision, it must monitor the level of the Signal to Noise Ration. If the SNR drops below the configured threshold value, step (11) would be invoked. However, in this case, the UMTS "Coverage Discovery Stage” is also invoked to avoid the MN being out of coverage for a long time which could cause more severe upper-layer disconnection. This is due to the fact that in the case of layer 3 horizontal handover situation, the
Figure 6.3, The proposed location-aided decision algorithm. The user is connected to visited WLAN Network

user may move out of the current coverage area before receiving a decision from the location-based evaluation running at the UTRAN.

Therefore when the SNR drops below the threshold, an upward handover is triggered to reestablish the connection with the UMTS network, since it is less risky than blindly triggering the layer 3 horizontal handover (since it may result in only a brief visit and causes performance degradation). After the MN reestablishes connection with UMTS, if the RSS of the discovered WLAN coverage continues to be strong, the Location-based Evaluation Decision Stage would be
triggered again. However, in this case, the MN continues to transfer data over the UMTS connection (without the risk of going out of range) as it waits for the handover decision to be sent in a notification from the UTRAN as described in Step(8) and step(9) of section 6.2.1

6.2.3 UPWARD HANDOVER DECISION

Vertical handover has been discussed in the previous chapter. The proposed location-based decision algorithm neither contributed nor hindered the upward vertical handover for the fact RF measurement at the PHY layer is adequate for such handovers.

6.3 WLAN COVERAGE DATA BASE

6.3.1 DATABASE STRUCTURE

To be able to get location-based information, the proposed approach made use of the location information in database. This information is stored up in a database in form of a footprint representation of WLAN coverage information. With this information available, the location-based evaluation implemented at the UTRAN queries the database for spatial information regarding the target hotspots that an MN is trying to handover to. An individual database for each RNS or for a small group of RNC is preferred in order to reduce the query response time. However, a distributed database can also be used to minimize the cost of implementing the proposed approach. Already, 3GPP defined architecture to support location-based services and the components of the architecture have been mentioned in this work. Each cell in the UTRAN has a cell ID that the MN receives through beacons from the cell’s Node-B. Therefore, when the cell ID received at the MN does not match the stored ID, the MN sends an update to the network of its new location. Therefore, the network will always be aware of the user’s cell location when the user is active. This cell ID can be used to narrow down the query and minimize the associated lookup response time. In addition, for the purposes of the proposed approach, a hotspot (comprised of one or more APs’ coverage areas) should be assigned a hotspot ID that distinguishes it from other hotspots within a UMTS cell. This would further narrow down the database search space and reduce the associated delay. However, more importantly, the concept of the hotspot ID was necessary for the correct functionality of the algorithm. A tag such as hotspot ID was needed to distinguish between two different WLAN hotspots. This tag could be embedded into the AP beacon or probe response frame to be received by the MN, who would forward such value to the UMTS network to be included in the Location-based evaluation.

As for the overall querying, updating, or managing of the database, the Structured Query Language (SQL) was the recommended method for spatial databases. The actual queries may use any coordinate system for an index, so long as the indexing format is compatible with or derivable from the coordinates of the user’ positioning tool (e.g., U-TDOA, GPS, etc.). The database structure is expected to be very simple since it merely consists of spatial records/entries stored at a certain resolution and encoded with a hotspot ID.
The concept of using a geo-location database to support and improve the handover process between heterogeneous wireless overlay networks has been investigated by other research efforts. The effort demonstrated in some of the research work described the use of a secure distributed geo-location database as well as fuzzy logic to determine when to trigger an layer 1 and layer 2 downward handover (i.e., for discovery purposes). The database was queried to estimate (using fuzzy logic) whether the user was near or far from the WLAN coverage area in order to determine when the MN should activate the WLAN adapter to prepare for a downward handover.

### 6.3.2 WLAN FOOT PRINT ACCURACY

![Figure 6.4 An Interface of Site Survey Software Package. Source: Airmagnet.](image)

The fault tolerance of the algorithm accommodates less accurate coverage footprints stored in the database. With information collected from a site survey, the accuracy of the coverage representation will be enhanced. In addition, some software packages are capable of generating RF coverage map information in binary form. The binary form information also enhances the accuracy of RF coverage maps since it can be ported directly into the WLAN coverage database to reduce the data entry efforts associated with encoding the spatial information. A snapshot of a user interface of site survey software is given in figure 6.4.

### 6.3.3 GRANULARITY

Although a highly granular coverage map would produce the most accurate result, a tradeoff between that and the query response time (which depends on the size of the database) is necessary in the context of the proposed approach. The proposed algorithm requires that a location-based decision be reached in real-time, thus it is time sensitive. This issue was addressed in the literature [68] by researchers whose effort included the use of a WLAN coverage database that provided coverage information at a 1.5m resolution, for the purposes of...
triggering location-based services depending on the user’s location within a hotspot of considerable size (i.e., university building, airport, etc).

6.4 MATHMATICAL MODEL OF ALGORITHM’S PERFORMANCE

This section provides the calculation involved in obtaining the parameter for estimating the user’s path (trajectory). The result of the calculation is then used to predict whether the user would remain in WLAN coverage for at least the required visit duration (RVD). If it is determined that the user will remain in the WLAN coverage area long enough to satisfy the RVD, a notification would be sent to the MN to indicate the suitability of the layer 3 downward or horizontal handover to a particular WLAN hotspot.

6.4.1 ASSUMPTIONS

In designing the algorithm, the following assumptions were made.

- A dual mode MN with two adapters each for WLAN and UMTS. The adapters should be able to exchange control messages with UTRAN at any point in time.
- The MN is active. A non-real-time application is running on it.
- A positioning method should be able to provide positioning with acceptable degree of accuracy say within 3-5m.
- A WLAN coverage footprint database should be available. This database would be queried for information regarding the user’s visit duration to the target WLAN hotspot whose AP beacon(s) caused the MN to trigger the Location-based evaluation at the UTRAN.

6.4.2 REQUIRED VISIT DURATION

In determining the minimum required time that a user must be in WLAN coverage in order for a handover to be beneficial, the following parameters were used. It can be recalled that vertical horizontal handover to a visited WLAN hotspot involves some costs which is in form of throughput degradation and decrease in application response time. However, a handover to a network with a higher data rate neutralizes the handover cost compensating with higher throughput and lower application response time. For this to happen, however, the user must be in the in the new visited network for a period long enough to allow for the following events to take place:

- The location-based evaluation to successfully complete, and for its decision notification to arrive at the MN. Based on the current technology, the algorithm’s latency is heavily
dependent on the latency of position acquisition delay, which depends which depends on the method of positioning

- The Address Configuration Stage to successfully complete (i.e., DHCP & DAD).
- The Mobile IP and AAA procedures to successfully complete
- The stabilization period to be over
- A sufficient amount of data to be transferred over the WLAN link to compensate for the increase in application response time due to handover latency. The latencies associated with most of these events are network dependent. For instance, the discovery latency is dependent on the intervals between beacons and the frequency of router/agent advertisements. Furthermore, the registration latency, which encompasses both the Mobile IP and the AAA latencies, is a function of the network’s MAC layer characteristics for the specific inter-networked UMTS-WLAN pair. Moreover, a single cycle of the location-based evaluation has a latency that would depend primarily on the position acquisition time and the database query latency or response time, which are also network dependent assuming a network-based positioning method such as U-TDOA. Furthermore, the amount of data that can be transferred after the stabilization stage depends on the throughput of both the current and target network. In addition to these network-dependent variation factors, which have the most impact on latency, there are some MN-dependent or adapter-dependent factors that may vary the latencies. An example of that is the address configuration latency, which depends on the MN’s processing capability and the “state of the interface” [31]. In light of the argument made in the previous paragraph, it is recommended that the latency values be calculated periodically for each 3G/UMTS-WLAN pair. The throughput of the current and target networks is an important factor in calculating the RVD as described below. Therefore, the UMTS network should periodically probe the WLAN hotspots that are inter-networked with it, for the values of their current average throughput. The throughput in any network can vary depending on the network load, which varies based on the user density and the type of applications currently activated by the users. Therefore, periodic updates are necessary in order to maintain a reasonably accurate estimate of the current throughput in the WLAN networks within a UMTS cell. Table 6.1 provides the list of acronyms used and their brief description.

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>La</td>
<td>The latency of a single cycle of the location-based evaluation (at the UTRAN) in addition to the time required to trigger the evaluation and receive the decision notification after the decision is reached</td>
</tr>
<tr>
<td>Lc</td>
<td>The latency associated with the addresses configuration procedures, which consist of obtaining a temporary IP address from a DHCP server at the new network</td>
</tr>
</tbody>
</table>
network after it has been checked using the Duplicate Address Detection
Mechanism (DAD) procedure.

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laaa</td>
<td>AAA protocol latency, which depends on the number of proxies that must be traversed in order for the foreign network’s AAA server to communicate with the user’s home AAA server (HAAA). This communication is important in order for the FAAA to negotiate the user’s credentials before granting certain authorizations. It also depends on the latency associated with the AAA sub-procedures, such as session key generation.</td>
</tr>
<tr>
<td>Lmip</td>
<td>Mobile IP latency associated with the registration step, where the visited network’s FA sends a request to the user’s HA to update it with the user’s new CoA and receive an acknowledgement from the HA (Section 3.2.1.4).</td>
</tr>
<tr>
<td>Ls</td>
<td>Time required for TCP to recover from the residual backoff time caused by the long communication pauses during the handover procedure, even after the L3 handover procedures have completed.</td>
</tr>
<tr>
<td>τ</td>
<td>Time required for receiving data equivalent to the data that would have been received in UMTS during the pause during handover procedures where no user data was transferred (or negligible user data was transferred due to TCP back-off) This is to avoid a situation where: WLAN visit duration = Handover latency</td>
</tr>
<tr>
<td>Rc</td>
<td>Throughput (in Mbps or Kbps) of the current network where the user is located, which could either be a UMTS network or a visited WLAN network</td>
</tr>
<tr>
<td>Rt</td>
<td>Throughput (in Mbps or Kbps) of the target network whose strong WLAN beacon was responsible for triggering the L1 and L2 handover procedures as well as the location-based decision mechanism at the UTRAN.</td>
</tr>
</tbody>
</table>

Table 6.1 Acronym used in the equation and their description

To compute the mandatory time period a user in a visited WLAN network must stay to benefit from a downward handover, the individual latencies (as listed in the table) are summed up to get the aggregate latency. The aggregate latency is equal to the minimum time period or the RVD.

Thus minimum RVD = L_a + (L_c + L_mip + L_aaa + L_s) + τ……………..(1).

Since minimum RVD depends on the throughput of both current and target network, the ratio of these two throughput is used to determine the time when data is not transmitted during handover procedure. This time represent the time data would have been transmitted had the handover not commenced, ie the user is still in UMTS network.

τ = (L_c + L_mip + L_aaa + L_s) x (R_c / R_t)…………………………………….. (2)

That is to say, τ= (handover latency without algorithm) x (Rc/Rt).
The minimum RVD is adjustable to accommodate any envisaged causes of latencies. Network operators can do this, or even increase the value to improve the chances of achieving benefit from the handover to a WLAN hotspot. However, it is recommended that the customized value of RVD remain reasonably low in order to avoid wasting opportunities for transferring a user to WLAN if the handover is suitable.

6.4.3 MN’s TRAVEL SPEED

In the proposed algorithm, the positioning mechanism would be triggered when the conditions necessary for that is met. The positioning method would obtain one or more position fixes for the MN to be used for position, speed and direction calculations. These fixes in the future may be obtained periodically and may not require specific triggering initiated by the algorithm. This will no doubt improve the accuracy and response time. Until then, the proposed approach assumes that position fixes would be acquired when step 2 of the proposed algorithm is invoked. Based on this assumption, a tradeoff between the accuracy and response time is important and depends on the type of positioning method and capabilities available.

6.4.4 PREDICTED PATH LENGTH (PPL)

Given the value of the minimum RVD and the user’s average travel speed (Ms), the algorithm computes the Predicted Path Length (PPL). PPL is the estimated distance that would be traveled by the user in RVD second, at the user’s average speed. This provides the distance of the user’s trajectory for a period equal to the minimum RVD, assuming the user maintain fairly fixed speed. This parameter will be used in later in the algorithm. The PPL is simply calculated as follows:

\[ \text{PPL} = \text{RVD} \times \text{Ms} \]

6.4.5 USER TRAVEL DIRECTION

The approach proposed in this study relies heavily on the algorithm’s knowledge of a reasonably accurate estimation of the user’s current travel direction. Further details regarding the use of the user’s travel direction is the following subsection. A minimum of two fixes must be obtained in order to compute the user’s velocity vector which represents the user’s speed and direction. Figure 6.5 shows the visual representation of the velocity vectors. The number of fixes obtained depends on the accuracy-response time tradeoff which in turn depends on the capability of the positioning method. If the user’s position was tracked to support other location-based services, then depending on the frequency of position updates, such tracking information could reduce the latency of computing the user’s location-based parameters and reduce the algorithm’s overall latency.
6.4.6 COVERAGE QUERY

Given the user’s speed, which is not negligible as determined by the algorithm, the algorithm would then predict whether the user would remain within the same WLAN coverage area for at least the minimum RVD. Note that, at this point in the algorithm, the user’s position, speed and direction have been computed. Furthermore, the minimum RVD and the user’s average speed were used to calculate the PPL. The algorithm then calculates the user’s Trajectory End Point (TEP) coordinates \((x_2, y_2)\), shown in figure 6.6 based on the user’s position coordinates \((x_1, y_1)\), direction \((\theta)\), and the calculated PPL as follows:

\[
\text{Sin}^{-1} (\theta) = \frac{\text{opposite}}{\text{hypotenuse}} = \frac{a}{\text{PPL}} \tag{4}
\]

\[
\rightarrow a = \text{PPL} \times \text{Sin}^{-1}(\theta)
\]

\[
\text{Cos}^{-1}(\theta) = \frac{\text{adjacent}}{\text{hypotenuse}} = \frac{b}{\text{PPL}} \tag{5}
\]
\[ b = \text{PPL} \times \cos^{-1}(\theta) \]

The coordinate of TEP \((x_2, y_2)\) depend on the user’s direction \((\theta)\)

If \(0 < \theta < 90\) \(\rightarrow x_2 = x_1 + b\) and \(y_2 = y_1 + a\)

If \(90 < \theta < 180\) \(\rightarrow x_2 = x_1 - b\) and \(y_2 = y_1 + a\).

If \(180 < \theta < 270\) \(\rightarrow x_2 = x_1 - b\) and \(y_2 = y_1 - a\).

If \(270 < \theta < 360\) \(\rightarrow x_2 = x_1 + b\) and \(y_2 = y_1 - a\)

\[ \text{Figure 6.6 Obtaining the TEP coordinates from the user’s position, direction and PPL.} \]

After the calculation of TEP coordinates is completed, the algorithm would trigger a query to the WLAN coverage database. The trigger would include the UMTS cell ID where the user is located at that time, the WLAN hotspot ID of the target network, and the TEP coordinates. The cell ID and hotspot ID would then be used to retrieve the list of hotspot records that represent the target hotspot. This list varies in length depending on the WLAN footprint representation resolution. A mapping mechanism would then link the TEP coordinates to the record/entry that encompasses the location represented by the coordinates. If the TEP coordinates do not fall
within the WLAN hotspot footprint, then the algorithm does not recommend the handover, and step (5) in the proposed algorithm would be triggered. On the other hand, if the TEP coordinates were successfully linked to a record in the list shown, for instance, in Figure 6.7, step (8) would be invoked. Recall that this step would include sending a notification to the MN to recommend the triggering of handover to the target hotspot.

**Figure 6.7 Visual representation of the coverage query mechanism based on TEP coordinates**

WLAN coverage database has been used widely in recording WLAN footprints for various purposes such as planning and security. Some research has also described positioning approach that was based on the idea of “received signal strength finger printing” rather than signal propagation (e.g TDOA). The design involved collecting RSS fingerprints at a series of locations within WLAN hotspots and storing them in a database at a 1.5m resolution, to represent the WLAN hotspot’s footprint. These databases were then indexed using RSS fingerprints, which were obtained through RF measurements at a WLAN adapter. Finally, a pattern recognition mechanism was used to link the RSS fingerprint, obtained at the MN, to a geographic location within the WLAN coverage area, thus obtaining the user’s position.
Some positioning concept re-enforced the idea proposed in this section, which similarly involved the use of a WLAN coverage database. However, beyond the basic concepts associated with using such database, there are differences in approaches, particularly with respect to objective, functionality, and underlying calculations.

### 6.5 THE PROPOSED ARCHITECTURE

3GPP described the architecture for location services as well as how it is integrated with the UTRAN to accommodate the entities that must be added to the infrastructure. It will be recalled that the proposed handover decision algorithm acts as public land mobile network operator client. This client is one out of a number of LCS clients described by 3GPP for location services. PLMN uses location information to optimize performance of the network. Therefore, the LCS architecture can be further augmented as shown in figure 6.8 to support approach by adding two main entities: a Vertical Handover Decision Controller (VHDC) and a WLAN Coverage Database Server (WCDS).

Figure 6.8 Architecture for the proposed solution.

3GPP described the architecture for location services as well as how it is integrated with the UTRAN to accommodate the entities that must be added to the infrastructure. It will be recalled that the proposed handover decision algorithm acts as public land mobile network operator client. This client is one out of a number of LCS clients described by 3GPP for location services. PLMN uses location information to optimize performance of the network. Therefore, the LCS architecture can be further augmented as shown in figure 6.8 to support approach by adding two main entities: a Vertical Handover Decision Controller (VHDC) and a WLAN Coverage Database Server (WCDS).
The VHDC provides the functionality required to run the proposed algorithm in terms triggering the necessary actions in the UTRAN to obtain the algorithm’s key input parameters (eg MN location fixes, speed, and direction). It also performs the computation for determining the duration that a user must be within WLAN coverage in order for the handover to be beneficial. It then queries the WCDS over the defined Iu-u interface to determine whether the user’s travel path, for the required visit duration, will have sufficient WLAN coverage.

The WDCS is a database server that holds the virtual 2-D maps of areas served by the SRNC in the UTRAN. It communicates with the VHDC to respond to the queries regarding the existence of WLAN coverage in the user’s path. The WCDS was placed in the UTRAN, rather than the UMTS core network or in the external network, to reduce the query response time. Another benefit of having a local WCDS in the UTRAN is to simplify the mechanism for updating the information regarding the WLAN coverage in the area served by the SRNC. Depending on the number of cells (i.e., Node-Bs) served by the SRNC, the size of such cells, the user-density and WLAN density in these cells, the WCDS may be able to accommodate more than one SRNC. On the other hand, in some cases where these attributes are high (e.g., urban areas with a large number of hotspots), additional WCDS units may be required to ensure a low query response time. The VHDC functionality was placed in the SRNC for a number of reasons. The main reason is that the RNC is the central control unit in the UTRAN and is responsible for handover control, among other responsibilities. Therefore, VHDC functionality will act as a natural extension to the handover control mechanisms already established in the SRNC, thereby, reducing the modifications necessary to support the proposed approach.

6.6 COMMUNICATION PROTOCOL

In this section, a protocol is proposed for the proposed algorithm. The protocol is needed for exchanging commands and information between the MN and the UTRAN. Four different notification messages are needed to support the proposed algorithm as shown in figure 6.9.

![Figure 6.9](Image)

Figure 6.9 the protocol specifying the messages exchanged between MN and the UMTS network to support the proposed location-aware/location-aided decision algorithm
The following is the description of each message type and the reactions that would cause it to be generated:

(a) This message would be sent to the UMTS network to trigger the location-based evaluation when the MN’s WLAN adapter discovers WLAN coverage via a strong beacon/response. This message must carry the hotspot ID.

(b) This message would be sent to the MN to recommend the triggering of the layer 3 handover procedures (i.e., Mobile IP, AAA, etc).

(c) This message would be sent to the UMTS network to indicate the MN’s desire to trigger an upward handover due to the user exiting the WLAN coverage. Note that this message was not strictly used to support the proposed algorithm. It would be needed, however, with any algorithm that assumes an automatic handover between UMTS and WLAN.

(d) This message would be sent to the UMTS network to terminate/abort the location based evaluation running at the UTRAN if the user moves away from the WLAN coverage area. As previously described, this prevents infinite loops from occurring since the network would have no other way of knowing. A potentially suitable approach for supporting these messages is to utilize the already existing protocols to carry such messages between the MN and the UTRAN. The RRC layer in the UTRAN protocol stack is the most suitable for such a task due to its existing support for handover procedures in UMTS.

6.7 SUMMARY

In this chapter, a description of location-aided algorithm design was provided to meet the research objectives. The algorithm for the conventional approach was combined with the algorithm with the proposed approach for triggering layer 3 handover procedures in the heterogeneous wireless overlay networks. The algorithm mainly supported downward handover from UMTS to WLAN and layer 3 horizontal handover to a visited WLAN. Upward handover from WLAN to UMTS was not supported nor hindered by the proposed algorithm nor hindered by the proposed algorithm.

The conventional approach was augmented without changing the basic flow of events that comprises the IEEE 802.11 specifications with regards to PHY and MAC layer handover. The location-based evaluation was designed to wait for a trigger from the MN when WLAN coverage was discovered via RF measurement at the MN’s WLAN adapter. The location-based evaluation would then run on the network side and notify the MN if the handover suitable. Otherwise, the evaluation would continue monitoring the user’s position and computes the TEP if changes occur. This continues until the user exits WLAN coverage or the user’s location-based information indicates that he/she will remain within WLAN coverage for the minimum RVD, thus triggering the handover to the discovered WLAN.
The underlying assumption, specifications, requirements, and calculations were also addressed in this chapter. The main underlying mechanism for obtaining and computing the minimum RVD, PPL, and TEP-coordinates was described in detail. Furthermore, the concept of using a WLAN coverage database to predict whether the user would remain within the same WLAN coverage for the RVD was also described in this chapter. In addition, the modifications that would be imposed on the UTRAN architecture were addressed with respect to the 3GPP standard architecture for location-based services (LCS). Finally, the basic control messages that would be required to support the proposed algorithm’s functionality, in terms of a communication protocol between the MN and the UTRAN, were briefly described in an effort to provide a more comprehensive look at the proposed algorithm’s requirements.
CHAPTER 7
ANALYSIS

7.0 INTRODUCTION

In chapter 6, subsection 6.2.1; the concept of the proposed algorithm for the location-based decision algorithm was illustrated. It can be seen from the diagram that the proposed algorithm was integrated with conventional RF measurement based handover algorithm. The proposed algorithm makes use of the functionalities of RF measurement handover algorithm to discover the existence of WLAN coverage during the WLAN “Coverage Discovery” stage. After the completion of the “Coverage Discovery”, the “location-based evaluation and decision” stage is triggered at the MN followed by the UTRAN being sent a notice to invoke the location-based evaluation for assessing the user’s velocity, and the average handover latency between the two networks. This information is used along with the spatial information of the discovered WLAN to predict whether the user’s visit to such a coverage area will be long enough to allow the application to benefit from higher data rates. Therefore, the location-aided algorithm pre-examines the suitability of the handover based on more parameters than merely the received signal strength of a beacon, which is the basis for the conventional method.

The functionality, capabilities, benefits and limitations of the proposed algorithm will be discussed in this chapter. Section 7.1 demonstrates the algorithm’s functionality and performance by comparing the outcome of the algorithm in two primary scenarios to the outcome of the conventional approach. It also addresses the algorithm’s limitations by examining its potential performance under worst-case scenarios. The following section demonstrates the algorithm’s performance in a quantitative manner by addressing the result/benefit of applying the algorithm. Section 7.3 will address the environments where the algorithm would be most suitable and beneficial based on the evaluation of its capabilities and limitation. The chapter concludes with a summary.

7.1 QUALITATIVE ANALYSIS OF THE ALGORITHM’S PERFORMANCE

This section describes two primary scenarios necessary for demonstrating the performance and functionality of the location-aided handover decision algorithm. The scenarios were used to show how the proposed algorithm differs from the conventional handover decision method in terms of reaching a different decision under the same conditions. Essentially, the conventional approach is divided into “Coverage Discovery” stage and the handover “Execution “stage.

The handover “Execution” stage begins when the WLAN adapter receives a strong beacon or probe response at the WLAN adapter. The handover “Execution” stage has already been shown in chapter 6. As already mention, the conventional handover procedure relies much on received
signal strength. Consequently, handover is initiated based on beacon’s strength at the location where the WLAN adapter received the beacon. Therefore, at negligible speed of an MN, both conventional approach and the algorithm proposed in this study will yield the same outcome, which is to trigger handover to the WLAN. Recall that in step (3), that the location-based evaluation is running at the UTRAN. This step is responsible for checking the user’s speed to determine whether it is negligible and to generate, send a notice to the MN to trigger the handover to the discovered WLAN upon receiving notice. In cases where the user is static or moving at negligible speed, no benefit is gained from the use of the algorithm instead of the algorithm.

However, if the MN is moving at non-negligible speed, the algorithm’s functionality will be invoked, thereby benefiting the process of handover. Recall that the algorithm is meant to serve as a kind of utility function to the conventional handover method under mobility conditions. Furthermore, UMTS and WLAN are the heterogeneous networks within which this algorithm is designed to operate in areas where the probability of WLAN visit duration is long enough to benefit from handover procedures. The location-assisted handover algorithm is particularly important in these areas because the conventional handover algorithm could lead to performance degradation especially at the application level.

7.1.1 IDEAL ALGORITHM PERFORMANCE

The following scenarios demonstrate the performance of the algorithm during user mobility. These scenarios were developed under an assumption of an error free environment (ideal conditions) to eliminate irrelevant factors and highlight the performance of the algorithm compared to the conventional method.

7.1.1.1 SCENARIO 1 (DOWNWARD VERTICAL HANDOVER)

(a) Conventional Method Only

The scenario in figure 7.1 shows an MN connected to the user’s home 3G/UMTS network, which overlaps with the WLAN coverage area (the circle highlighted in blue). A non real time application is assumed to be running in the MN. Upon activating this application, the WLAN “Coverage Discovery” step is triggered. In this scenario, the user is already within the coverage area of a WLAN in the building illustrated in the figure. Upon discovering the existence of the WLAN, the handover “Execution” stage is triggered, which includes the layer 3 handover procedures already described.

The problem in this case is that the conventional method does not take into consideration that the user is headed out of the discovered WLAN coverage area. This is due to the fact that it is unable to examine the user’s position, speed, and direction, which will result in a visit shorter in
duration than the period required for completing the layer 3 handover procedures. In other words, the user’s visit duration (UVD) is less than the minimum Required Visit Duration (RVD) to compensate for the cost/impact of the handover. The minimum RVD is essentially the amount of time that the user must remain within the discovered WLAN’s coverage area to complete the handover procedures and transfer data over the new WLAN connection.

As the user moves towards the exit of the building and out of the coverage area, the MN’s WLAN adapter notices degradation in the signal to noise ratio (SNR) value from the current AP. Therefore, it triggers the WLAN “Coverage Discovery Stage” to search for better WLAN coverage. However, in this scenario, no other APs are available since the user is exiting the building. The SNR value drops below the configured threshold thus triggering the UMTS “Coverage Discovery Stage” at the MN in an attempt to perform an upward handover back to UMTS. In this case, the downward handover execution was triggered inappropriately and did not result in any benefit since no data was transferred over the new WLAN connection. Furthermore, this handover would in fact result in a negative impact on performance as further addressed in Section 7.2.

Figure 7.1 Scenario 1, limitation of the conventional approach.
(b) PROPOSED LOCATION-ASSISTED ALGORITHM APPLIED

In section 6.2 where the proposed handover decision algorithm was discussed, the location-based Evaluation and Decision step is triggered when the first advertisement from the router associated with the discovered WLAN AP is received. Thus, the location-based evaluation is triggered at the UTRAN. Before then, the algorithm would have obtained the user’s position, mean speed, and the direction; then utilize the information along with the minimum RVD value to calculate the predicted path length for the user as shown below:

\[ \tau = (\text{Handoff Latency}) \times (Rc / Rt) \]

\[ \text{RVD} = L_a + (\text{Handoff Latency}) + \tau \]

\[ \text{PPL} = \text{RVD} \times \text{Ms} \] (Where Ms is the user’s mean travel speed).

![Figure 7.2 Scenario1, location-aided algorithm applied.](image)

The PPL value and the user’s direction angle would have then been utilized for computing the TEP coordinates. The WLAN coverage database serving the user’s current UMTS cell would have then been queried using the UMTS cell ID, and the TEP coordinates for the user.
Note that the grid shown in Figure 7.2 merely provides a visual representation of how the WLAN coverage is stored in the WLAN coverage database. The TEP coordinates are mapped to the appropriate grid block (in this case the block is highlighted in red). If the grid block belongs to the discovered WLAN coverage area, which is not the case in this scenario, then the algorithm triggers the layer 3 handover procedures to the discovered WLAN.

In this scenario, however, the algorithm running at the UTRAN determined that the user’s current TEP is outside of the WLAN coverage area. Therefore, instead of triggering the handoff “Execution Stage,” the algorithm sets a “wait” timer (Step (5) in the decision algorithm and iterates after it expires to obtain new TEP coordinates and use them for a new database query. In the meantime, the MN refrains from triggering handover “Execution State” and continues to transfer data over the UMTS network while waiting for a decision from the UTRAN.

Looking at figure 7.2, the user is heading towards a building exit and thus is heading away from the WLAN coverage area. Therefore, eventually the value of the SNR degrades, thus triggering the WLAN “Coverage Discovery Stage” in an attempt to find better WLAN coverage. When the SNR drops below the configured threshold, which occurs here due to exiting the building, the algorithm at the MN sends a notification to the UTRAN to set the “Abort” flag, which causes the algorithm to terminate. It then returns to the “Coverage Discovery Stage” to search for other WLAN coverage.

Essentially, this scenario showed that if the location-aided handover decision algorithm was applied to in place of the conventional method, the downward handover from UMTS to the discovered WLAN would not have been triggered, thus the handover-related performance degradation would have been prevented.

7.1.1.2 SCENARIO 2 LAYER 3 HORIZONTAL HANDOVER.

(a) ONLY CONVENTIONAL METHOD APPLIED

Another scenario is shown in Figure 7.3. This scenario illustrates the same shortcoming of the conventional approach but in the context of a layer-3 horizontal handover to a discovered WLAN. In this scenario, the user has already performed a handover from its home 3G/UMTS network to a visited WLAN (hotspot 1) and is currently receiving data over the established WLAN connection. As the user moves towards the border of the current WLAN coverage area, the SNR begins to degrade thus triggering the WLAN “Coverage Discovery Stage” to search for other coverage. In this scenario, the MN’s WLAN adapter receives a strong beacon from an AP that belongs to a different WLAN (Hotspot-2) as shown in Figure 7.3. Upon receiving the strong beacon and discovering the AP’s associated router, the conventional approach triggers the handover.
“Execution Stage” (i.e., layer 3 handover procedures) to establish the connection through Hotspot-2. However, in this scenario, the user is moving at a speed and direction away from newly discovered WLAN coverage area as well.

As with the previous scenario, the MN’s WLAN adapter noticed degradation in the SNR and triggered the WLAN “Coverage Discovery” stage, which failed to find any other coverage. When the SNR finally dropped below the configured threshold, the UMTS “Coverage Discovery” step was triggered to perform an upward handover back to UMTS. Therefore, the handover from Hotspot-1 to Hotspot-2 was wasteful of resources (i.e., bandwidth for handover control messages, and processing power at the MN). It also contributed to performance degradation rather than enhancement.

(b) THE PROPOSED ALOGORITHM APPLIED

Because the conventional method was applied as shown in figure 7.3, the user’s position, speed and direction was not taken into consideration. Also, the minimum required visit duration (RVD) was not made use of. Based on such parameters, the TEP coordinates would have been computed, and then used to query the WLAN coverage database. As shown in Figure 7.4 the grid block containing the TEP is outside of the coverage area of Hotspot-2. Therefore the algorithm would not have triggered the handover “Execution” step- i.e., layer 3 handover procedures.
Figure 7.4 Scenario2 the location-aided algorithm in operation.

The MN in this case would have continued to transfer data over the WLAN connection established with Hotspot-1 until the SNR value of that WLAN degrades below the configured threshold. Upon dropping below the threshold, the UMTS “Coverage Discovery” stage would have been triggered to attempt an upward handover from WLAN (Hotspot-1) directly to the 3G/UMTS network, without wasting resources in performing an inefficient and unnecessary handoff to Hotspot-2.

Figure 7.5 performance of the algorithm when the user enters Hotspot2 while transferring data over the 3G/UMTS connection.
This scenario demonstrated the algorithm’s performance during a situation where the user was already connected to another WLAN (i.e., Hotspot-1) when entering the boundaries of a new WLAN (i.e., Hotspot-2). However, the algorithm’s outcome would have been the same had the user been connected to 3G/UMTS when entering the coverage area of Hotspot-2, as shown in Figure 7.5.

7.1.2 PROPOSED ALGORITHM LIMITATIONS

The location-aided algorithm has limitations. The performance limitation is presented to show the weakness observed in the worst-case scenario. The limitations are illustrated using scenarios similar to those used in the previous section.

7.1.2.1 PHYSICAL BOUNDARY LIMITATION

As shown in Figure 7.6, the physical boundary limitation or the “no-exit” problem is mainly observed in indoor WLAN coverage areas.

It occurs when the proposed algorithm is triggered as the user moves towards the boundaries of the building, which could also be the WLAN’s coverage area.

Figure 7.6 Scenario demonstrating the physical boundary limitation.

In this scenario, the algorithm’s computations determine that the user’s TEP falls outside of the WLAN coverage area. Therefore, the location-based evaluation at the UTRAN does not notify the MN to trigger a handover to the discovered WLAN. However, since this is an indoor WLAN
coverage area whose boundaries are defined by the physical space (i.e., the building) in which it is deployed, the user cannot exit the WLAN coverage area unless a building exit is available. Therefore, in such situation, the algorithm’s initial outcome is to set the “wait” timer and not trigger the handover. Since there is no building exit in this case, refraining from the handover is considered wasteful. However, this decision does not persist since the user may change direction.

Other limitations include limitations due to sudden change in direction. This change could also occur in form of rapid increase in speed of the user. If the algorithm has determined that the user is going to remain in the WLAN long enough to benefit from downward handover, the algorithm cannot adjust immediately to sudden change in direction whereby the user exits the WLAN coverage, or increase speed which ultimately leads to leaving the WLAN coverage.

Also, the algorithm’s latency could result from delays associated with obtaining the position fixes, long processing time at the LMU and obstructions to the GPS. This latency is also a limitation the algorithm has not been designed to handle.

7.2 SUMMARY.

The focus of this chapter was to demonstrate the operation of the proposed algorithm. In doing this, the differences in capabilities between the conventional handover algorithm and the proposed algorithm were highlighted. The proposed algorithm was shown to be able to predict the user’s duration of stay in a visited WLAN network, and consequently invoke the necessary downward handover process depending on whether the time of stay meets minimum required time or not. However, in contrast, the conventional algorithm which uses only RF measurement lacks this predictive capability. The algorithm’s limitations were discussed to show the conditions the algorithm’s design could not handle adequately.
CHAPTER 8

CONCLUSION

8.0 THESIS SUMMARY

The study of heterogeneous wireless overlay networks such as 3G/UMTS and WLAN/IEEE 802.11 and how to interwork them have been the focus of a great deal of research efforts due to the variety of emerging applications that require different levels of QoS. In addition, the complementing nature of these technologies makes them suitable for integration to enhance the overall capacity of the 3G/UMTS network and provide better performance to users via the higher WLAN data rates. In an environment where such networks are loosely coupled and their coverage areas overlap, a dual-mode MN is expected to handover between such networks depending on WLAN coverage availability. The conventional method for discovering WLAN coverage and making the handover decision is based primarily on RF measurements such as Received Signal Strength (RSS). This approach has been the source of a number of problems (e.g., ping pong effect, near far effect, etc.) that have been addressed in many publications. This study addressed another handover-related issue that stems from relying entirely on an RF-based method for making the handover decision between heterogeneous networks.

A number of factors were expected to contribute to the increase in the rate of handover between heterogeneous networks in the observable future. The increase in the number of wireless users requires that wireless cells become smaller in order to maintain an acceptable level of QoS. Furthermore, the increase in the number of WLAN deployed in areas of higher user mobility, for instance airports, parks, malls and campuses were expected to result in users spending only a short period of time in any one WLAN’s coverage area. In such cases, the RF-based handover decision method was expected to trigger a downward handover upon discovering WLAN coverage then trigger a handover to another WLAN or back to 3G/UMTS upon exiting the current coverage area. The occurrence of such handover events depends on the size, shape, and number of WLAN cells in a given area as well as each user’s speed and direction of travel in such area. The nature of handoff, particularly between heterogeneous networks that typically belong to different administrative domains, was found to cause substantial interruptions in data transfer. However, handoff to WLAN was expected to compensate for such interruptions by completing the interrupted task at a significantly higher data rate. However, due to the expected increase in short visits to a WLAN’s coverage area the user could exit the discovered coverage area before the completion of the handover procedures and the transfer of sufficient data to compensate for the interruption. In such cases, the interruptions caused by the handover to and from the discovered WLAN merely caused degradation in performance and wasted resources. Recall that the handover from a 3G/UMTS network to a WLAN is mainly triggered for performance optimization purposes, not out of necessity. Therefore, performance degradation...
caused by one or more inefficient and unnecessary handover would essentially defeat the original purpose for internetworking 3G/UMTS networks with WLANs.

The technologies involved in this research effort (i.e., IEEE 802.11, UMTS, etc.) were studied during the initial research period in order to determine the capabilities and limitations of such technologies. Upon establishing a better understanding of these technologies’ specifications, a clearer view of the problem and the potential solution was achieved. The RF-based conventional handover decision method was sufficient for low user mobility and relatively large WLAN cells (e.g., corporate use). However, it was clearly lacking the means to accurately assess the available WLAN coverage for users traveling through areas characterized by small cells and higher user mobility. This was mainly due to the conventional method’s lack of awareness of the user’s position, speed, and direction. Therefore, there was a need for a solution that took into consideration the user’s position and velocity, the handover delays between the inter-networked 3G-WLAN pair, and the spatial properties/attributes of the discovered WLAN. An algorithm was developed to support an environment where short visits to WLAN coverage areas were likely to occur and the handover rate was high. The algorithm was not designed to entirely replace the current RF-based method. Instead, it utilized the RF-based measurements to discover coverage and trigger the proposed evaluation stage, rather than triggering the handover procedures automatically. The majority of the location-based evaluation ran at the UTRAN and took into account the user’s position, speed, and direction. These parameters were obtained via one of the positioning methods that were included in the 3GPP standard for Location-based Services (LCS) such as A-GPS or U-TDOA. The architecture introduced by 3GPP to support LCS consisted of specific nodes for supporting the calculation of the user’s position, speed, and direction.

The algorithm also took into account the values of the average delays associated with a handover between the inter-networked 3G/UMTS and WLAN pair. These delays were used to compute the minimum period that a user must remain in the coverage area of the discovered WLAN in order for the application to benefit from the higher data rates. The value obtained from this calculation as well as the user’s average speed and direction was used to estimate the user’s potential Trajectory End Point (TEP) coordinates.

The TEP coordinates along with the discovered WLAN’s ID and the UMTS cell’s ID were used to query a WLAN coverage database. The database used a mechanism to map the TEP to a particular grid block or entry. If the TEP was found to be within the boundaries of the discovered WLAN’s coverage area, then the UTRAN sent a notification to the MN to trigger the handover to the discovered WLAN. Otherwise, the location-based evaluation was designed to continue iterating until the TEP coordinates fell within the boundaries of the discovered WLAN’s coverage area, or the user exited the coverage area and notifies the UTRAN to abort the evaluation.

The location-based evaluation running at the UTRAN was the same regardless of whether the user was transferring data over a 3G/UMTS connection or another WLAN connection. In either
case, the MN communicated with the 3G/UMTS network over the control plane (C-plane) to perform the location-based evaluation. On the other hand, the procedures running at the MN were different depending on whether the MN was transferring data over a 3G/UMTS connection or another WLAN’s connection when discovering new WLAN coverage.

Two primary scenarios were discussed to demonstrate the capabilities and performance of the algorithm in cases where the MN was connected to a 3G/UMTS network or when the MN was connected to another WLAN, upon discovering new WLAN coverage. The algorithm does not provide any improvements over the RF-based handover decision method in cases where the decision is to trigger the handover. Therefore, the two primary scenarios consisted of situations that would produce contrary decisions using the conventional handover method and the proposed location-aided algorithm. In both scenarios the RF-based decision was to trigger the handover to the discovered WLAN, whereas the proposed location-aided algorithm’s decision was to refrain from triggering the handover. The logic and functionality of the algorithm were demonstrated and discussed using these scenarios. In both cases, the algorithm’s ability to take into consideration the user’s position, speed, and direction as well as the handover delay values allowed it to predict that the user’s visit to the discovered WLAN was not long enough to accommodate the handover delays, the stabilization period, and the transfer of a sufficient amount of data.

8.1 CONCLUSION

The importance of the proposed algorithm lies on its predictive ability. With conventional approach, a user who visited WLAN coverage area is handed over to the WLAN network to satisfy the benefit of extending UMTS with WLAN. It could be recalled that downward handover is triggered between WLAN and UMTS for the purpose of extending the UMTS. But this is done without regard to the user’s duration of stay in the WLAN coverage area. Consequently, such handovers often waste resourceful in the situations the user’s duration of stay in the WLAN coverage area is very short.

The proposed algorithm is designed to predict, using some parameters, the time the user will probably spend in a visited WLAN network. With this capability, the algorithm instructs, or otherwise the MN and the UTRAN to either or not trigger handover process. The algorithm’s contributions were merely observed in terms of application-level throughput and application response time. The impact of consecutive handovers which result from short visits to a discovered WLAN has been demonstrated. These short visits caused interruptions in data transfer that could have lasted for a period as long as twenty seconds. These interruptions were shown to significantly throttle the application level throughput level throughput and cause an increase in application-response time. Depending on the type of application, the length of the interruption period, and the frequency of interruptions, the increase in application response time could become noticeable by the user. Therefore, the algorithm’s ability to predict short visits allowed it to prevent the occurrence of handovers that would result in such performance.
degradation. The algorithm was designed to perform its location-based evaluation in the background (i.e., at the UTRAN) while the MN continued to transfer data over its current connection and without interruptions. Only when the user’s visit to a WLAN’s coverage area was foreseen to be long enough to result in benefit does the algorithm trigger the handover to such WLAN. This allowed the task, for which the data transfer was initiated, to complete quicker than if the short visit to the discovered WLAN’s coverage area had triggered the handover (i.e., using the conventional method). The channels used to transfer data were made available after quickly completing data transfer due to high data rate of the WLAN, thus it allowed the network to offer more bandwidth to complete other users’ task sooner.

Furthermore, the algorithm’s ability to withhold triggering handover associated with short visits avoided the unnecessary loading of the servers used to support mobility for example servers such as DHCP servers, AAA servers, Mobile IP servers, etc. Moreover, it prevented the unnecessary consumption of bandwidth and transmission power associated with exchanging handover control messages. In addition to the algorithm’s benefits, its limitations were also addressed in this study. One such limitation is the situation where the algorithm’s limitations could result in undesirable or inaccurate handover decisions. The algorithm’s limitations were not expected to have a significant impact on the outcome. The algorithm’s mechanisms were designed to recover from errors and take self-correcting measures. Moreover, one algorithm limitation, which depended mainly on the latency of the algorithm itself, was expected to become obsolete as a result of advancements in user tracking technologies and location-based services in the observable future. Finally, the objective of this research effort was satisfied as demonstrated and addressed throughout this document. An extensive analysis of the problem, which highlighted the need for a location-aware handover decision approach, was conducted during the course of this study. Based on such analysis, a location-aided handoff decision algorithm was developed to make better handover decisions in heterogeneous environments characterized by higher user density, smaller cells, and higher user mobility, which result in shorter user visits to WLAN coverage areas (i.e., a higher rate of unnecessary handovers). The proposed algorithm was shown to meet the underlying objective, which was to reduce or eliminate the triggering of handovers that result in performance degradation and wasting of resources. By applying the proposed algorithm, the only handovers allowed to trigger are the ones that have the potential to improve performance, which supports the original objective of internetworking heterogeneous technologies such as 3G/UMTS and IEEE 802.11 networks.

8.2 FUTURE WORK

In this study, a WLAN coverage database was employed to provide coverage information stored up in the database as WLAN coverage footprint. This concept no doubt impacts on the access time to WLAN network. The addition of the database makes the overall WLAN system more granular. Therefore, future study may concentrate on studying the tradeoff and the balance likely to produce the best results in terms of making accurate decision within an acceptable amount of time. Furthermore, research is needed to determine the best method for assessing the value of the
“mean handover latency” associated with completing the handover procedures between an inter-networked UMTS/WLAN pair. This value was needed in the calculation of min required visit duration, which was used in the computation of TEP coordinates. Therefore, increasing the accuracy of the RVD value would result in an increase in the accuracy of the location-aided handover decision. In most location-based services, including the algorithm/service described in this research effort, there is a need to obtain and store (at least temporarily) the user’s position for processing at the network. Furthermore, such information as well as other location based information, which is typically stored in a database, must be shared among different network access providers/operators in order to support location-based services on large scale. This requires strict security measures in order to ensure the user’s privacy and minimize the risk associated with user tracking. Therefore, future work is needed to establish the requirements and measures needed to support location-based services.

Finally, this research focused on mitigating a handover-related issue by using location-based information. This is only one aspect of the benefits that could results from utilizing location-based information in the wireless domain. As the accuracy, reliability, and availability of location-based information increases with the increase in user demand, such information can be used to improve other aspect of wireless communication. For instance, future work is needed to study the potential and benefit of using location-based information to improve coverage discovery in homogeneous/heterogeneous networks, heterogeneous network load-balancing, and heterogeneous network/cell design and planning.
REFERENCES

1. THEODORE S RAPPAPORT “Wireless Communications Principles and Practices” second edition

2. JOCHEN SCHILLER “Mobile Communications” second edition.


4. Lane Pool “Cellular Communication Explained From Basic to 3G”


10. Richard A. Good; “Implementing Vertical Handovers between IEEE 802.11 and UMTS Cellular Network”, Department of Electrical Engineering, University of Cape Town


13. 3GPP TS 24.030: “Stage 3 Location Services (LCS); Supplementary service operations”, Version 9.0.0, Release 9, December 2009


