AN INTELLIGENT VERTICAL HANDOFF DECISION ALGORITHM IN NEXT GENERATION WIRELESS NETWORKS

by

Yaw Nkansah-Gyekye

A dissertation submitted in fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY in the Department of Computer Science, University of the Western Cape

Supervisor: Prof. Johnson I. Agbinya

May 2010

Declaration

I declare that *An Intelligent Vertical Handoff Decision Algorithm In Next Generation Wireless Networks* is my own work, that it has not been submitted before for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged as complete references.

Full Name: Yaw Nkansah-Gyekye

Signature: ……………………….

Date: May 2010

Abstract

Seamless mobility is the missing ingredient needed to address the inefficient communication problems faced by the field workforces of service companies that are using field workforce automation solutions to streamline and optimise the operations of their field workforces in an increasingly competitive market place. The key enabling function for achieving seamless mobility and seamless service continuity is seamless handoffs across heterogeneous wireless access networks. A challenging issue in the multi-service next generation wireless network (NGWN) is to design intelligent and optimal vertical handoff decision algorithms, beyond traditional ones that are based on only signal strength, to determine when to perform a handoff and to provide optimal choice of access network technology among all available access networks for users equipped with multimode mobile terminals. The objective of the thesis research is to design such vertical handoff decision algorithms in order for mobile field workers and other mobile users equipped with contemporary multimode mobile devices to communicate seamlessly in the NGWN. In order to tackle this research objective, we used fuzzy logic and fuzzy inference systems to design a suitable handoff initiation algorithm that can handle imprecision and uncertainties in data and process multiple vertical handoff initiation parameters (criteria); used the fuzzy multiple attributes decision making method and context awareness to design a suitable access network selection function that can handle a tradeoff among many handoff metrics including quality of service requirements (such as network conditions and system performance), mobile terminal conditions, power requirements, application types, user preferences, and a price model; used genetic algorithms and simulated annealing to optimise the access network selection function in order to dynamically select the optimal available access network for handoff; and we focused in particular on an interesting use case: vertical handoff decision between mobile WiMAX and UMTS access networks. The implementation of our handoff decision algorithm will provide a network selection mechanism to help mobile users select the best wireless access network among all available wireless access networks, that is, one that provides always best connected services to users.

KEY WORDS – Field workforce, Seamless mobility, Seamless communication, Vertical handoff, Vertical handoff decision, Intelligent and optimal vertical handoff decision algorithm, Next generation wireless network, Multimode mobile terminal, Access network selection, Fuzzy logic, Fuzzy multiple attributes decision making, Mathematical optimization, Genetic algorithm, Simulated annealing.

Acknowledgments

The quest for a Ph.D. degree has indeed been a long journey. I am eternally grateful to God for providing me all the blessings of life to complete the research.

My sincere thanks go to my supervisor, Professor Johnson Ihyeh Agbinya, for giving me the opportunity to conduct my PhD research at the University of the Western Cape. I am grateful to him for his guidance, steadfast support and encouragement.

I acknowledge the financial support I received during the early years of the research work from the South Africa National Research Foundation and Telkom South Africa Limited.

I appreciate the support and help I received from Professor I M Venter, Mr. Michael Norman, the entire academic and non-academic staff of the Computer Science department, and a good number of staff in other departments and offices at the University of the Western Cape which made it possible to do the research. I acknowledge the friendship of my fellow PhD students in the Computer Science department.

I am grateful to several friends including Professor James Darkwa (and his family), Professor Francis Benyah, Mr. Anthony Duah, Mr. Johnson Ampoful and others for their diverse support and assistance during my stay in South Africa. I will definitely miss the sumptuous lunches and dinners, the moral support, and the fruitful discussions we had on several occasions.

I am finally grateful to my parents, my brothers and sisters, my wife Yvonne, my children, and other relatives for their love, warmth, sacrifices and support for my education. Without them this research work would never have come into existence.

Publications

Conference Proceedings

Y. Nkansah-Gyekye and J. I. Agbinya, “Vertical Handoff Decision Algorithms Using Fuzzy Logic”, *Proceedings of the First Australian Conference on Wireless Broadband and Ultra Wideband Communications (AusWireless 2006)*, Sydney, Australia, 13-16 March 2006.

Y. Nkansah-Gyekye and J. I. Agbinya, “Vertical Handoff between WWAN and WLAN’, *Proceedings of the 5th International Conference on Networking (ICN 2006)*, Mauritius, 23-28 April 2006.

Y. Nkansah-Gyekye and J. I. Agbinya, “Vertical Handoff Decision Using Fuzzy Logic’, *Proceedings of Southern African Telecommunication Networks & Applications Conference* (*SATNAC 2006)*, Western Cape, South Africa, 3-6 September 2006.

Y. Nkansah-Gyekye and J. I. Agbinya, “Vertical Handoff Decision Algorithm for UMTS-WLAN Using Fuzzy Logic”, *Proceedings of the 2nd Australian Conference on Wireless Broadband and Ultra Wideband Communications (AusWireless 2007)*, Sydney, Australia, 27-30 August 2007.

Y. Nkansah-Gyekye and J. I. Agbinya, “Vertical Handoff Decision Algorithm Based on Fuzzy Logic and Genetic Algorithm”, *Proceedings of Southern African Telecommunication Networks & Applications Conference (SATNAC 2008)*, Wild Coast Sun, Eastern Cape, South Africa, 7-10 September 2008.

Y. Nkansah-Gyekye and J. I. Agbinya, “An Optimum Vertical Handoff Decision Algorithm for UMTS-WiMAX”, *Proceedings of the 2nd IFIP Symposium on Wireless Communications and Information Technology in Developing Countries (WCITD 2008)*, Pretoria, South Africa, 6-7 October 2008.

Y. Nkansah-Gyekye and J. I. Agbinya, “A Vertical Handoff Decision Algorithm for Next Generation Wireless Networks”, *Proceedings of the 3rd International Conference on Broadband Communications, Information Technology & Biomedical Applications (BroadCom 2008)*, Pretoria, South Africa, 23-26 November 2008.

Book Chapter

Y. Nkansah-Gyekye and J. I. Agbinya, “Vertical Handoffs in Fourth Generation Wireless Networks” in “*Advances in Broadband Communication and Networks*”, Editors: J. I. Agbinya, O.Sevimli, S. Lal, S. Selvadurai, A. Al-Jumaily, Y. Li, and S. Reisenfeld, River Publishers, Denmark, 2008, ISBN: 978-8792329004.

Abbreviations

2G – Second Generation

3G – Third Generation

3GPP – Third Generation Partnership Project

4G – Fourth Generation

AAA – Authentication, Authorization, and Accounting

ABC – Always Best Connected

ACK – Acknowledgment

AHP – Analytic Hierarchy Process

ANDSF – Access Network Discovery and Selection Function

ANS – Access Network Selection

ANSF – Access Network Selection Function

AP – Access Point

ASCONF – Address Configuration

BS – Base Station

BT – British Telecommunications

BWA – Broadband Wireless Access

CH – Correspondent Host

CN – Correspondent Node

DAR – Dynamic Address Reconfiguration

DHCP – Dynamic Host Configuration Protocol

FA – Foreign Agent

FIS – Fuzzy Inference System

FL – Fuzzy Logic

FMADM – Fuzzy Multiple Attributes Decision Making

FWA – Field Workforce Automation

GA – Genetic Algorithm

GRA – Grey Relational Analysis

GPRS – General Packet Radio Service

HA – Home Agent

HIA – Handoff Initiation Algorithm

HM – Handoff Management

IP – Internet Protocol

MA – Mobility Agent

MADM – Multiple Attributes Decision Making

MCDM – Multiple Criteria Decision Making

MCHO – Mobile-Controlled Handoff

MEW – Multiplicative Exponent Weighting

MF – Membership Function

MIP – Mobile IP

MODM – Multiple Objective Decision Making

MT – Mobile Terminal

mSCTP – Mobile Stream Control Transmission Protocol

NAHO – Network-Assisted Handoff

NGWN – Next Generation Wireless Network

NCHO – Network-Controlled Handoff

NIM – Network Interface Management

NN – Neural Network

QoS – Quality of Service

R – Real Line

RAT – Radio Access Technology

RSS – Received Signal Strength

RSSI – Received Signal Strength Indication

SA – Simulated Annealing

SAW – Simple Additive Weighting

SCTP – Stream Control Transmission Protocol

SIP – Session Initiation Protocol

T1 FS – Type-1 Fuzzy Set

T2 FS – Type-2 Fuzzy Set

TCP – Transmission Control Protocol

TOPSIS – Technique for Order Preference by Similarity to Ideal Solution

UDP – User Datagram Protocol

UMTS – Universal Mobile Telecommunications System

VHD – Vertical Handoff Decision

VHDA – Vertical Handoff Decision Algorithm

VoIP – Voice-over-IP

Wi-Fi – Wireless Fidelity

WiMAX – Worldwide Interoperability for Microwave Access

WLAN – Wireless Local Area Network

WMAN – Wireless Metropolitan Area Network

WON – Wireless Overlay Network

WPAN – Wireless Personal Area Network

WWAN – Wireless Wide Area Network

Table of Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

This chapter presents the background of and motivation for the thesis. An overview of seamless mobility in the next generation wireless network is given. The chapter explains the problem statement and objective, and provides the thesis outline.

1.1 Background and Motivation

The business world is being transformed by workforces who are becoming increasingly mobile, are away from their desks regularly and use mobile computing devices to perform their duties in various locations. Faced with the need for higher worker productivity, the pressure to reduce operating costs, and increasing customer demand for faster and efficient service, telecommunications companies and other companies providing field service operations are looking for ways to streamline and optimise the operations of their field workforces in an increasingly competitive market place. Consequently, these companies have started to automate their field workforces through intelligent field workforce automation (FWA) solutions including scheduling and dispatch components as a way of providing a platform for making internal communication more efficient, reducing cost and improving quality of service.

Field workforce automation (FWA) is about getting the right engineer, technician, or skilled personnel to the right place at the right time with the right equipment/information in an optimum manner and least cost to the company [1, 2]. Among other functions, the FWA software automates the work allocation, the work dispatch, and the work monitoring of the field workforce.

The FWA software becomes a full field workforce automation solution that delivers business benefits when it is combined with the appropriate mobile devices, seamless mobility in a next generation wireless network that provides seamless communications and a variety of services for both the location and business requirements, and an effective change management programme [3].

Telkom South Africa Limited provides telecommunications products and services to millions of business and residential customers across South Africa, as well as maintaining and repairing the enormous network infrastructure that backs up those services. There are currently thousands of field technicians in Telkom, whose daily duties are mainly the fixing of telephone lines, repairing networking faults, and service provision to customers. Just about seven years ago, each field technician used a Husky hand-held computer to make dial-up connections to Web Force, Telkom's national workforce management tool, either from a landline or through a cellular phone and data cable. This procedure resulted in an astronomical amount of unproductive time spent, as well as very high billing and operational costs, as evaluated by Telkom. Although Telkom has installed an improved automated workforce management system, and upgraded devices for the field technicians, communication between the field technicians and the national workforce management tool is still not very efficient.

The problems associated with the field workforce management of Telkom South Africa Limited are similar to field workforce problems facing telecommunications companies and other service industries worldwide. A top priority of many service organizations is ensuring the robustness of the connection between the field workforce and the back office. This is due to many enterprises continuing to struggle with inadequate data management resulting from manual and paper-based procedures in the field, and high operational costs and unproductive time spent by the mobile workforce due to inefficient communication between the field workforce and the back office. As a result, more and more companies are automating their mobile workforces with field workforce automation solutions, or workforce management systems, as a way to improve connectivity to the field workforce and reduce reliance on manual processes. These companies realise that linking the mobile workforce with the enterprise and its data resources is key to enhancing productivity, accuracy, profitability and, ultimately, customer satisfaction. For instance, in 1992, British Telecommunications (BT) launched Work Manager [1] – an information system automating work-management and field communication processes – in order to allocate work more efficiently for the thousands of field engineers BT employs across the United Kingdom to maintain networks, repair faults and provide service to customers. Work Manager was later marketed as ***a.p.solve****'s* intelligent fieldforce management called *TASKFORCE* [4].

The competitive business climate of service organizations demands a field workforce that can provide efficient customer services anytime, solve problems on the go, and stay connected anywhere, anytime with seamless communications. Therefore, the field workforce must be provided with the tools and applications they need to be able to work anywhere, anytime, with the same access, availability and services as if they were back at their office desks. True mobility means having the freedom to communicate and access information using one's device of choice.

Mobile handheld devices and mobile applications are transforming the ability of companies to keep in touch with and manage their remote field service personnel, but running a smooth operation involves a lot more – including having the right field workforce automation software (or enterprise management systems) back at base, wireless field applications accessible via mobile devices, and the ability to provide a seamless mobility experience that enables seamless communications for the field workforce in a next generation wireless network. However, a major issue that is often overlooked is the provision of seamless mobility, and thus seamless communications, for an efficient mobile field workforce.

The research work presented in this thesis started in response to a call by Telkom South Africa Limited to examine and provide a solution to the inefficient communication problems between the field technicians and the national workforce management tool. The key objective of our "IP-centric Field Workforce Automation" research project is to ensure global access and mobility between heterogeneous wireless access networks, and to allow a user to use one mobile device across many types of networks. We want to ensure that users equipped with multimode mobile terminals (MTs) in a next generation wireless network (NGWN) environment will experience seamless mobility and enjoy *seamless communications* and ubiquitous access to applications in an *always best connected* (ABC) mode that employs the most efficient combination of available access systems. Seamless communication involves the ability of the MT to successfully or simultaneously attach to different points of attachment in the NGWN infrastructure in a way that makes the physical movement transparent and preserves application-level connectivity unaltered.

The rapid evolutions in both wireless broadband technologies and handheld devices are driving an evolution towards seamless communications and seamless services delivery with ubiquitous and high end-user mobility across an overlay of wireless access technologies in the next generation wireless networks. The next generation of cellular/wireless communications is expected to be purely Internet Protocol (IP)-based and consist of a converged core network and multiple wireless access networks to deliver IP-based services (including voice, video, multimedia, and data) for multimode MTs while moving between the various access technologies and roaming between various operator networks.

The evolving NGWN will seamlessly integrate various types of existing and emerging wireless access networks that can be categorized to their coverage areas as wireless personal area networks (such as ultra wideband and Bluetooth), wireless local area networks (such as the IEEE 802.11x family), wireless metropolitan area networks (WMANs) (such as the IEEE 802.16 / WiMAX (Worldwide Interoperability for Microwave Access)), wireless wide area networks (such as existing 2G/3G cellular access technologies (e.g., General Packet Radio Service (GPRS) and Universal Mobile Telecommunications System (UMTS)), and Long-Term Evolution that will offer higher data rates), and regional/global area networks including radio and television broadcasting, and satellite communications.

These heterogeneous wireless access networks typically differ in terms of signal strength, coverage area, data rate, latency, and packet loss rate. Therefore, each of them is practically designed to support a different set of specific services and handheld devices. The limitations of these complementary wireless access networks can be overcome through the integration of the different technologies into a single unified platform (that is, a NGWN system) that will empower mobile users equipped with multimode MTs to be connected to the NGWN system using the best available access network that suits their needs. Wireless networks make it possible for fixed and mobile users to access a variety of services.

In the integrated NGWN it is important to provide a seamless mobility experience to the user (i.e., to make the transition from one access network to another as transparent as possible to the user). Seamless mobility support is the basis of providing seamless communications and uninterrupted services to mobile users roaming between various wireless access technologies in a heterogeneous NGWN environment. In other words, users can realize the goal of seamless communications and seamless services delivery through achieving the seamless mobility objective. With seamless mobility, users can exploit all the access technologies to best meet their service costs and QoS requirements by automatically selecting the best access network through changing weight factors and constraints in a single objective optimization function.

Therefore, we are challenged, at the highest level, to provide seamless mobility for the field workforce of service companies such as Telkom SA using contemporary multimode mobile handheld devices in an integrated WLAN-WiMAX-UMTS heterogeneous wireless access network so that the FWA solution becomes a full field workforce automation solution that delivers business benefits. This challenge constitutes the core motivation of the thesis. Of course this challenge is very difficult as some key issues need to be considered before achieving the seamless mobility experience.

Seamless mobility can be achieved by enabling a multimode MT to conduct a seamless handoff with low latency and minimal packet loss across the heterogeneous wireless access networks, for example, across mobile WiMAX and 3G access networks, seamlessly transferring and continuing its ongoing session from one access network to the best available access network.

Set against this background, our primary motivation can now be stated concisely. **To resolve the problem of inefficient communication between the field workforce and the central communication system of a telecommunications company or another service industry, we shall design an intelligent and optimal vertical handoff algorithm for mobile field workers equipped with contemporary multimode mobile devices to communicate seamlessly with their central company management system and be provided with always best connected services in an integrated heterogeneous WLAN-WiMAX-UMTS wireless network system.**

1.2 Overview of Seamless Mobility in a Next Generation Wireless Network

Field workforces (or mobile workers) require a seamless, real-time and intuitive mobility experience and seamless service continuity when communicating with the back office and accessing business applications using appropriate mobile devices in a heterogeneous access network environment.

The next generation wireless network (NGWN) is expected to allow a multimode mobile terminal to simultaneously access multiple networks for a variety of services (or applications) with an appropriate quality of service. The key characteristics of the NGWN include [5, 6, 7]:

- Multiple networks: The evolving next generation wireless network (NGWN) will seamlessly integrate various types of existing and emerging wireless access networks with their corresponding radio access technologies (RATs) since no single RAT is able to optimally cater for all the different wireless communications scenarios. The access networks include wireless personal area networks (such as ultra wideband and Bluetooth) that provide range-limited ad hoc wireless service to users; wireless local area networks (such as the IEEE 802.11x family) that are designed to provide wireless access in areas with cell radius up to hundred meters; wireless metropolitan area networks (WMANs) (such as the IEEE 802.16 / WiMAX (Worldwide Interoperability for Microwave Access) that provide broadband connectivity to both fixed and mobile users in a WMAN environment); wireless wide area networks (such as existing second generation / third generation (2G/3G) cellular access technologies (e.g., General Packet Radio Service (GPRS) and Universal Mobile Telecommunications System (UMTS)) that provide long-range cellular voice and limited-throughput data services to users with high mobility, and Long-Term Evolution that will offer higher data rates); and regional/global area networks including radio and television broadcasting, and satellite communications.

- Multiple services: Users must be able to access, with seamless mobility and quality of service guarantee, a variety of services, or applications including voice, messaging, e-mail, Internet fax, information services, and multimedia.
- Multimode mobile terminal (MT): Intelligent multimode mobile terminals are being designed. The intelligent multimode MT must be equipped with multiple radio interfaces to access different access technologies. A multimode MT may include interfaces for WLAN, WiMAX, UMTS, and so on.

The objective of achieving seamless integration of the heterogeneous wireless access networks in the NGWN is to offer seamless service continuity and a seamless mobility experience to the user (that is, to make the transition from one access network to another as transparent as possible to the user). The seamless mobility experience preserves the application-level connectivity unaltered and conceals the heterogeneity of the NGWN that is conceived as an intelligent system capable of manipulating its available resources to provide the best service to the user without any user intervention. While this kind of experience is typically provided by all terrestrial wireless technologies in homogeneous environments, for example, by all kinds of 3G networks, which facilitate the mobility of a user across 3G cells in a fashion transparent to the user, the NGWN will need to extend this capability to heterogeneous architectures that encompass multiple access technologies such as mobile WiMAX and UMTS.

The vision of seamless mobility and seamless service continuity is the ability for users equipped with multimode mobile terminals to remain connected while roaming across different access networks in an NGWN. Seamless mobility envisions easy, universal, uninterrupted access to communication and other services anytime, anywhere for users across different access networks and locations. In a heterogeneous NGWN environment, seamless mobility support is the basis of providing seamless communications and uninterrupted always best connected services to mobile users roaming between various wireless access networks. In other words, users can realize the goal of seamless communications and seamless always best connected service delivery through achieving the seamless mobility objective. With seamless mobility, users can exploit all the access technologies to best meet their service costs and quality of service (QoS) requirements by automatically selecting the best access network through changing weight factors and constraints in a single objective optimization function [8, 9, 10]. Seamless mobility provides always best connected services to the mobile user.

There are several issues to consider before an integrated heterogeneous wireless access system in the NGWN can provide the vision of seamless mobility, including: maintaining connectivity for users on the go; choosing among different access network types based on network characteristics, services offered, need, and user preferences; power management to prolong battery life of multimode devices; and dynamic spectrum allocation.

Addressing these issues require key functionalities that need to be performed, including: monitoring of the current serving network, access network discovery, handoff decision, and handoff execution.

Seamless mobility and seamless service continuity can be achieved by enabling mobile terminals to conduct seamless handoffs with low latency and minimal packet loss across heterogeneous wireless access networks, for example, across mobile WiMAX and 3G (such as UMTS) access networks, seamlessly transferring and continuing their ongoing sessions from one access network to the best available target access network. In other words, seamless handoff is the key enabling function for seamless mobility and seamless service continuity among the heterogeneous wireless technologies in the NGWN. Users have experienced seamless horizontal handoff during cell phone calls when connected to single cellular network interfaces for several years now. In the future, however, seamless vertical handoff should become commonplace when MTs are connected to multiple interfaces of the NGWN. A seamless handoff is typically characterized by two performance requirements:

- The handoff latency should be low and no more than a few hundreds of milliseconds.
- The QoS provided by the source and target access networks should be nearly identical in order to sustain the same communication experience.

These two performance requirements are not trivial to satisfy when two or more heterogeneous access networks are combined in a single architecture. Soft handoff can eliminate handoff latency and instability when a multimode MT is used. In order to offer seamless handoff, a crucial issue to be addressed is the provision of efficient vertical handoff algorithms covering the three sequential phases of the handoff process: access network discovery, handoff decision, and handoff execution. Vertical handoff decision is the main study of this thesis.

1.3 Problem Statement and Objective

The main research problem is to address the inefficient communication problems between the field workforces and the workforce management tools of service companies that are automating their field workforces through intelligent field workforce automation (FWA) solutions as a way of providing a platform for making internal communication more efficient, reducing cost and improving quality of service.

The objective of our research project is to offer seamless service continuity and a seamless mobility experience to the mobile worker (that is, to make the transition from one access network to another as transparent as possible to the worker) so that the FWA software becomes a full field workforce automation solution that delivers business benefits. Even though service companies are rolling out field workforce automation solutions, seamless mobility is the missing ingredient needed to create an effective field (or mobile) workforce as mobility continues to play a key role in helping organizations improve customer service, increase productivity, and reduce costs.

Seamless mobility and seamless service continuity can be achieved by enabling mobile terminals to conduct seamless handoffs with low latency and minimal packet loss across heterogeneous wireless access networks, for example, across mobile WiMAX and UMTS access networks, seamlessly transferring and continuing their ongoing sessions from one access network to another.

Therefore, our main research objective is to design an intelligent and optimal vertical handoff decision algorithm in order for mobile field workers and other mobile users equipped with contemporary multimode mobile devices to communicate seamlessly in the NGWN system. In order to tackle this research objective, the study addresses the following issues:

1. The challenges such as imprecision and complexity constraints that characterize the handoff decision process.
2. The design of a suitable handoff initiation algorithm that can handle imprecision and uncertainties in data and process multiple vertical handoff initiation parameters (criteria).
3. The design of a suitable access network selection function that can handle a tradeoff among many handoff metrics including quality of service (QoS) requirements (such as network conditions and system performance), mobile terminal conditions, power requirements, application types, user preferences, and a price model.
4. Optimisation of the access network selection function in order to dynamically select the optimal available access network for handoff.
5. Provision of a network selection mechanism to help mobile users select the best wireless access network among all available wireless access networks, that is, one that provides always best connected services to users.

In order to address these research issues, we:

- used fuzzy logic and fuzzy inference systems to design a suitable handoff initiation algorithm that can handle imprecision and uncertainties in data and process multiple vertical handoff initiation parameters (criteria);
- used the fuzzy multiple attributes decision making method and context awareness to design a suitable access network selection function that can handle a tradeoff among many handoff metrics including quality of service requirements (such as network conditions and system performance), mobile terminal conditions, power requirements, application types, user preferences, and a price model;
- used genetic algorithms and simulated annealing to optimise the access network selection function in order to dynamically select the optimal available access network for handoff, that is, one that provides always best connected services to users; and
- focused in particular on an interesting use case: vertical handoff decision between mobile WiMAX and UMTS access networks.

1.4 Thesis Outline

This thesis is organised as follows.

In Chapter 1, the background and motivation for the research is presented with the research problem.

Chapter 2 gives a brief introduction to the next generation wireless networks, and seamless mobility. The chapter also presents an overview of vertical handoff in an NGWN and a review of vertical handoff related literature.

In Chapter 3, fuzzy logic and fuzzy inference system are presented with our proposed seamless vertical handoff algorithm in the next generation wireless network. A key phase of the algorithm is the vertical handoff decision algorithm which is composed of two parts: handoff initiation, and access network selection. Finally, the chapter discusses the handoff initiation algorithm using fuzzy inference systems.

Chapter 4 first presents the fuzzy multiple attribute decision making (FMADM) method as a suitable model for making decisions under uncertainty. The FMADM method is then used to develop the access network selection component of the vertical handoff decision algorithm studied in the thesis.

Chapter 5 introduces the notion of optimisation and defines formally the combinatorial, continuous, and mixed-variable optimisation problems. It also explains how the characteristics of these problems impact the types of algorithms used to tackle them. A review of the genetic algorithm (GA) is given in this chapter. Then the GA is used to optimize the access network selection function of the access network selection component of the vertical handoff decision algorithm with the aim of selecting the optimal available access network for handing off.

A review of the simulated annealing (SA) method is presented in Chapter 6. Then the SA is used to optimize the access network selection function of the access network selection component of the vertical handoff decision algorithm. Finally, the Chapter presents a performance comparison of the GA and the SA.

A summary of and the major contributions of the thesis report and future work are presented in Chapter 7.

CHAPTER 2

VERTICAL HANDOFFS IN NEXT GENERATION WIRELESS NETWORKS

This chapter begins with an introduction, and then gives an overview of the next generation wireless networks and an overview of seamless mobility. Handoff in the next generation wireless networks is presented followed by a review of related literature on vertical handoff.

2.1 Introduction

Service providers such as telecommunications companies are using field workforce automation solutions to streamline and optimise the operations of their field workforces. The provision of seamless mobility, and thus seamless communications, for an efficient mobile field workforce equipped with appropriate multimode mobile devices in a next generation wireless network is the missing component needed to transform the field workforce automation software into a full workforce automation solution. Field workforces and other mobile users require a seamless, real-time and intuitive mobility experience and seamless service continuity when communicating with the back office and accessing business applications using appropriate multimode mobile devices in a heterogeneous access network environment.

The rapid evolutions in both wireless broadband technologies and handheld devices and the growing consumer demand for communication services anywhere and anytime are driving an evolution towards the seamless integration of heterogeneous wireless access networks with their corresponding radio access technologies (RATs) in the next generation wireless networks (NGWNs) with the objective of offering seamless service continuity and a seamless mobility experience to the user, that is, to make the transition from one access network to another as transparent as possible to the user. The NGWN is proposed to integrate the multiple RATs in a common network that can ease the shortage of radio resources facing users today since the available radio resources of a single RAT are still far from satisfying the requirements of increasingly mobile services [10]. The NGWN will provide significantly higher data rates, and offer a variety of services and applications. The NGWN is expected to allow a multimode mobile terminal to simultaneously access multiple networks for a variety of services with appropriate qualities of service and to choose the most suitable access network among the available access networks. In other words, a main feature of the NGWN is to provide always best connected services to users, that is, users can choose the best available access networks in a way that suit their needs, and to change to another best network if conditions change.

Mobile users, including field workforces, can realize the goal of seamless communications and seamless always best connected service delivery through achieving the seamless mobility objective. Seamless communication involves the ability of the MT to successfully or simultaneously attach to different access points in the NGWN infrastructure in a way that makes the physical movement transparent and preserves application-level connectivity unaltered. The NGWN mobile users will enjoy seamless mobility and ubiquitous access to applications in an *always best connected* (ABC) mode that employs the most efficient combination of available access systems. The goal of the mobile user is to be always best connected, that is, to be not only connected, but also connected through the best available device and access technology at all times [11].

Seamless mobility and seamless service continuity can be achieved by enabling multimode mobile terminals to conduct seamless handoffs with low latency and minimal packet loss across heterogeneous wireless access networks, for example, across mobile WiMAX and UMTS access networks, seamlessly transferring and continuing their ongoing sessions from one access network to the best available target access network. A handoff algorithm is required to preserve connectivity as the mobile terminals move about, and at the same time curtail disturbance to ongoing transfers. Consequently, seamless handoff, with low latency and minimal packet loss, has become a crucial factor for field workforces and other mobile users who wish to receive continuous and reliable services. One of the challenging issues in the multi-service NGWN is to design intelligent and optimal vertical handoff decision algorithms, beyond traditional ones that are based on only signal strength, to determine when to perform a handoff and to provide an optimal selection of access network technology among all available access networks for users equipped with multimode mobile terminals. In the process, an intelligent and optimal vertical handoff decision algorithm will enable user applications to switch automatically between active interfaces that best suit them based on application requirements and interface capabilities. In this thesis we address the design of such vertical handoff decision algorithms, and we focus in particular on an interesting use case: vertical handoff decision between mobile WiMAX and UMTS access networks.

2.2 Next Generation Wireless Networks

The next generation wireless networks (NGWNs) will be based on a heterogeneous infrastructure comprising different wireless (and wired) access networks with their corresponding radio access technologies in a complementary manner to provide ubiquitous high speed wireless connectivity to mobile terminals. Some key features of the NGWN include [5, 6, 7]:

- Integrated heterogeneous access networks with a common IP-based core.
- Support for telecommunication, data and multimedia services with low transmission cost.
- Use of multimodal devices capable of supporting various types of network access technologies.
- High usability: anytime, anywhere, and with any technology.
- User friendliness and user personalization.
- Support for integrated service access from various service providers.

The term *all-IP multi-access* networks together with multimode terminals and smart access selection mechanisms that allow the user to be in some sense *always best connected* (ABC) is often used as a synonym for next-generation systems. The term *always best connected* refers to the concept of defining a set of access selection criteria and mechanisms that allow users to get connected to various services in a nearly optimal manner [12]. For a person to be *always best connected* means that he or she is always connected through the best available device and access technology at all times [11].

The heterogeneous access networks, multiple services, and multimode terminals are discussed below.

2.2.1 Heterogeneous Wireless Access Networks

The next generation wireless network (NGWN) is expected to be purely IP-based and consist of a converged core network and multiple wireless access networks to deliver IP-based services (including voice, video, multimedia, and data) for multimode MTs while moving between the various access technologies and roaming between various operator networks.

The evolving NGWN will seamlessly integrate various types of existing and emerging wireless access networks that can be categorized to their coverage areas as:

- Wireless personal area networks (WPANs) – WPANs, such as ultra wideband and Bluetooth, provide range-limited ad hoc wireless service to users.
- Wireless local area networks (WLANs) – WLANs are designed to provide wireless access in areas with cell radius up to hundred meters, and are used mostly in home and office environments. They provide high-throughput connections for stationary/quasi-stationary wireless users without the costly infrastructure of 3G. The most notable WLAN standard is the IEEE 802.11x family. The IEEE 802.11a and IEEE 802.11g provide data rates up to 54 Mbps per access point, and the IEEE 802.11n provides up to 540 Mbps.
- Wireless metropolitan area networks (WMANs) – WMANs are designed to cover wider areas, generally as large as entire cities, with large numbers of LANs and WLANs. The IEEE 802.16 / WiMAX (Worldwide Interoperability for Microwave Access) intends to provide broadband connectivity to both fixed and mobile users in a WMAN environment. It provides data rates up to 50 Mbps. It is optimised to provide high-rate wireless connectivity for a large set of services and applications that require QoS guarantees.
- Wireless wide area networks (WWANs) – WWANs, such as existing 2G/3G cellular access technologies (e.g., General Packet Radio Service (GPRS) and Universal Mobile Telecommunications System (UMTS)) that provide long-range

cellular voice and limited-throughput data services to users with high mobility, and Long-Term Evolution that will offer higher data rates.

- Regional/global area networks – These include radio and television broadcasting, satellite communications, etc.

These heterogeneous wireless access networks typically differ in terms of signal strength, coverage area, data rate, latency, and loss rate. Table 2.1 shows the characteristic differences between 3G, Wi-Fi, and WiMAX. Therefore, each of them is practically designed to support a different set of specific services and devices. However, these networks will coexist and use a common IP core to offer services ranging from low-data-rate non-real-time applications to high-speed real-time multimedia applications to end users since the networks have characteristics that complement each other. The limitations of these complementary wireless access networks can be overcome through the integration of the different technologies into a single unified platform (that is, an NGWN system) that will empower mobile users equipped with multimode MTs to be connected to the NGWN system using the best available access network that suits their needs. For example, given the complementary characteristics of WLAN (faster data rate and short-distance access) and a 3G system such as UMTS (slower data rate and long-range access), it is compelling to combine them to provide ubiquitous broadband wireless access. Similarly, a mobile WiMAX can be overlaid onto a 3G system to provide QoS to wireless Internet users in a cost-effective manner even though 3G systems are designed primarily for mobile voice and data users while WiMAX systems are optimized to provide high-rate broadband wireless access for a large set of services and applications (e.g., with multimedia traffic) that require high security and QoS guarantees.

Cells of the heterogeneous access networks are overlaid within each other to form larger wireless overlay networks (WONs). A WON has a hierarchical structure with different levels [13]. Higher levels in the hierarchy cover a large area but provide lower bandwidth whilst lower levels are comprised of high bandwidth wireless cells that provide a smaller coverage area. WONs solve the problem of providing network connectivity to a large number of mobile users in an efficient and scalable way. The integration and internetworking of the heterogeneous wireless access networks in the

NGWN system requires the design of intelligent handoff and location management schemes to enable mobile users to switch network access and experience uninterrupted service continuity anywhere and anytime.

The main requirements for achieving seamless integration of the heterogeneous wireless access networks in the NGWN are to provide an appropriate AAA (authentication, authorization, and accounting) infrastructure, minimize handoff interruption, preserve the QoS as the MT moves between the current and the candidate network access technologies, and enable inter-operator roaming [9].

Table 2.1: *Comparison of 3G, Wi-Fi, and WiMAX*

Traits	3G	Wi-Fi	WiMAX
Standard	W-CDMA, CDMA2000	IEEE 802.11g	IEEE 802.16e
Objective	To provide voice and data service to mobile users	To provide BWA to fixed and nomadic users	To provide BWA to fixed and mobile users
Data rate	3 Mb/s	54 Mb/s	10-50 Mb/s
Coverage	30 m-20 Km	50-60 m	Up to 50 Km
Mobile speed	Up to 120 Km/h	Up to 1 Km/h	60 Km/h
Mobility	Full mobility functions	Fixed and nomadic	IP mobility
Cost	High	Low	Low
Advantages	Range, mobility	Speed, cheap	Speed, range, mobility

2.2.2 Multiple Services

Field workforces and other mobile users should be able to access a variety of services or applications with seamless mobility and quality of service (QoS) guarantee. The available services would include:

- Web browsing (Internet access) is the basic service that must be supported by the NGWN. In general, the QoS features (except bandwidth), such as delay, jitter, and packet loss, are not required for many web browsing applications. For basic Internet access, an essential security requirement is how to protect the confidentiality of messages between a correspondent node and a BS. Due to the nature of wireless communications, the wireless channel and wireless links are error prone and can fail due to various reasons. Therefore, it is important that the reliability is taken into account in the design and use of next generation wireless access networks.

- Data downloading may need a higher bandwidth especially if the content is a multimedia application and if a large amount of data is to be downloaded.
- Voice-over-IP (VoIP) is a real-time low-bandwidth application that is very sensitive to delay (and therefore requires low network latency) and jitter but can withstand some packet loss.
- Streaming application – Being a multimedia service, a streaming application requires a higher bandwidth than VoIP. Therefore, available bandwidth, transport cost, and current utilization are important factors. It is less vulnerable to delay and jitter than VoIP because of the ability to buffer longer duration of data before playback. Sensitivity to packet loss is similar to VoIP.

2.2.3 Next Generation Multimode Terminals

The evolution toward the NGWNs will necessitate a user-centric approach where users can access different access networks and services using a single device equipped with multiple radio interfaces such as WPAN, WLAN, WMAN (e.g., mobile WiMAX), and WWAN (e.g., UMTS). Terminals and devices capable of supporting different types of access technologies are being designed to eventually replace simple cellular phones, Personal Digital Assistants, Smart Phones, and all other current mobile devices. The next generation of mobile terminals includes devices capable of supporting multiple access systems by incorporating several interface cards and appropriate software for switching between multiple interface technologies. An intelligent multimode terminal should be able to decide autonomously the active interface that is best for an application session and to select the appropriate radio interface as the user moves in and out of the vicinity of a particular access technology. The decision regarding the switching of the interface and the handoff of the active sessions to the new active interface may be decided based on network conditions, QoS requirements of the running applications, and user preferences.

Requirements that need to be fulfilled in order to design intelligent multimode mobile terminals include [14]:

- The terminal should operate with minimum inputs from the user. From the perspective of a user experience, it is preferable to carry out these decisions in an automated manner rather than querying the user every time a new interface becomes available or an old interface disappears.
- Radio access interfaces should be selected based on network conditions, QoS requirements of applications, and user preferences.
- The requirements of applications should be determined and then a decision made whether an application could benefit from changing interfaces.
- Traffic should be balanced while changing the active interface in a way that is transparent to the user, that is, as seamlessly as possible.

The multimode mobile terminal must be capable of [15]:

- Detecting the availability of access networks;
- Finding, receiving and processing measurements regarding the characteristics of available access networks;
- Accessing, modifying and storing the user profile;
- Allowing the user to dynamically redefine his/her preferences; and
- Supporting the applications in seamlessly handing off the existing connections from one access network to another.

2.3 Seamless Mobility in Next Generation Wireless Networks

The availability of broadband wireless technologies has ushered in a new era of a unified communication infrastructure that will provide us with easy, universal, uninterrupted access to communication, information and entertainment when, where and how we want it. The objective of achieving seamless integration of the heterogeneous wireless access networks in the NGWN is to offer seamless service continuity and a seamless mobility experience to the mobile user (that is, to make the transition from one access network to another as transparent as possible to the user). From the user's perspective, the seamless mobility experience makes the physical movement transparent, preserves the application-level connectivity unaltered and conceals the heterogeneity of the NGWN that is conceived as an intelligent system capable of handling its available resources to provide the best service to the user without any user intervention [9]. While this kind of experience is typically provided by all wireless technologies in homogeneous environments, for example, by all kinds of 3G networks, which facilitate the mobility of a user across 3G cells in a fashion transparent to the user, the NGWN will need to extend this capability to the NGWN heterogeneous architecture that encompass multiple access technologies such as mobile WiMAX and UMTS.

The vision of seamless mobility and seamless service continuity is the ability for mobile users equipped with multimode mobile terminals to remain connected while roaming across different access networks and technologies in an NGWN and be served uninterruptedly with multiple media from multiple vendors. Seamless mobility envisions easy, universal, uninterrupted access to communication and other services anytime, anywhere for users across different access networks, devices and locations. Seamless mobility also means that our mobile devices and networks will understand us better as they will learn our preferences, sense the world around us, and use that knowledge to make life easier [16].

Achieving seamless mobility will transform industries and enhance convenience and productivity. It will allow everyone to communicate, interact with and collaborate on personal digital content that is customised according to user personal preferences, location, circumstances and availability. Seamless mobility will meet the needs of [16]:

- Consumers who are demanding personalised services that make it easy to access and share digital content when, where and how they want.
- Businesses that are demanding more support for their increasingly mobile workforces and customers who require anytime, anywhere communications and ubiquitous access to data, resources and applications. This will enhance productivity, increase customer response time and spark business innovation.

Some of the attributes of seamless mobility are [16]:

- Full mobility, that is, transparent connectivity across different heterogeneous access networks.
- Always on, always here capability, via sessions that cross access networks and mobile devices automatically and transparently.
- Seamless communication and seamless service delivery.
- Personalised, context-aware applications.
- User-centric content that is context- and device-sensitive.
- Flexible and friendly user-interface architectures and technologies that deliver intuitive, friendly and personal user experiences.
- Privacy, safety and security technologies to enable trusted communications.

In a heterogeneous NGWN environment, seamless mobility support is the basis of providing seamless communications and uninterrupted always best connected services to mobile users roaming between various wireless access networks. With seamless mobility, users can exploit all the access technologies to best meet their service costs and QoS requirements by automatically selecting the best access network through changing weight factors and constraints in a single objective optimization function [8, 9, 10].

There are several issues to consider before an integrated mobile WiMAX/UMTS system can provide the vision for seamless mobility, including [17]:

- Maintaining connectivity for users on the go.
- Choosing among different access network types based on network characteristics, services offered, need, and user preferences.
- Preserving key user-centric capabilities across boundaries between networks, devices and services.
- Configurable, power-efficient multimode mobile devices that can accommodate new wireless access technologies in a variety of frequency bands.
- Dynamic use of licensed and unlicensed spectrum.

Addressing these issues require key functionalities that need to be performed, including [5, 17, 18]:

- Monitoring of the current serving network – obtaining timely information and regular updates on the status of the current serving access network in order to minimise service disruption and optimise the quality of service.
- Network discovery – finding available access networks in the vicinity of the mobile terminal.
- Handoff decision – making the decision on whether to initiate a handoff to the access network(s) discovered and choosing the best target access network to connect to.
- Handoff execution – transferring of data packets to a new wireless link in order to reroute a mobile user's connection path to the new point of attachment.

Seamless mobility and seamless service continuity can be achieved by enabling mobile terminals to conduct seamless handoffs with low latency and minimal packet loss across heterogeneous wireless access networks, for example, across mobile WiMAX and 3G access networks, seamlessly transferring and continuing their ongoing sessions from one access network to the best access network among all available target access networks [9]. In other words, seamless handoff is the key enabling function for seamless mobility and seamless service continuity among the heterogeneous wireless technologies in the NGWN. Users have experienced seamless horizontal handoff during cell phone calls when connected to single cellular network interfaces for several years now. In the future, however, seamless vertical handoff should become commonplace when MTs are connected to multiple interfaces of the NGWN.

A seamless handoff is typically characterized by two performance requirements [9]:

- The handoff latency should be no more than a few hundreds of milliseconds.
- The QoS provided by the source and target access networks should be nearly identical in order to sustain the same communication experience.

These two performance requirements are not trivial to satisfy when two or more heterogeneous access networks are combined in a single architecture. Soft handoff can eliminate handoff latency and instability when a multimode MT is used. In order to offer seamless handoff, a crucial issue to be addressed is the provision of efficient, intelligent and optimal vertical handoff decision algorithms which is the study of this thesis.

2.4 Handoffs in the Next Generation Wireless Networks

Mobility management is a main challenge in the evolving multi-service heterogeneous NGWN. It consists of two components: location management and handoff management [19]. Location management tracks and locates the mobile terminal (MT) for successful information delivery. Handoff management maintains the active connections for roaming MTs as they change their point of attachment to the network. Seamless mobility can be achieved by enabling MTs to conduct seamless handoffs across the heterogeneous wireless access networks, for example, across mobile WiMAX and 3G access networks.

Each MT is at all times within range of at least one network point of attachment. In cellular voice telephony and cellular data networks, such a point of attachment is called a base station (BS), and in wireless local area networks, it is called an access point (AP). It is most beneficial for overall network performance that each MT is connected to the most suitable BS or AP while accessing services provided by the network. A criterion for determining which one of the points of attachment is the most suitable one for a given application at any particular location and moment of time is for the MT to select a BS or an AP that gives the maximum received signal strength. This selection fulfills the QoS requirements of that particular application and usually also maximizes the quality of the connection provided that no other signal degrading impairments exist.

When an MT moves away from a BS or an AP, the signal level degrades and there will be a need to switch communications to another BS or AP. *Handoff* is the mechanism by which an ongoing connection between an MT and a correspondent terminal is transferred from one point of attachment to the network to another [20]. That is, handoff is the mechanism by which an MT keeps its connection active when it migrates from the coverage area of one network attachment point to another.

Handoff is a complex process that involves several important issues such as control, methodology, algorithms, metrics, protocol, and performance measures. The focus of this thesis is the handoff decision algorithm and related metrics.

2.4.1 Handoff Classification

Handoffs can be classified using the network type involved into horizontal (intra-system) and vertical (inter-system) cases as an MT moves within or between different overlays of a WON. Figure 2.1 shows horizontal and vertical handoffs in a WON [21]; in this case, cell B and cell C are using the same network technology while cell A is using a different network technology.

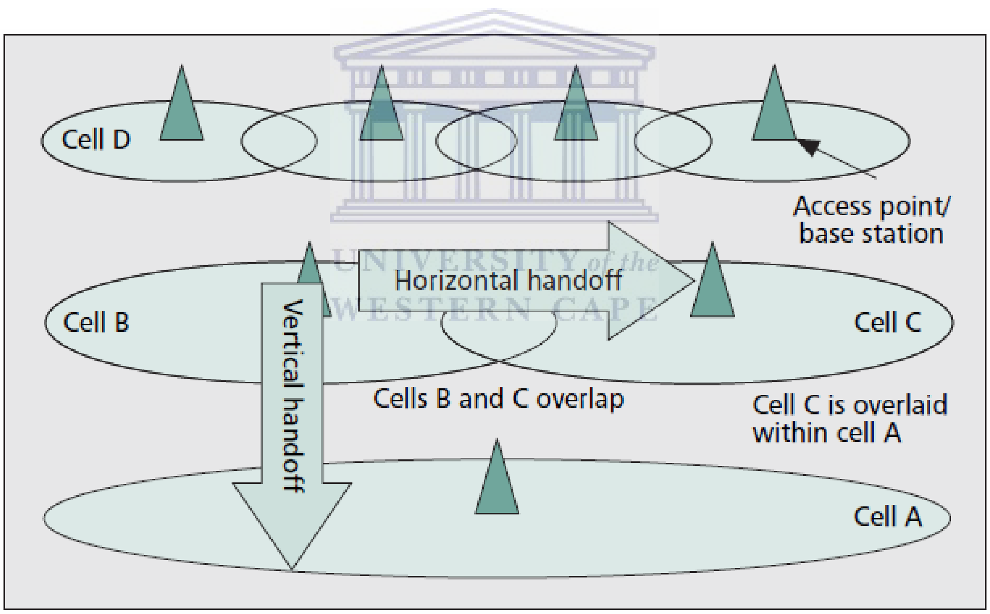


Figure 2.1: *Horizontal and Vertical Handoffs [21]*

Horizontal handoff or *intra-system handoff* is a handoff that occurs between the APs or the BSs of the same network technology. In other words, a horizontal handoff occurs between the homogeneous cells of a wireless access system. For example, the changeover of signal transmission of an MT from an IEEE 802.16e BS to a geographically neighbouring IEEE 802.16e BS is a horizontal handoff process. The network automatically exchanges the coverage responsibility from one point of attachment to another every time an MT crosses from one cell into a neighbouring cell supporting the same network technology. Horizontal handoffs are mandatory since the MT cannot continue its communication without performing it.

Vertical handoff or *inter-system handoff* is a handoff that occurs between the different points of attachment belonging to different network technologies. For example, the changeover of signal transmission from an IEEE 802.16e BS to the BS of an underlay 3G cellular network is a vertical handoff process. Thus, vertical handoffs are implemented across heterogeneous cells of wireless access systems, which differ in several aspects such as received signal strength (RSS), bandwidth, data rate, coverage area, and frequency of operation. The implementation of vertical handoffs is more challenging as compared to horizontal handoffs because of the different characteristics of the multiple access networks involved.

In general, there are two types of vertical handoff: upward and downward. An *upward* vertical handoff is a handoff to a wireless underlay with a larger cell size and lower bandwidth. A *downward* vertical handoff is a handoff to a wireless overlay with a smaller cell size and larger bandwidth. Thus, a mobile device performing an upward vertical handoff disconnects from a network providing smaller coverage area and higher access speed (for example, WLAN) to a new one providing broader coverage but lower access speed (for example, WWAN), while a mobile device performing a downward vertical handoff disconnects from a network providing broader coverage area and lower access speed to a new one providing limited coverage but higher access speed.

Handoffs can be characterized using the number of connections involved as soft or hard. A handoff is *hard* if the MT can be associated with only one point of attachment at a time. In other words, an MT may set up a new connection at the target point of attachment after the old connection has been torn down.

A *soft* handoff or a *make before break* handoff occurs if the MT can communicate with more than one point of attachment during handoff. In this case, the MT's connection may be created at the target point of attachment before the old point of attachment connection is released. For example, an MT equipped with multiple network interfaces can simultaneously connect to multiple points of attachment in different networks during soft handoff.

A *seamless handoff* means that the transition to the new network point of attachment is transparent to the user (that is, there is no perceptible service degradation). It is a handoff that is both *fast* (low handoff latency) and *smooth* (minimal packet loss).

In a *passive handoff* the user of a MT has no control over the handoff process. Passive handoff is the most common in 2G and 3G wireless networks.

In a *proactive handoff* the user is allowed to decide when to handoff. The handoff decision can be based on preferences specified by the user. This type of handoff is expected to feature prominently in 4G wireless networks.

2.4.2 Handoff Decision Mechanism

The handoff decision mechanism or handoff control may be centralized (that is, the handoff decision may be located in the MT itself (as in mobile data and WLANs) or in a network entity (as in cellular voice)). These cases are called mobile-controlled handoff (MCHO) (or terminal-controlled handoff) and network-controlled handoff (NCHO), respectively.

In NCHO, the network makes a handoff decision based on measurements of the RSSs of the MT at a number of BSs. Information about the signal quality for all users is available at a single point in the network that facilitates appropriate resource allocation. It is beneficial when the handoff decision is made by the network because [9]:

- The network can redirect the MT to another radio site or frequency that has enough capacity to handle its ongoing communications, and
- The network can also coordinate the mobility of all MTs in a way that overall traffic is evenly distributed across all radio resources, congestion is minimized, and total throughput is maximized.

The disadvantage of the NCHO is that the radio network may lack some parameters that impact the handoff decision, such as user preferences, the exact type of services active on the MT, and some operator policies pertaining to mobility between mobile WiMAX and 3GPP accesses.

In MCHO, the MT is completely in control of the handoff process. This type of handoff has a short reaction time (in the order of 0.1 sec). The MT itself first discovers all the available networks. It then measures the signal strengths from surrounding BSs and interference levels on all channels, and makes the evaluations for the handoff decision. A handoff can be initiated if the signal strength of the serving BS is lower than that of another BS by a certain threshold. The MCHO will be choice of the future as the telecommunications market migrates from a network-centric approach to a user-centric approach. Indeed, the 3GPP Release 8 specifications mandate that the MT make the decision for handoff between 3GPP and mobile WiMAX. The MCHO has some key advantages [9]:

- The MT is the entity that is aware of the different access technologies in its surroundings.
- The MT can make a handoff decision based on its up-to-date radio measurements, preconfigured user preferences, and all downloadable operator mobility policies,
- The MT does not need to send any inter-technology radio measurement to the network,
- The impact on the 3G and mobile WiMAX access networks is minimized because, for example, there is no need for 3G radio access to receive measurement reports for WiMAX cells and make decisions on handoff to WiMAX.

In network-assisted handoff (NAHO), the network assists the MT in the decision process by performing data collection and analysis. The MT can also provide its location and any other information that could be considered by the network in the analysis. The network only assists the MT in the decision process and the final decision is done by the MT.

2.4.3 Desirable Handoff Features

In the NGWN, an efficient handoff algorithm can achieve many desirable features by trading off different operating characteristics. Some of the major desirable features of a handoff algorithm are described below [21, 22, 23]:

Fast: A handoff algorithm should be fast so that the mobile device does not experience service degradation or interruption. Service degradation may be due to a continuous reduction in signal strength or an increase in co-channel interference (CCI). Service interruption may be due to a "break before make" approach of handoff.

Reliable: A handoff algorithm should be reliable. This means that the service should have good quality after handoff. Many factors help in determining the potential service quality of a candidate BS or AP. Some of these factors include received signal strength (RSS), signal-to-interference ratio (SIR), signal-to-noise ratio (SNR), and bit error rate (BER).

Communication quality: The communication quality should be maximized through minimizing the number of handoffs. Excessive handoffs lead to heavy handoff processing loads and poor communication quality. The more attempts at handoff, the more chances that a call will be denied access to a channel, resulting in a higher handoff call dropping probability.

Traffic balancing: The handoff procedure should balance traffic in adjacent cells, thus eliminating the need for channel borrowing, simplifying cell planning and operation, and reducing the probability of new call blocking.

Interference prevention: A handoff algorithm should minimise global interference. Transmission of bare minimum power and maintenance of planned cellular borders can help achieve this goal.

Context-awareness: A handoff algorithm should be context-aware. The algorithm should adapt to its surroundings and acquire and utilise user, mobile terminal, and network information to improve QoS, connectivity and maintain a high level of user satisfaction.

Multi-service: NGWNs will support multiple services such as voice, e-mail, and video streaming. Supporting multiple services in heterogeneous wireless networks is not only because it is the main reason for user transition and will enhance user satisfaction, but also because it will give network operators access to new levels of traffic and increase revenues for network and service operators. Therefore, handoff algorithms need to support multiple services.

Scalability: The recent increase in demand for multimedia services is likely to grow at a faster pace due to advances in multimedia distribution services. Therefore, handoff algorithms must support an expanded system capacity of heterogeneous wireless networks in terms of data rate, geographical coverage, and the number of subscribers supported. A handoff algorithm must also accommodate increase in size of individual wireless access networks and the integration of a large number of heterogeneous wireless access networks.

Simplicity: Implementation cost and the scalability issues require that a handoff algorithm should be as simple as possible. A simple handoff algorithm will not incur additional delay in the network since the algorithm will have a low computational overhead. However, an overly simple vertical handoff decision algorithm (VHDA) may not achieve high network resource utilization. For good QoS and efficient network resource utilization, a sophisticated VHDA is required to support multiple services, especially in a scenario where users are roaming across different access networks. Therefore there is a tradeoff between simplicity and the efficiency of VHDAs.

2.5 Handoff Process

Both horizontal and vertical handoff processes may be divided into three sequential phases [18]: network discovery, handoff decision, and handoff execution. Several handoff algorithms exist that can be used to complete the ongoing handoff process.

2.5.1 Network Discovery

Network discovery is the process where a mobile terminal (MT) equipped with multiple interfaces searches for reachable wireless access networks. As the multimode MT moves across the network, it has to discover other available access technologies in its surroundings which might be preferable to the currently used access network. For example, a multimode MT using a UMTS access network in an NGWN system needs to discover when mobile WiMAX access networks become available and possibly handoff to a mobile WiMAX if this is more preferable to the operator and/or user, or if the radio signal from its serving UMTS cell starts to deteriorate significantly. The network discovery phase collects information about the network, mobile devices, access points, and user preferences to be processed and used for making decisions in the handoff decision phase.

2.5.2 Handoff Decision

Handoff decision is the ability to decide when to perform the horizontal or vertical handoff and determine the best handoff candidate access network. It includes access network selection and drives the handoff execution. Suppose an MT that is using a UMTS cell has discovered its available neighbour WiMAX cells. The next issue is whether the MT needs to initiate a handoff to a discovered WiMAX cell. Several vertical handoff decision algorithms exist that can be used to make the correct decision to handoff the ongoing connection.

Handoff metrics are used to indicate whether or not a handoff is needed. In traditional horizontal handoffs which happen in homogeneous networks, only information obtained from the radio-link layer such as the RSS and channel availability are considered for handoff decisions. A handoff is made if the RSS from a neighbouring BS exceeds the RSS from the current BS by a predetermined threshold value.

A traditional handoff decision time algorithm uses the RSS measurement and optionally the threshold, hysteresis, or dwell timer to make the handoff decision as follows [20]:

- **Received signal strength:** The BS whose signal is being received with the largest strength is selected (choose BS B_{new} if $RSS_{new} > RSS_{old}$).
- **Received signal strength with threshold:** A handoff is made if the RSS of a new BS exceeds that of the current one and the signal strength of the current BS is below a threshold T (choose B_{new} if $RSS_{new} > RSS_{old}$ and $RSS_{old} < T$).
- **Received signal strength with hysteresis:** A handoff is made if the RSS of a new BS is greater than that of the old BS by a hysteresis margin H (choose B_{new} if $RSS_{new} > RSS_{old} + H$).
- **Received signal strength with hysteresis and threshold:** A handoff is made if the RSS of a new BS exceeds that of the current BS by a hysteresis margin H and the signal strength of the current BS is below a threshold T (choose B_{new} if $RSS_{new} > RSS_{old} + H$ and $RSS_{old} < T$).
- **Algorithm with Dwell timer:** Sometimes a dwell timer is used with the above algorithms. A timer is started the instant the condition in the algorithm is true. If the condition continues to be true until the timer expires, a handoff is performed.

In vertical handoffs, many network characteristics have an effect on whether or not a handoff should take place. Traditional handoff decision metrics based on the received signal strength indication (RSSI) and other physical layer parameters used for horizontal handoff in cellular systems are insufficient for the challenges of the next generation heterogeneous wireless systems. The RSS alone cannot be used for vertical handoff decisions because the RSSs of different networks cannot be compared directly due to the different characteristics of the overlay heterogeneous wireless networks involved. In order to perform intelligent handoff decisions in the next generation heterogeneous wireless environment and provide seamless vertical handoff, the following metrics are suggested in addition to the RSS [18, 24], as illustrated in Figure 2.2:

(a) *Quality of service*. Handing off to a network with better network conditions and higher performance would usually provide improved service levels.
   (i) *Network conditions*. Network-related parameters such as traffic, available bandwidth, network latency, and congestion (packet loss) may need to be considered for effective network usage.
   (ii) *System performance*. To guarantee the system performance and provision of improved service levels, a variety of parameters can be measured and employed in the handoff decision, such as the RSS, channel propagation characteristics, path loss, interchannel interference, signal-to-noise ratio (SNR), and the bit error rate (BER).
(b) *Cost of Service*. The cost of services offered is a major consideration to users since different network operators and service providers may employ different billing plans and strategies that may affect the user's choice of access network and consequently handoff decision.
(c) *Battery power*. Battery power may be a significant factor for handoff in some cases since wireless devices operate on limited battery power. For example, when the battery level decreases, handing off to a network with lower power requirements would be a better decision.
(d) *Security*. The ability of a network (including operator networks and corporate networks) to resist attack from software virus, intruders and hackers, and to

protect network infrastructure, services and confidentiality and integrity of customers' data is a major issue and could sometimes be a decisive factor in the choice of a network. The most significant source of risks in wireless networks is that the technology's underlying communications medium, the airwave, is open to intruders. A network with high encryption is preferred when the information exchanged is confidential.

(e) *Mobile terminal conditions*. Mobile terminal conditions include the screen size, portability/weight, performance (processing power, memory, and storage space), bandwidth requirements, networks supported, and dynamic factors such as velocity, moving pattern, and location information. The velocity attribute has a necessary effect and larger weight on vertical handoff decision than in horizontal handoff. Handing off to an embedded network in an overlaid architecture of heterogeneous networks is discouraged when traveling at a high speed since a handoff back to the original network will occur very shortly afterward when the mobile terminal leaves the smaller embedded network.

(f) *User preferences*. User preferences (such as preferred network operator, preferred technology type, preferred maximum cost) can be used to cater special requests for one type of network over another. For instance, if the target network to which a mobile node performs a handoff does not offer high security, the user may still decide to use the current network. Depending upon coverage, a user may wish to use a secure and expensive access network (such as UMTS) for his official e-mail traffic but may still opt for a cheaper network (for example, WLAN) to access web information.

(g) *Application types*. Different types of applications or services such as voice, data and multimedia applications require different levels of data rate, network latency, reliability, and security. Data-intensive applications such as video streaming will perform better when higher bandwidth is available. Real-time applications will need low network latency while non-real-time applications will not be so sensitive to network latency.

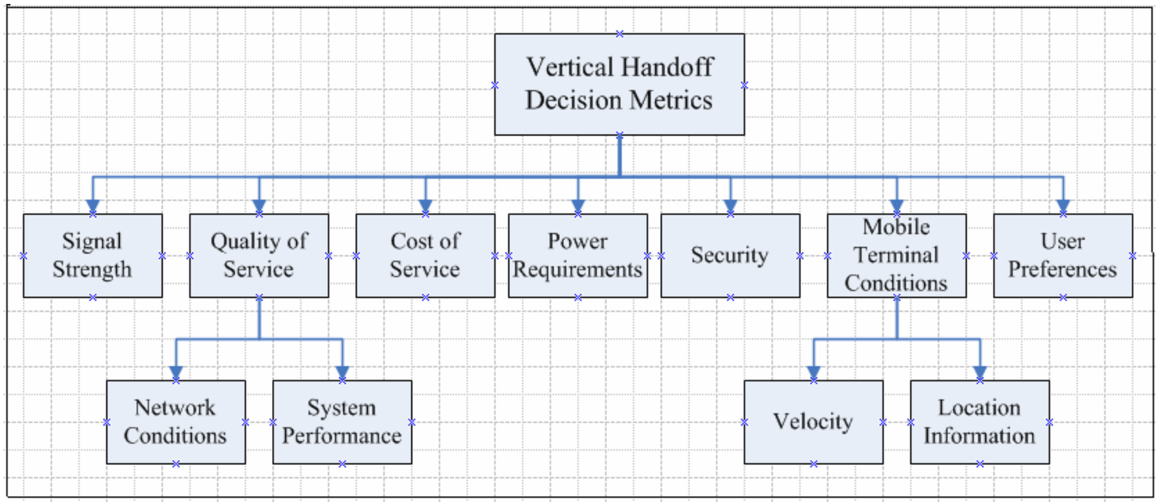


Figure 2.2: *Vertical Handoff Decision Metrics*

2.5.3 Handoff Execution

This phase executes the vertical handoff procedure to associate the mobile terminal with the new wireless access network. Handoff execution requires the actual transfer of data packets to a new wireless link in order to reroute a mobile user's connection path to the new point of attachment. It can be implemented by protocols such as Mobile IP and Stream Control Transmission Protocol.

Many proposals have been made for implementing handoffs while roaming across heterogeneous wireless networks. These approaches operate at different layers of the Internet protocol stack. We shall review some of these approaches in the next section.

2.6 Review of Vertical Handoff Related Literature

As an important step towards achieving the objective of enabling seamless vertical handoff, the IEEE 802.21 or Media Independent Handover (MIH) [25] standard has been proposed as a framework to provide interoperability, generic link layer intelligence and to enable efficient handoff management between heterogeneous access networks. This section provides a review of relevant vertical handoff related proposals and algorithms made by researchers aimed at providing seamless vertical handoffs. Vertical handoff decision algorithms are discussed in subsection 2.6.1. The relevant literature on vertical handoff execution is presented in subsection 2.6.2.

2.6.1 Vertical Handoff Decision Algorithms (VHDAs)

The complexity of the handoff decision process has led to the proposal of a number of vertical handoff decision algorithms for wireless access networks. In this section, we discuss vertical handoff decision algorithms proposed in the research literature.

2.6.1.1 Received Signal Strength (RSS) based VHDAs

The first approach is based on the traditional strategies of using the RSS that may be combined with other parameters. The RSS based vertical handoff decision algorithms compare the RSS of the current point of attachment with the RSS of the available point of attachment to make handoff decisions.

In [26], Ylianttila *et al*. present an algorithm to compute an optimization policy for the dwelling timer according to the available data rates in both networks. They show that the optimal value for the dwelling timer is dependent on the difference between the available data rates in both networks. In [27], Zahran and Liang proposed a signal threshold adaptation algorithm for vertical handoff in heterogeneous networks. They investigated the effect of an application-based signal strength threshold on adaptive preferred-network lifetime-based handoff strategy in terms of the signaling load, available bandwidth, and packet delay for a roaming mobile in a 3G-WLAN inter-network.

2.6.1.2 Cost function-based VHDAs

A vertical handoff decision cost function is a measurement of the benefit obtained by handing off to a particular network. It is evaluated for each network that covers the service area of a user. It is derived from a number of parameters and it is efficient, flexible and has a low implementation complexity.

In [28], the authors propose a policy-enabled handoff across a heterogeneous network environment using different parameters such as available bandwidth B_n, power consumption P_n, and cost C_n. The cost function f_n of the network n is given by

$$f_n = w_b \cdot N(1/B_n) + w_p \cdot N(P_n) + w_c \cdot N(C_n) \quad (2.1)$$

$$= w_b \cdot \ln(1/B_n) + w_p \cdot \ln(P_n) + w_c \cdot \ln(C_n), \quad (2.2)$$

with $N(x)$ as the normalization function of the parameter x, and where w_b, w_p, and w_c are the weights of the parameters such that $\sum w_i = 1$. The cost function is estimated for the available access networks and then used in the handoff decision of the MT. Using a similar approach as in [28], a cost function-based vertical handoff decision algorithm for multi-services handoff was presented in [24]. The selection of the optimal network, *n_opt*, is based on

$$n\_opt = argmin(f^n) \;\; \forall n, \tag{2.3}$$

where f^n is the handoff cost function for network *n,* and is calculated as

$$f^n = \sum_s \left(\prod_i E^n_{s;i}\right) \sum_j f_{s;j}(w_{s;j})\, N(Q^n_{s;j}), \tag{2.4}$$

where $N(Q^n_{s;j})$ is the normalized QoS parameter, $Q^n_{s;j}$, representing the cost in the *j*th parameter to carry out service *s* on network *n,* $f_{s;j}(w_{s;j})$ is the *j*th weighting function for service *s* and $E^n_{s;i}$ is the *i*th network elimination factor of service *s*. The available network with lowest cost function value becomes the handoff target. However, only the available bandwidth and the RSS of the available networks were considered in the handoff decision performance comparisons. In [15], Koutsorodi *et al.* present a mobile terminal architecture for devices operating in heterogeneous environments, which incorporates intelligence for supporting mobility and roaming across access networks. They compute the function:

$$OF(p, q) = w_q \times \text{Quality}(p, q) + w_o \times \text{Operator}(p) + w_t \times \text{Technology}(p) - w_c \times \text{Cost}(p, q), \tag{2.5}$$

for all $p \in P = \{p_1, p_2, \ldots, p_n\}$, $n \in \mathbb{Z}$, and $q \in Q(p) = \{q_1, q_2, \ldots, q_m\}$, $m \in \mathbb{Z}$; where P is the set of attachment points that the terminal perceives, and $Q(p)$ is the set of quality levels at which attachment point p can offer the service under consideration. The optimal attachment point and quality level for each of the requested/running services is the determination of: $\max_{\forall p \in P}\{\max_{\forall q \in Q(p)}\{OF(p, q)\}\}$.

2.6.1.3 Multiple Criteria VHDAs

Multiple criteria VHD algorithms make handoff decisions based on several handoff criteria such as signal strength, network coverage area, network security, and service cost. They combine the various criteria in order to select the most suitable access network for a new service request or handoff. They are efficient and flexible, and have a medium implementation complexity. The handoff decision problem is a typical multiple criteria decision making (MCDM) problem since the handoff decision problem involves selecting a suitable access network among a number of candidate access networks with respect to several criteria (attributes). The cost function based algorithm and the computational intelligence based algorithm are based on multiple criteria. These algorithms are the currently used approaches for combining many decision criteria for vertical handoff decision. The MCDM can be divided into multiple attribute decision making (MADM) that deals with the problem of choosing an alternative from a set of alternatives which are characterized in terms of their attributes, and multiple objective decision making (MODM) that consists of a set of conflicting goals which cannot be achieved simultaneously [29].

The decision about access network selection in a heterogeneous wireless environment can be solved using specific multiple attribute decision making (MADM) algorithms such as: Weighted Sum Model (WSM) [30] or Simple Additive Weighting (SAW), Weighted Product Model (WPM) [30] or Multiplicative Exponent Weighting (MEW), Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [30], Analytic Hierarchy Process (AHP) [30], and Grey Relational Analysis (GRA) [31]. In [32], the authors compare the performance between four fuzzy MADM-based vertical handoff decision algorithms. The imprecise data are first converted to crisp numbers, and then the classical MADM methods, including SAW, MEW, TOPSIS and GRA are used to calculate the score. It showed that SAW, MEW and TOPSIS provide similar performance to the traffic classes used, and GRA provides a slightly higher bandwidth and lower delay for interactive and background traffic. An integrated network selection algorithm using AHP and GRA is presented in [33] with a number of parameters, and UMTS and WLAN considered as alternative networks. The AHP is used to achieve weighting of QoS parameters based on service application and user preferences, and the GRA is used to rank the network alternatives. However, the AHP method demands certain pre-known relations between the QoS parameters, but there is limited explanation of the way these things come up in [33]. Charilas *et al.* [34] expand the methodology described in [33] and define the required QoS parameters in packet-switched networks by conducting and making use of real measurements. In [21], Nasser *et al.* describe handoff decisions, radio link transfer, and channel assignment as stages of the handoff process. They propose a vertical handoff decision function Q_i that is defined in terms of a number of network parameters (including cost of service (*C*), security (*S*), power consumption (*P*), network conditions (*D*), and network performance (*F*)) and provides handoff decision when roaming across heterogeneous wireless networks:

$$Q_i = f\,(w_c/C_i,\ w_s S_i,\ w_p/P_i,\ w_d D_i,\ w_f F_i) \qquad (2.6)$$

where w_c, w_s, w_p, w_d, and w_f are weights for each of the network parameters.
However, computational intelligence techniques were not used.

2.6.1.4 Computational Intelligence VHDAs

The fourth category of handoff decision algorithm uses computational intelligence techniques. Computational intelligence based handoff decision algorithms choose an access network for vertical handoff by applying a computational intelligence technique, such as Fuzzy Logic (FL), Fuzzy Multiple Attribute Decision Making (FMADM), Neural Networks (NNs), and Genetic Algorithm, to some vertical handoff decision criteria. Most computational-intelligence-based VHD algorithms incorporate fuzzy logic. Formulating the vertical handoff decision as a FMADM problem allows the use of fuzzy logic to deal with imprecise information that the decision attributes could contain as well as to combine and evaluate multiple attributes. Computational intelligence based handoff decision algorithms have high efficiency, and improve users' satisfaction in using heterogeneous wireless networks. However, they have high implementation complexity.

In [20], Pahlavan *et al*. present a neural networks-based approach that detects signal decay and makes handoff decision. The authors showed that the NN architecture performs better than traditional handover decision algorithms in terms of handover delay and number of unnecessary handovers. However, the NN architecture requires prior knowledge of the radio environment and needs more configuration before deployment.

In [35], Chan *et al.* propose a mobility management in a packet-oriented multi-segment made up of terrestrial (GPRS and UMTS) and satellite mobile networks using Mobile IP and fuzzy logic concepts. Handover is separated into initiation, decision and execution phases. MIP is used in the handover execution phase, a fuzzy logic system is applied with four different system criteria (bit error rate, signal strength, network coverage, and perceived QoS) in the handover initiation phase, and fuzzy logic and multiple objective decision making (MODM) concepts are applied during the handover decision phase to select an optimum network using inputs from the system and user. The handover decision phase involves two different stages: (1) fuzzy ordinal ranking procedures and weighting of the criteria, and (2) the MODM to select the optimum network segment. However, the handover management is for vertical handoff between different wide area networks. A vertical handover decision making algorithm using fuzzy logic for an integrated radio and optical wireless system is presented in [36]. Three input fuzzy variables were used: the probability of a short interruption, the failure probability of handover to radio, and the size of unsent messages. The paper shows that the proposed algorithm can achieve excellent performance of packet transfer delay. In [37] the vertical handoff decision process is formulated as a FMADM problem. The proposed handoff scheme consists of two parts: the first one is to process multiple criteria using a Mamdani fuzzy logic inference system, while the second one applies a genetic algorithm to optimise a FMADM access network selection function to select a suitable access network.

In [38], Guo *et al.* use an algorithm consisting of a Modified Elman Neural Network for the number of users predicted after the handover (considered as an input of the adaptive multi-criteria decision) and an FIS that makes the analysis of relevant criteria and does the final decision according to the inputs and the IF-THEN rules. The authors showed by simulations that the adaptive multi-criteria VHDA can make handover decision for UMTS and WLAN networks and performs better compared to the conventional algorithm based on the RSS criterion.

2.6.1.5 Context-aware approaches

The context-aware handoff decision concept is based on the idea that the behaviour of the handoff algorithm should be governed by its surroundings. Context-awareness exploits device, network, and user information to improve connectivity, QoS and maintain a high level of user satisfaction. Context information, which is relevant to the handoff decision algorithm, is related to the network (such as QoS, coverage), terminal (such as its location, capabilities), the service (such as QoS requirements, service type), and user preferences. In [39], Wei *et al.* propose an approach which integrates a context management framework, a programmable platform and a service deployment scheme to provide the functionality needed for context-aware handover. The context management framework is in charge of collecting the relevant context information for different services and managing the context information, the programmable platform is used to download and install suitable modules for context exchange and processing, while the service deployment scheme is used to synchronise and manage the work of the involved active nodes. The approach shows through a prototype evaluation that mobile nodes are able to execute handovers more efficiently when context information is considered. In [40], Kassar *et al.* propose an intelligent handover management system for future generation wireless networks using a context-aware vertical handover decision strategy. The paper uses a combination of fuzzy logic and the AHP for network selection. It attempts to address the issue of handling imprecise data during handovers.

2.6.1.6 Comparison of the Proposed Handoff Decision Algorithms

The handoff decision approaches in the research literature that were reviewed in the previous sub-sections are compared in Table 2.2 according to user consideration, multiple criteria, efficiency, flexibility, and computational complexity. We deduce from the comparison in Table 2.2, that multiple criteria, computational intelligence, and context-aware vertical handoff decision approaches have more advantages. Therefore, a suitable vertical handoff decision algorithm should be designed by taking the good advantages of multiple criteria, computational intelligence, and context-awareness into consideration.

Table 2.2: *Comparison of Proposed Handoff Decision Algorithms*

Vertical Handoff Decision Approach	Traditional (RSS-based)	Cost Function	Multiple Criteria	Computational Intelligence	Context-Aware
User consideration	No	Medium	Medium	Medium	High
Multiple criteria	No	Yes	Yes	Yes for FL No for NN	Yes
Application type supported	Non-real-time	Real-time and non-real-time	Real-time and non-real-time	Real-time and non-real-time	Real-time and non-real-time
Efficiency	Low	Medium	High	High	High
Flexibility	Low	High	High	Medium	High
Computational complexity	Low	Low	Medium	Medium to high	Medium

2.6.2 Vertical Handoff Management Techniques

Many proposals have been made for implementing handoffs while roaming across heterogeneous wireless networks. These approaches operate at different layers of the Internet protocol stack.

When designing a new architecture for implementing vertical handoff, it is important to limit the modifications required to existing wireless systems, and to minimise the amount of network traffic needed. The host mobility problem may be attacked from many layers of the Internet suite of protocols. In [41], Eddy addresses the issue of which layer in the IP protocol stack mobility belongs to. He discusses the various strengths and weaknesses of implementing mobility at three different layers of the protocol stack (network layer, transport layer, and session layer) using the three evaluation criteria of seamless transitions, location management, and infrastructure-free requirement (see Table 2.3). He concludes that the transport layer is the most likely place for a mobility protocol, but the best approach may be a cross-layer approach where interlayer communication is used.

Table 2.3: *A summary of differences between mobility approaches at various layers [41]*

Layer	Seamless transitions	Location management	Required infrastructure
Network layer (MIP)	Transport layer must deal with losses and path changes	Included	Deployment of HAs and router support for fast/ smooth handovers
Transport layer	Included	Requires external location manager	Little or none
Session layer	Included	May be included	Little or none

2.6.2.1 Network Layer Handoff Approach

Network layer solutions provide mobility-related features at the IP layer. Vertical handoff involves changing the access network interface, which typically results in changing the mobile node's IP address and administrative domain. The Mobile IP (MIP) protocol is therefore a natural candidate to support smooth vertical handoff.

MIP is a mobility management protocol proposed to solve the problem of node mobility by redirecting packets to the mobile node's current location. Its main functional entities (mobility agents (MAs)) are a home agent (HA) located in the home network and a number of foreign agents (FAs) located in visited foreign networks. MIP provides IP layer mobility by enabling a mobile node (MN) that originates from its home network to be addressable by the same home IP address across different foreign networks the MN is visiting. This is realized by maintaining a binding between the MN's home IP address and the care-of-address (CoA), which is the IP address allocated to the MN in the currently visited foreign network. The binding is created as a result of the MN registering its new CoA with the MN's HA as soon as it detects that its location has changed. Data traffic originated from and addressed to the MN is redirected between the HA and the FAs by means of IP-in-IP encapsulation. There are certain routing inefficiencies in MIP including triangle routing, triangle registration, encapsulation and need for home addresses. MIPv6 eliminates triangular routing and enables the correspondent node to reroute packets on a direct path to the MT.

In [42], Floroiu *et al*. provide a quantitative analysis of a Mobile IPv4-based WLAN-GPRS (General Packet Radio Service) handover prototype, and identify a number of side effects related to the link layer and routing mechanisms. Figure 2.3 shows a MN roaming between a WLAN access network and a GPRS access network while communicating to a correspondent node located in the home network. The datagrams exchanged between the mobile and correspondent nodes are redirected over the HA-FA and FA-MA segments using reverse tunneling and IP-in-IP encapsulation, and additional Point-to-Point Protocol/Gateway Tunneling Protocol tunneling over the GPRS Gateway Service Node (GGSN)-mobile node segment.

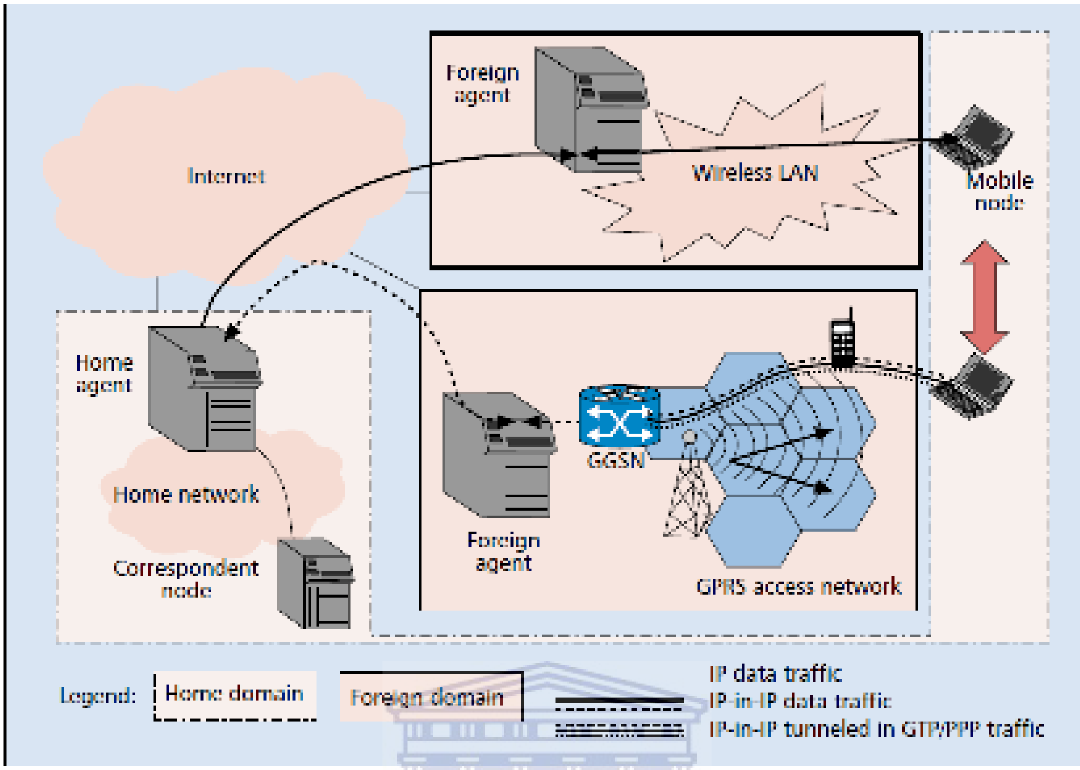


Figure 2.3: *Mobile IP in a WLAN-GPRS environment [42].*

2.6.2.2 Application Layer Handoff Approach

The Session Initiation Protocol (SIP)-based handoff [43] approach is an application-layer solution that provides personal and terminal mobility management in heterogeneous networks. SIP is an application-layer control protocol for establishing, modifying, and terminating multimedia sessions in IP-based networks between two or more participants. It also allows inviting another participant to ongoing sessions. The main entities in SIP are user agents, proxy servers, and redirect servers. Terminal mobility requires SIP to establish a connection either during the start of a new session (*pre-call mobility*), when the MT has already moved to a different location, or in the middle of a session (*mid-call mobility*). For mid-call mobility, the MT sends another INVITE message about the MT's new IP address and updated session parameters to the correspondent host (CH). Performing a vertical handoff during an ongoing session is similar to mid-call mobility. In [44], Wu *et al.* analyse the delay associated with vertical handoff using SIP in the WLAN-UMTS internetwork. An MT performs two key functions to initiate a WLAN-to-UMTS handoff [44]: data connection setup, and a SIP message exchange that re-establishes the connection. For a UMTS-to-WLAN vertical handoff, the MT goes through the following steps to update its location with the CH: DHCP registration procedure, and SIP message exchange. Numerical analysis showed that the WLAN-to-UMTS handoff due to error-prone and bandwidth-limited wireless links incurred much larger delay than the UMTS-to-WLAN handoff.

A major limitation of SIP-based handoff is that the processing of SIP messages in the intermediate and destination servers may take considerable time and introduce unacceptable handoff delays.

2.6.2.3 Transport Layer Handoff Approach

There are several benefits associated with implementing handoffs at the transport layer including [18, 41, 45]:

- Simplified network infrastructure. There is no requirement of a home network or additional infrastructure beyond Dynamic Host Configuration Protocol (DHCP) and Domain Name Server (DNS), which are already deployed as part of IP networks.
- Inherent route optimization. With mobility anchored at the transport layer, there is implicit route optimization. There are no triangular routes since packets move directly from the end source to the end destination with no indirection.
- Smooth handoffs. With transport layer mobility, transport protocols are explicitly aware of the changes in their network attachment status. This allows transport protocols to take proper action in order to smooth transitions into new networks.
- Immune to spurious agents. The transport layer mobility architecture does not use home agents or foreign agents and is therefore immune to spurious agents.

A drawback of the transport layer handoff approaches is their dependence on other layers for location management. Another drawback is that if each transport protocol is to implement binding updates, then each of the protocols requires an authentication scheme to prevent spoofing.

The transport layer approach requires a means to detect and reconfigure mobile hosts as they move from one network type to another. This includes the detection of new networks and the allocation of new IP addresses. These tasks are often handled by Dynamic Host Configuration Protocol (DHCP) or Router/Neighbor Discovery methods.

Recent transport-level management protocols that have been proposed include the Stream Control Transmission Protocol (SCTP) and the Datagram Congestion Control Protocol (DCCP).

Similar to the Transmission Control Protocol (TCP) and User Datagram Protocol (UDP), SCTP [46] is an IP-based transport protocol providing end-to-end communication over IP networks. Like TCP, SCTP provides a connection-oriented, reliable, full-duplex, congestion and flow-controlled layer 4 channel. However, unlike both TCP and UDP, SCTP offers new delivery options that better match diverse applications needs.

An SCTP connection, called an *association*, provides novel services such as *multi-homing*, which allows the end points of a single association to have multiple IP addresses, and *multi-streaming*, which allows for independent delivery among data streams. The built-in support for multihomed endpoints by SCTP is especially useful in environments that require high availability of the applications. Multistreaming is used to alleviate the head-of-line blocking effect resulting from TCP's strict byte-order delivery policy.

Kernel-level SCTP implementations are available on various operating system platforms, such as Linux and Solaris/OpenSolaris. We installed and tested the SCTP throughput performance of the OpenSS7 Linux kernel implementation [47] on two Pentium 4 machines in our network laboratory. One of the machines was equipped with two interface cards. Both machines were running Red Hat Linux 9 and the kernel-2.4.20-2.8.9.sctp.0.2.19.1.i386.rpm [48].

An extension of SCTP's multihoming feature [49] allows on-the-fly layer 3 address reconfiguration. This Dynamic Address Reconfiguration (DAR) extension for SCTP enables each end point to dynamically add or delete an IP address to or from an existing association, and to change the primary IP address for an active SCTP association using address configuration (ASCONF) messages. Due to the multi-homing feature of mobile SCTP (mSCTP), that is, an SCTP implementation with its DAR extension, an end point's network interface can be added into the current association if it is possible for the interface to establish a connection to the Internet via an IP address.

The capabilities of mSCTP to add, to change, and to delete the IP addresses dynamically during an active SCTP association provides an end-to-end vertical handoff solution between two IP access networks such as UMTS and WLAN [50]. Both the MT, or mobile client, and a fixed correspondent node, or fixed server, are assumed to implement mSCTP as shown in Figure 2.4. The multimode MT supports both UMTS and WLAN at the physical and data link layers. The handoff execution procedure has three basic steps: Add IP address, Vertical handoff triggering, and Delete IP address.

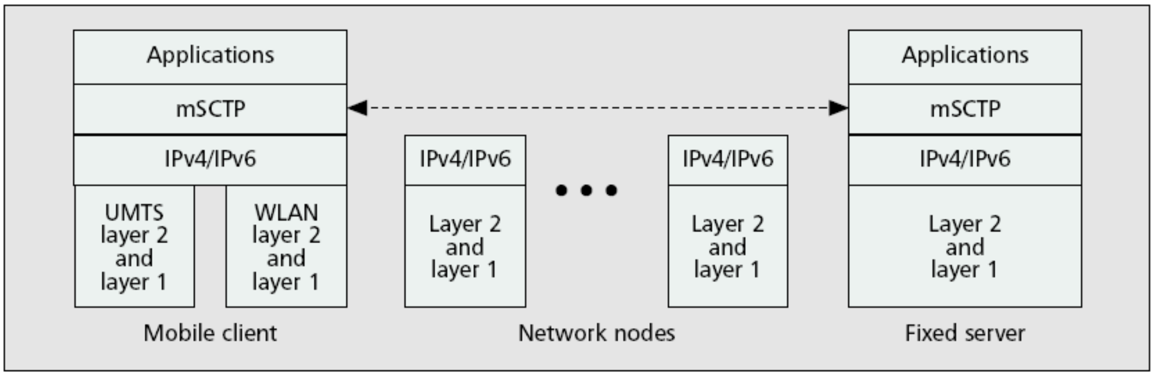


Figure 2.4: *Protocol architecture using mSCTP [50]*

2.7 Summary

An overview of the next generation wireless networks and seamless mobility was given in this Chapter. There was a thorough discussion of handoffs in the next generation wireless networks. There was an explanation of the three sequential handoff phases: access network discovery, handoff decision, and handoff execution. Vertical handoff related literature in the research field was reviewed. Vertical handoff decision algorithms (VHDAs) proposed in the research literature include RSS based approaches, cost function based VHDAs, multiple criteria VHDAs, computational intelligence VHDAs, and context-aware approaches. Vertical handoff implementation (execution) approaches are mainly at the network layer, application layer, and transport layer of the Internet protocol stack.

CHAPTER 3

VERTICAL HANDOFF INITIATION USING FUZZY INFERENCE SYSTEMS

This research exploits the capability of fuzzy logic to develop adaptive intelligent vertical handoff decision algorithms. In this chapter, an introduction to the principal concepts and mathematical notations of fuzzy set theory and fuzzy inference systems is given. We then present the design of a novel seamless vertical handoff management architecture that can achieve seamless mobility in the next generation heterogeneous wireless IP access networks for the mobile field workforce, and indeed for all mobile users, equipped with multimode mobile terminals. We also present our intelligent vertical handoff decision algorithm and its two components. The first component – vertical handoff initiation – is developed using a fuzzy inference system.

3.1 Introduction

Seamless handoff, with low latency and minimal packet loss, has become a crucial factor for mobile users who wish to receive continuous and reliable services. In traditional horizontal handoffs which happen in homogeneous networks, only information obtained from the radio-link layer such as the received signal strength (RSS) and channel availability are considered for handoff decisions. In contrast to this approach, in addition to the RSS, many parameters such as network coverage area, data rate, service cost, reliability, security, battery power, mobile terminal velocity, and network latency may be used in the vertical handoff decision process. This makes the entire vertical handoff process more complex and ambiguous since various factors should simultaneously be taken into account to make a successful handoff decision. One of the challenging issues in the multi-service NGWN is to design intelligent vertical handoff decision algorithms to determine when to perform a handoff and to provide an optimal choice of access network technology among all available access networks for users equipped with multimode mobile terminals. The design of such an intelligent vertical handoff decision algorithm is very challenging due to complexity of the different handoff decision metrics, modularity of the handoff decision, information imprecision, and interpretability issues. In this thesis, we propose fuzzy logic as an effective means of meeting these challenges, as far as both knowledge representation and inferencing are concerned. As an incomplete knowledge representation technique, fuzzy logic is well suited for addressing imprecision issues [51, 52].

A fuzzy inference system (FIS), or a fuzzy logic system, is a nonlinear mapping of input data vector into a scalar output. It is a computing framework based on the concepts of fuzzy set theory and fuzzy logic. Two types of fuzzy inference systems that can be implemented are the Mamdani-type and the Sugeno-type. Fuzzy inference systems have been successfully applied in fields such as automatic control, data classification, decision analysis, and expert systems.

Fuzzy logic is a multi-valued logic that deals with approximate reasoning. It adds to Boolean or two-valued logic an important capability – the capability to reason precisely with imperfect information, that is, information which is, in one or more respects, imprecise, uncertain, incomplete, unreliable, vague, or partially true [53]. Fuzzy logic comprises fuzzy sets and operations on fuzzy sets used to make inferences in fuzzy logic. Fuzzy sets model the properties of imprecision, approximation, or vagueness.

Our observations, experiences, and communications almost always include a large measure of uncertainty. Two main types of uncertainty are random and linguistic. The former is associated with unpredictability, and the latter is associated with words and it is based on ambiguity, imprecision, and/or vagueness. Probability theory is used to handle random uncertainty and fuzzy sets are used to handle linguistic uncertainty. A linguistic variable is a variable whose values are words, phrases, or sentences in natural language. It provides a way of expressing a phenomenon that is too complex or ill-defined to describe in a conventional crisp manner.

Fuzzy sets are associated with terms that appear in the antecedents or consequents of if-then rules and with the inputs and outputs of the FIS. Membership functions (MFs) are used to describe these fuzzy sets. Two kinds of fuzzy sets can be used in an FIS, type-1 and type-2. Type-1 fuzzy sets are described by MFs that are totally certain, whereas type-2 fuzzy sets are described by MFs that are themselves fuzzy. An FIS that is described completely in terms of type-1 fuzzy sets is called a *type-1 FIS*, whereas an FIS that is described using at least one type-2 fuzzy set is called a *type-2 FIS*.

3.2 Fuzzy Sets and Fuzzy Logic

3.2.1 Formal Definitions of Fuzzy Set

A fuzzy set is a generalization to classical set that allow objects to take partial membership in vague concepts [53]. Thus, a fuzzy set is a set that does not have a crisp boundary and is characterised by a membership function (MF) which defines the degree of membership of elements in the set. Different kinds of fuzzy sets can be distinguished. For example, a traditional fuzzy set or a type-1 fuzzy set (T1 FS) has a grade of membership that is crisp, whereas a type-2 fuzzy set (T2 FS) has grades of membership that are fuzzy. Although T1 FSs have been successfully used in many applications, however, such FSs have limited capabilities to directly handle lots of data uncertainties which can be handled by T2 FSs [54, 55, 56]. We use T1 FSs in this thesis.

Formally, if X is a collection of objects x (that is, X is a universe of discourse), then a traditional *fuzzy set (or type-1 fuzzy set)* A in X is defined as a set of ordered pairs:

$$A = \{(x, \mu_A(x)) \mid x \in X\}, \tag{3.1}$$

where $\mu_A : X \to [0, 1]$ is the *membership function* (MF) of the fuzzy set A. The MF defines how each point $x \in X$ is mapped to a degree of membership (or membership value) $\mu_A(x)$ in the interval [0, 1], where 0 and 1 correspond to full non-membership and full membership, respectively.

A *type-2 fuzzy set A* is characterized by a type-2 membership function $\mu_A(x, u)$, where $x \in X$ and $u \in J_x \subseteq [0, 1]$, that is,

$$A = \{((x, u), \mu_A(x, u)) \mid \forall x \in X, \forall u \in J_x \subseteq [0, 1]\}, \tag{3.2}$$

in which $0 \le \mu_A(x, u) \le 1$.

3.2.2 Fuzzy Set Characteristics

The *support* of a fuzzy set A is the set of all elements in the universe of discourse, X, that belongs to A with non-zero membership, that is,

$$support(A) = \{x \in X \mid \mu_A(x) > 0\} \tag{3.3}$$

A *fuzzy singleton* is a fuzzy set whose support is a single point in X with $\mu_A(x) = 1$.

The *core* (or *kernel*) of a fuzzy set A is the set of all elements in the universe of discourse, X, that belongs to A with membership degree 1, that is,

$$core(A) = \{x \in X \mid \mu_A(x) = 1\} \tag{3.4}$$

The *boundaries* of a fuzzy set A is the set of all elements in the universe of discourse, X, that belongs to A with non-zero membership but not complete membership, that is,

$$boundaries(A) = \{x \in X \mid 0 < \mu_A(x) < 1\} \tag{3.5}$$

A *normal* fuzzy set is one whose MF has at least one element x in the universe whose membership value is unity. That is, the fuzzy set A is normal if

$$\exists x \in X \mid \mu_A(x) = 1.$$

A fuzzy set A is said to be a *convex* fuzzy set if, for any elements x, y, and z in A, the relation $x < y < z$ implies that

$$\mu_A(y) \geq \min[\mu_A(x), \mu_A(z)] \tag{3.6}$$

Alternatively, a fuzzy set A is said to be a *convex* fuzzy set if and only if

$$\mu_A[\lambda x_1 + (1 - \lambda)x_2] \geq \min(\mu_A(x_1), \mu_A(x_2)) \quad \forall x_1, x_2 \in X \text{ and } \lambda \in [0, 1] \tag{3.7}$$

The *height* of a fuzzy set A is the maximum value of the MF, that is,

$$height(A) = \max\{\mu_A(x)\}.$$

3.2.3 Membership Functions

A membership function, also referred to as a characteristic function, defines the fuzzy set. It is used to associate a degree of membership of each of the elements of the domain to the corresponding fuzzy set.

While membership functions for fuzzy sets can be of any shape or type, these functions must satisfy the following constraints:

- The range of a MF must be in the interval [0, 1].
- For each $x \in X$, $\mu_A(x)$ must be unique.

The most common forms of MFs are those that are normal and convex. Commonly, triangular, trapezoidal, and Gaussian functions are used as MFs. These parameterized functions are introduced below.

Triangular MF: A triangular MF is a function of a vector, **x**, and depends on three scalar parameters *a, b,* and *c* with $a < b < c$, as follows:

$$trimf(x; a, b, c) = \begin{cases} 0, & x \le a \\ (x-a)/(b-a), & a \le x \le b \\ (c-x)/(c-b), & b \le x \le c \\ 0, & c \le x \end{cases} \tag{3.8}$$

or more compactly as

$$trimf(x; a, b, c) = \max(\min((x-a)/(b-a), (c-x)/(c-b)), 0) \tag{3.9}$$

The parameters *a* and *c* locate the *x* coordinates of the feet of the triangle and the parameter *b* locates the *x* coordinate of the peak of the triangle.

Trapezoidal MF: A trapezoidal MF is a function of a vector, **x**, and depends on four scalar parameters *a, b, c*, and *d* with $a < b < c < d$, as follows:

$$trapmf(x; a, b, c, d) = \begin{cases} 0, & x \le a \\ (x-a)/(b-a), & a \le x \le b \\ 1, & b \le x \le c \\ (d-x)/(d-c), & c \le x \le d \\ 0, & d \le x \end{cases} \tag{3.10}$$

or more compactly as

$$trapmf(x; a, b, c, d) = \max(\min((x-a)/(b-a), 1, (d-x)/(d-c)), 0) \tag{3.11}$$

The parameters *a* and *d* locate the *x* coordinates of the feet of the trapezoid and the parameters *b* ad *c* locate the *x* coordinates of the shoulders.

Gaussian MF: A Gaussian MF is specified by two parameters σ and *c* as follows:

$$gaussmf(x; \sigma, c) = \exp(-((x-c)/\sigma)^2) \tag{3.12}$$

where *c* determines the center of the MF and σ represents the width.

Both the triangular MF and trapezoidal MF are linear functions and have been used extensively due to the easy computational requirement. Because of its concise notation, the Gaussian MF is popularly used in specifying fuzzy sets in complex systems, which has the advantage of being smooth and differentiable at all points. The three MFs are illustrated in Figure 3.1.

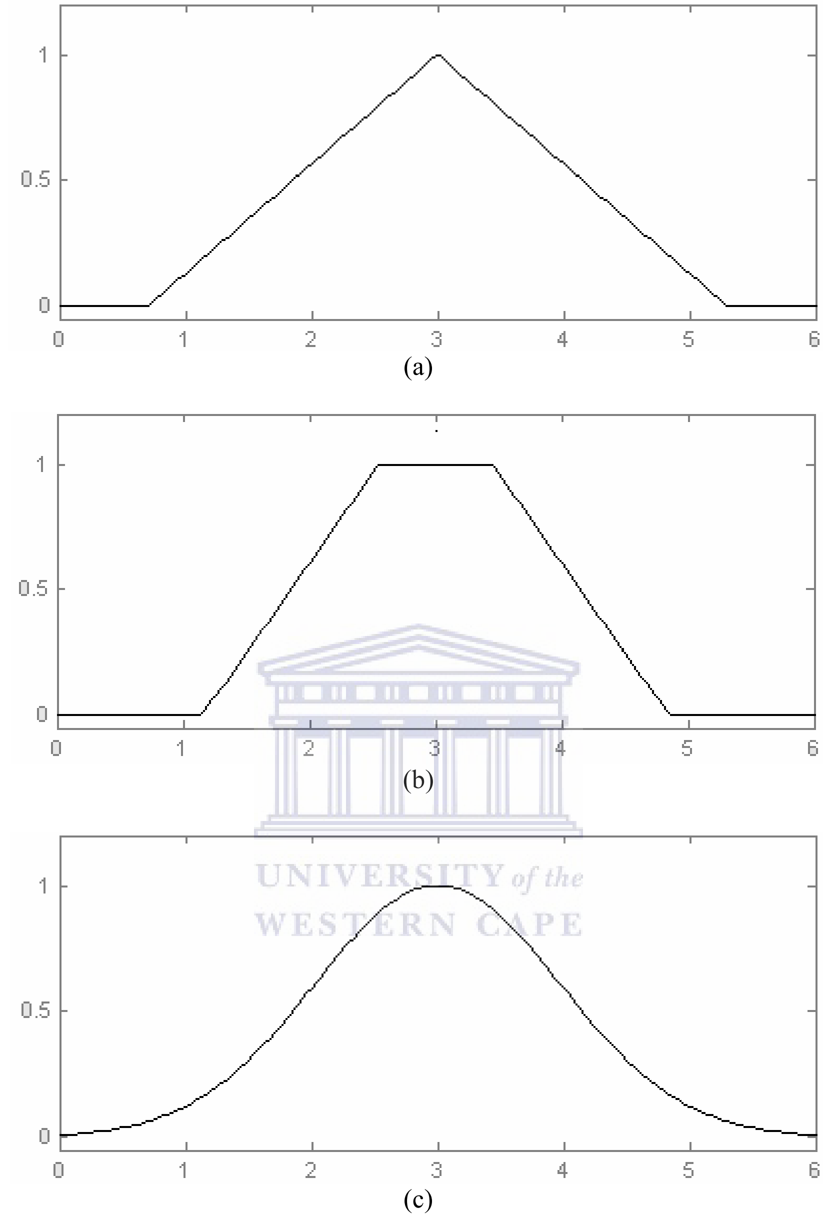


Figure 3.1: *(a) Triangular MF, (b) Trapezoidal MF, and (c) Gaussian MF*

3.2.4 Fuzzy Operators

Fuzzy set operators are needed to operate on linguistic variables. The set operations *intersection* and *union* correspond to the fuzzy logic operations, *conjunction* (*and*) and *disjunction* (*or*), respectively. There are many choices for these operations. The fuzzy conjunction operator and the fuzzy disjunction operator are dual with respect to negation.

Specifically, a fuzzy conjunction operator, denoted as $T(x, y)$, and a fuzzy disjunction operator, denoted as $S(x, y)$, form a dual pair if they satisfy the following condition:

$$1 - T(x, y) = S(1 - x, 1 - y)$$

The set of candidate fuzzy conjunction operators (including *fuzzy intersection*, *algebraic product*, *bounded product*, and *drastic product*) is called *triangular norms* or *t-norms*, and the set of candidate fuzzy disjunction operators is called *triangular conorms, t-conorms,* or *s-norms*.

A t-norm operator, denoted as $T(x, y)$, (or alternately, as $x * y$) is a function mapping from [0, 1] × [0, 1] to [0, 1] that is commutative, associative, monotonic and has 1 as unit.

A t-conorm operator, denoted as $S(x, y)$, (or alternately, as $x \oplus y$) is a function mapping from [0, 1] × [0, 1] to [0, 1] that is commutative, associative, monotonic and has 0 as unit.

In this chapter, we use *min* (minimum) for fuzzy conjunction and *max* (maximum) for fuzzy disjunction. The *intersection* (*conjunction*) of two fuzzy sets A and B is a fuzzy set C, written as $C = A \cap B$ or $C = A\ AND\ B$, whose MF is related to those of A and B by

$$\mu_C(x) = min(\mu_A(x), \mu_B(x)) = \mu_A(x) \wedge \mu_B(x) \qquad (3.13)$$

where $\wedge$ denotes the min operation between these two MFs.

The *union* (*disjunction*) of two fuzzy sets A and B is a fuzzy set C, written as $C = A \cup B$ or $C = A\ OR\ B$, whose MF is related to those of A and B by

$$\mu_C(x) = max(\mu_A(x), \mu_B(x)) = \mu_A(x) \vee \mu_B(x) \qquad (3.14)$$

where $\vee$ is the max operation between these two MFs.

The *complement* A' of a fuzzy set A is defined by

$$A'(x) = 1 - A(x). \qquad (3.15)$$

A fuzzy set A is *contained* in a fuzzy set B (or, equivalently, A is a *subset* of B) if and only if $\mu_A(x) \le \mu_B(x)$ for all x. In symbols,

$$A \subseteq B \Leftrightarrow \mu_A(x) \le \mu_B(x). \qquad (3.16)$$

3.2.5 Fuzzy Relations and Compositions

A fuzzy relation generalizes the classical notion of relation into a matter of degree. Formally, a *fuzzy relation* $R(X, Y)$ between the variables $x \in X$ and $y \in Y$ is defined by a function that maps ordered pairs in the product space $X \times Y$ to their degree of membership in the relation $R(X, Y)$, that is, $R: X \times Y \rightarrow [0, 1]$. The fuzzy relation $R(X, Y)$ is in fact a fuzzy set in the product space $X \times Y$, that is, it is a fuzzy subset of $X \times Y$, and has the MF $\mu_R(x, y)$ where $x \in X$ and $y \in Y$, that is,

$$R(X, Y) = \{((x, y), \mu_R(x, y)) \mid (x, y) \in X \times Y\} \tag{3.17}$$

More generally, a fuzzy n-ary relation $R(X_1, X_2, \ldots, X_n)$ in the variables $x_1, x_2, \ldots, x_n$ whose domains are $X_1, X_2, X_3, \ldots, X_n$, respectively, is defined by a function that maps the n-tuple $<x_1, x_2, x_3, \ldots, x_n>$ in $X_1 \times X_2 \times X_3 \times \ldots \times X_n$ to a number in the interval [0, 1], that is, $R: X_1 \times X_2 \times \ldots \times X_n \rightarrow [0, 1]$.

Let $R(X, Y)$ and $S(X, Y)$ be fuzzy relations in the same product space $X \times Y$. The *intersection* and *union* of R and S, which are *compositions* of the two relations, are then defined as

$$\mu_{R \cap S}(x, y) = \mu_R(x, y) * \mu_S(x, y) \tag{3.18}$$

$$\mu_{R \cup S}(x, y) = \mu_R(x, y) \oplus \mu_S(x, y) \tag{3.19}$$

where $*$ is any t-norm, and $\oplus$ is any t-conorm.

Let R and S be fuzzy relations in $U \times V$ and $V \times W$, respectively, and let $*$ be a t-norm. The **composition** $R \circ S$ of R and S with respect to $*$ is the fuzzy relation on $U \times W$ with membership function given by

$$\mu_{R \circ S}(u, w) = sup_{v \in V}[\mu_R(u, v) * \mu_S(v, w)] \tag{3.20}$$

When U, V and W are discrete universes of discourse, then the sup operation is the maximum. The composition of fuzzy relations is not uniquely defined. When $x * y = x \wedge y$, $R \circ S$ is referred to as a **max-min composition**. When $x * y = xy$, $R \circ S$ is a **max-product composition**.

Let A and B be fuzzy sets in U and V, respectively, and let $*$ be a t-norm. A fuzzy implication, denoted by $A \rightarrow B$, is a special kind of fuzzy relation in $U \times V$ with the following membership function:

$$\mu_{A \to B}(u, v) = \mu_A(u) * \mu_B(v). \quad (3.21)$$

This fuzzy implication is known as *fuzzy conjunction*.

3.2.6 Linguistic Variables and Linguistic Hedges

Two main aspects essential to fuzzy system modeling are the concepts that are to be quantified such as *height* and the characteristics of the identified concepts such as *short* and *tall* for height. The concepts quantified within a fuzzy inference system are called *linguistic variables* and the corresponding characteristics are called *linguistic terms*. A *linguistic variable* is a variable whose values are words or sentences in a natural or artificial language [57]. A linguistic variable is a variable whose value can be described: 1) qualitatively using an expression involving linguistic terms, and 2) quantitatively using a corresponding MF [58].

There are three main categories of linguistic variable:

- Quantification variables, e.g., all, most, many, few, none.
- Usuality variables, e.g., always, frequently, often, seldom, never.
- Likelihood variables, e.g., certain, likely, possible, uncertain, unlikely.

Linguistic variables that change the shape or position of a MF are called *linguistic hedges* [59]. A *linguistic hedge* or modifier is an operation that modifies the meaning of a term, or more generally, of a fuzzy set. A hedge can be viewed as an operator that acts upon a fuzzy set's membership function to modify it. Different kinds of hedges can be defined as follows:

- **Concentrations:** Concentrations tend to concentrate the membership values around points with higher membership degrees by reducing the degree of membership of all elements that are only "partly" in the fuzzy set. They can be defined, in general terms, as

  $$\mu_{con(A)}(x) = \mu_A(x)^p \text{ for } p > 1, \quad (3.22)$$

  where *con(A)* is the concentration of the set *A*. For example, if *weak RSSI* has membership function $\mu_{WR}(r)$, then *very weak RSSI* is a fuzzy set with membership function $[\mu_{WR}(r)]^2$, and *very very weak RSSI* is a fuzzy set with membership function $[\mu_{WR}(r)]^4$.

- **Dilations**: Dilations stretch or dilate a fuzzy set by increasing the membership degree of elements that are "partly" in the set. Dilation hedges are defined, in general, as

$$\mu_{dil(A)}(x) = \mu_A(x)^{1/p} \text{ for } p > 1. \tag{3.23}$$

If, for example, *weak RSSI* has membership function $\mu_{WR}(r)$, then *somewhat weak RSSI* is a fuzzy set with membership function $[\mu_{WR}(r)]^{1/2}$.

- **Intensification**: This operation acts in a combination of concentration and dilation. It increases the degree of membership of those elements in the set with original membership values greater than 0.5, and it decreases the degree of membership of those elements in the set with original membership values less than 0.5. This hedge is defined as

$$\mu_{int(A)}(x) = \begin{cases} 2^{p-1}\,\mu_A(x)^p & \text{for } \mu_A(x) \le 0.5 \\ 1 - 2^{p-1}\,(1 - \mu_A(x))^p & \text{for } \mu_A(x) > 0.5 \end{cases} \tag{3.24}$$

3.3 Fuzzy Inference System

A fuzzy inference system (FIS) is a nonlinear mapping of input data vector into a scalar output [54]. It is a computing framework based on the concepts of fuzzy set theory, fuzzy if-then rules, and fuzzy reasoning. It is able to simultaneously handle numerical data and linguistic knowledge using fuzzy logic whose real power is in its ability to efficiently handle and manipulate verbally-stated information [60].

Two types of fuzzy inference systems that can be implemented are the Mamdani-type and the Sugeno-type [61]. The differences between these fuzzy inference systems lie in the consequents of their fuzzy rules, and therefore their aggregation and defuzzification processes differ accordingly. In this thesis we use the Mamdani FIS which has the following advantages compared to other FISs [61]:

- It is intuitive.
- It has widespread acceptance.
- It is well suited to human input.

An FIS that is described completely in terms of type-1 fuzzy sets is called a *type-1 FIS*, whereas an FIS that is described using at least one type-2 fuzzy set is called a *type-2 FIS*. A type-1 FIS contains four components – a fuzzifier, a fuzzy rule base, a fuzzy inference engine, and a defuzzifier – that are inter-connected, as shown in Figure 3.2. The FIS maps crisp inputs into crisp outputs, and this mapping can be expressed quantitatively as $y = f(\boldsymbol{x})$, where the input and output of the fuzzy system are $\boldsymbol{x} \in R^N$ and $y \in R$, respectively.

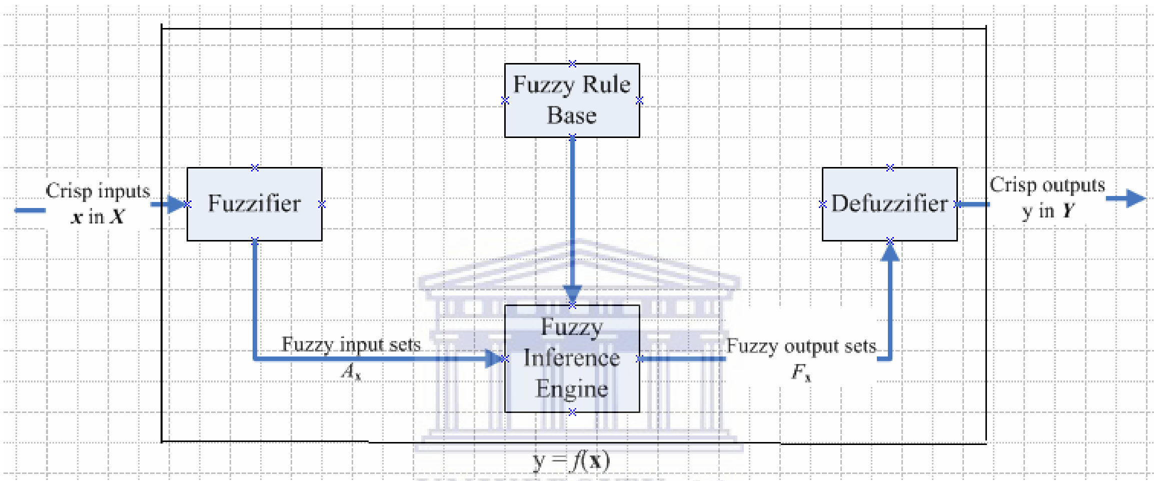


Figure 3.2: *Type-1 Fuzzy Inference System.*

We discuss the components of a T1 FIS below.

3.3.1 Fuzzifier

The fuzzifier transforms the crisp inputs into degrees of match with linguistic values. The fuzzifier maps a crisp point, $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T \in \boldsymbol{X}$, into a fuzzy set $A = \{(x_1, \mu_A(x_1)), (x_2, \mu_A(x_2)), \ldots, (x_n, \mu_A(x_n))\}$ in $\boldsymbol{X}$, where $\mu_A : \boldsymbol{X} \rightarrow [0, 1]$ is the membership function of the fuzzy set A and $\mu_A(x_i)$ is the membership degree of x_i in A. The fuzzifier is needed in order to activate rules which are expressed in terms of linguistic variables, which have fuzzy sets associated with them. The inputs to the FIS prior to fuzzification may be certain (for example, perfect measurements) or uncertain (for example, noisy measurements). Two choices for the fuzzifier are the *singleton fuzzifier* and the *non-singleton fuzzifier*.

- *Singleton fuzzifier* – The set A is a fuzzy singleton with support x_i, i.e., $\mu_A(x_i) = 1$ for $x = x_i$ and $\mu_A(x_i) = 0$ for all other $x_i \in X$ with $x \neq x_i$. The main advantage in using the singleton fuzzifier is the great simplicity of implementing the consequence part. It can be used with the Mamdani method to simplify considerably the deffuzification stage, whose task is reduced to the calculation of a weighted average with a restricted set of crisp values.

- *Non-singleton fuzzifier* – For this fuzzifier, $\mu_A(x_i) = 1$, and $\mu_A(x_i)$ decreases from 1 as x_i moves away from x. In non-singleton fuzzification, x_i is mapped into a fuzzy number, i.e., a fuzzy membership function is associated with it. A non-singleton fuzzifier is useful when the input measurements contain uncertainties (such as additive noise), and they are handled by treating them as fuzzy numbers.

3.3.2 Fuzzy Rule Base

It contains a number of fuzzy IF-THEN rules. If a Mamdani fuzzy inference system has n inputs and a single output, its fuzzy rule R_j is of the general form:

$$\text{IF } x_1 \text{ is } F_{1j} \text{ AND } x_2 \text{ is } F_{2j} \text{ AND } \ldots \text{ AND } x_n \text{ is } F_{nj}, \text{ THEN } y \text{ is } G_j. \quad (3.25)$$

F_{ij} and G_j are fuzzy sets in $X_i \subset R$ and $Y \subset R$, respectively. The variables $\boldsymbol{x} = [x_1, x_2, x_3, \ldots, x_n]^T \in X_1 \times X_2 \times \ldots \times X_n$ and $y \in Y$ are called the input and the output linguistic variables, respectively. The IF-part of the rule is called the *antecedent* or premise, and the THEN-part of the rule is called the *consequent* or conclusion.

In the Takagi-Sugeno model, the fuzzy rule R_j is of the general form:

$$\text{IF } x_1 \text{ is } F_{1j} \text{ AND } \ldots \text{ AND } x_n \text{ is } F_{nj}, \text{ THEN } y = f_j(x_1, \ldots, x_n). \quad (3.26)$$

3.3.3 Fuzzy Inference Engine

The fuzzy inference engine is the kernel of the FIS and is used in performing the inference operations on the fuzzy rules. It handles the way in which rules are activated and combined. The inferencing process maps the fuzzified inputs received from the fuzzification process to the rule base, and produces a fuzzified output for each rule. Fuzzy logic principles are used to combine fuzzy IF-THEN rules in the rule base, and fuzzy sets in $\boldsymbol{X} = X_1 \times \ldots \times X_n$ are mapped into fuzzy sets in Y. Let $A \in \boldsymbol{X}$ be the input to the fuzzy inference engine, and let $A = F_{1j} \times F_{2j} \times \ldots \times F_{nj}$ and $G_j = B$. Then the fuzzy rule R_j can be represented as the fuzzy implication [54]

$$R_j : F_{1j} \times F_{2j} \times \ldots \times F_{nj} \rightarrow G_j = A \rightarrow B, \quad (3.27)$$

in $X \times Y$. The fuzzy inference engine is treated as a system that maps fuzzy sets into fuzzy sets by means of $\mu_{A \to B}(\boldsymbol{x}, y)$.

Rule R_j is described by the MF $\mu_{Rj}(\boldsymbol{x}, y)$, where

$$\mu_{Rj}(\boldsymbol{x}, y) = \mu_{Rj}(x_1, x_2, \ldots, x_n, y) = \mu_{A \to B}(\boldsymbol{x}, y) \tag{3.28}$$

since $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$. Hence

$$\begin{aligned}\mu_{Rj}(\boldsymbol{x}, y) &= \mu_{A \to B}(\boldsymbol{x}, y) = \mu_{F1j \times F2j \times \ldots \times Fnj \to B}(\boldsymbol{x}, y) \\ &= \mu_{F1j \times F2j \times \ldots \times Fnj}(\boldsymbol{x}) * \mu_B(\mathrm{y}) \\ &= \mu_{F1j}(x_1) * \mu_{F2j}(x_2) * \ldots * \mu_{Fnj}(x_n) * \mu_B(\mathrm{y}) \\ &= [T^n_{i=1}\, \mu_{Fij}(x_\mathrm{i})] * \mu_B(\mathrm{y}),\end{aligned} \tag{3.29}$$

where T is short for a t-norm.

The n-dimensional input to R_j is given by the fuzzy set $A_\mathbf{x}$ whose MF is

$$\mu_{A\mathbf{x}}(\mathbf{x}) = \mu_{X1}(x_1) * \mu_{X2}(x_2) * \ldots * \mu_{Xn}(x_n) = T^n_{i=1}\, \mu_{Xi}(x_\mathrm{i}) \tag{3.30}$$

Each rule R_j determines a fuzzy set $B_j = A_\mathbf{x} \circ R_j$ in Y such that

$$\begin{aligned}\mu_{Bj}(\mathrm{y}) &= \mu_{A\mathbf{x} \circ Rj}(\mathrm{y}) \\ &= \sup_{\mathbf{x} \in \mathbf{X}} [\mu_{A\mathbf{x}}(\mathbf{x}) * \mu_{A \to B}(\boldsymbol{x}, y)], \quad y \in Y,\end{aligned} \tag{3.31}$$

on using Zadeh's sup-star composition. This equation is the input-output relationship in Figure 3.2 between the fuzzy set that excites a one-rule inference engine and the fuzzy set at the output of that engine.

Substituting (3.29) and (3.30) into (3.31) yields

$$\begin{aligned}\mu_{Bj}(\mathrm{y}) &= \sup_{\mathbf{x} \in \mathbf{X}} [\mu_{A\mathbf{x}}(\mathbf{x}) * \mu_{A \to B}(\boldsymbol{x}, y)] \\ &= \sup_{\mathbf{x} \in \mathbf{X}} [T^n_{i=1}\, \mu_{Xi}(x_\mathrm{i}) * [T^n_{i=1}\, \mu_{Fij}(x_\mathrm{i})] * \mu_B(\mathrm{y})] \\ &= \sup_{\mathbf{x} \in \mathbf{X}} \{[T^n_{i=1}\, \mu_{Xi}(x_\mathrm{i}) * \mu_{Fij}(x_\mathrm{i})] * \mu_B(\mathrm{y})\} \\ &= \{[\sup_{x_1 \in X_1} \mu_{X1}(x_1)\mu_{F1j}(x_1)] * \ldots * [\sup_{x_n \in X_n} \mu_{X\mathrm{n}}(x_n)\mu_{Fnj}(x_n)]\} * \mu_B(\mathrm{y}), \quad y \in Y. \\ &= \begin{cases} [T^n_{i=1}\, \mu_{Fij}(x'_\mathrm{i})] * \mu_B(\mathrm{y})], & \text{SF} \\ [T^n_{i=1}\, (\sup_{x_i \in X_i} \mu_{Xi}(x_\mathrm{i}) * \mu_{Fij}(x_\mathrm{i}))] * \mu_B(\mathrm{y}), & \text{NSF} \end{cases}\end{aligned} \tag{3.32}$$

for both singleton input and non-singleton input with associated singleton fuzzification (SF) and non-singleton fuzzification (NSF), respectively.

The final fuzzy set $B = [A_{\mathbf{x}} \circ R_1, R_2, \ldots, R_M]$, which is determined by all the rules in the rule base, is obtained by combining B_j and its associated MF $\mu_{A\mathbf{x} \circ Rj}(y)$ for all $j = 1, \ldots, M$.

3.3.4 Defuzzifier

A defuzzifier transforms the fuzzy results of the inference into a crisp output for the FIS. It maps fuzzy sets in V into a crisp output $y \in V$. One criterion for the choice of a defuzzifier is computational simplicity. This criterion has led to the following popular defuzzification methods [62]:

1) Centroid Defuzzifier: This defuzzifier returns the centre of gravity (centroid), y, of the fuzzy set B and uses this value as the output of the FIS

$$\underline{y} = \int_S y(\mu_B(y)) \,/ \int_S (\mu_B(y)), \tag{3.33}$$

where S denotes the support of $\mu_B(y)$. Often, S is discretised so that $\underline{y}$ is given by:

$$\underline{y} = \sum_{i=1}^{n} y_i\,(\mu_B(y_i)) \,/ \sum_{i=1}^{n} (\mu_B(y_i)). \tag{3.34}$$

2) Maximum Defuzzifier: This defuzzifier examines the fuzzy set B and chooses as its output the value of y for which $\mu_B(y)$ is a maximum.

3) Mean of Maxima Defuzzifier: This defuzzifier examines the fuzzy set B and first computes the n values of y for which $\mu_B(y)$ is a maximum. It then calculates the mean of these n values as its output:

$$\underline{y} = \sum_{i=1}^{n} y_i \,/\, n, \tag{3.35}$$

where y_i is the support value at the n points, whose MF reach the maximum value $\mu_B(y_i)$.

3.4 A Seamless Vertical Handoff Algorithm in the Next Generation Wireless Network

After discussing several issues pertaining to seamless mobility between UMTS and mobile WiMAX access networks and reviewing vertical handoff related literature in the last chapter, we are now ready to present the design of a novel seamless vertical handoff management architecture that can achieve seamless mobility in the next generation heterogeneous wireless IP access networks for the mobile field workforce, and indeed for all mobile users, equipped with multimode mobile terminals under the motivation of increasing user satisfaction while maintaining a satisfactory level of QoS and seamless connectivity.

3.4.1 Requirements and Features

A next generation terminal-controlled vertical handoff algorithm should satisfy the following requirements:

- Seamless Access – In the NGWN, the MT will need to move from one integrated access network to another seamlessly and transparently without risk of connection disruption. Therefore, a next generation VHDA must support efficient and seamless roaming of MTs among multiple wireless networks.
- High bandwidth utilization – The integration of heterogeneous wireless access networks is designed to offer higher bandwidth utilization for especially constraining applications to satisfy their QoS requirements and improve the system performances.
- Multi-service – Next generation wireless networks will support multiple classes of services such as voice, video, video streaming, web browsing. Supporting multiple services in the NGWNs will enhance the satisfaction of users since different users have different service requirements. Therefore, VHDAs need to support multiple services.
- Quality of Service requirements – Next generation wireless systems will consist of various access technologies with differing parameters interconnected with a common IP-core. These applications will have varying requirements. A next generation vertical handoff algorithm must provide proper QoS including satisfactory high bandwidth, high throughput, minimum packet loss rate, minimum handoff latency, high reliability, and minimum service costs in a heterogeneous mobile computing environment.

There have been a number of research efforts concerning seamless mobility and seamless vertical handoff as reviewed in Chapter 2. However, there are still limitations in the ongoing research efforts. First, there is not an efficient way to deal with imprecise information that the decision attributes could contain. Second, it is difficult to provide user personalization for optimal network access selection. Third, the few papers that used cost or objective functions did not employ optimization algorithms.

To solve some of these problems, we propose a novel terminal-controlled handoff management architecture that provides context-awareness and seamless handoff in next generation heterogeneous wireless IP access networks under the motivation of increasing user satisfaction while maintaining a satisfactory level of QoS and seamless connectivity. Our architecture has the following features:

- It supports seamless vertical handoff.
- It allows users to configure the operational mode of their MTs and prioritise their preferences through changing weight factors and constraints of a single-objective access network selection optimization function.
- It proposes an access network discovery process based on the Third Generation Partnership Project (3GPP) standardized access network discovery and selection function (ANDSF) [63]. This mechanism saves power consumption of the MT, minimizes the modifications of legacy radio systems and facilitates the discovery of available WiMAX cells.
- It uses fuzzy multiple attribute decision making (FMADM) because selecting a suitable access network among a number of candidate networks with respect to different criteria is a typical multiple attributes decision making (MADM) problem.
- It uses fuzzy logic to deal with imprecise data that the decision attributes might contain, and to combine and evaluate multiple attributes simultaneously.
- It uses optimization techniques including the Genetic Algorithm (GA) and Simulated Annealing (SA) to maximize or minimize key parameters in order to select an optimum target access network for possible vertical handoff.
- It supports context-awareness.

- It does not require significant changes to legacy systems since service providers do not want their existing systems to be modified due to cost and complexity.

3.4.2 Terminal-controlled Handoff Framework

Our terminal-controlled handoff architecture is based on the following assumptions:

- There is at least one 3G WWAN (UMTS) and one mobile WiMAX that overlap in the service area of the MT.
- All the wireless access networks in the architecture are connected to the Internet.
- All the mobile terminals are multi-interface.
- Mobile Stream Control Transmission Protocol (SCTP) is supported in all the wireless access networks and the Internet.

The terminal-controlled handoff management is based on the intelligent functionalities in the multimode MT. As illustrated in Figure 3.3, the main components of the terminal-controlled handoff architecture are:

- **Network Interface Management (NIM):** The NIM contains the protocol stack of all the access networks and monitors the interfaces periodically. It is responsible for collecting information on the collocated networks in its receiving range by using the current network interface, identifying the network interfaces present in the MT, monitoring the status of the network interfaces, and executing the selection and de-selection of the appropriate interface.
- **User Preferences:** The User Preferences module is responsible for storing, accessing and editing the user's preferences. A user-friendly Graphical User Interface is used to enter the user preferences, which is a rating among the parameters considered for handoff decision. These user-supplied input values are used by the Handoff Decision sub-module to generate weights for the Access Network Selection Function.
- **Handoff Management (HM):** The vertical handoff process required to provide seamless mobility may be divided into three sequential phases [18]: access network discovery, handoff decision, and handoff execution. Therefore the HM

module, which is responsible for handling all issues relating to handoff management, comprises of the following three sub-modules:

- Access Network Discovery: This is responsible for the discovery of the availability or unavailability of different wireless access networks in the vicinity of the MT. This is discussed further in subsection 3.4.3.
- Handoff Decision: This decides whether to initiate a vertical handoff and selects the most suitable wireless access network among all the available access networks for serving the needs of the user based on access network characteristics, MT conditions, and user preferences. This is discussed further in section 3.5.
- Handoff Execution: Handoff execution requires the actual transfer of data packets to a new wireless link in order to reroute a mobile user's connection path to the new point of attachment. This is discussed further in subsection 3.4.4.

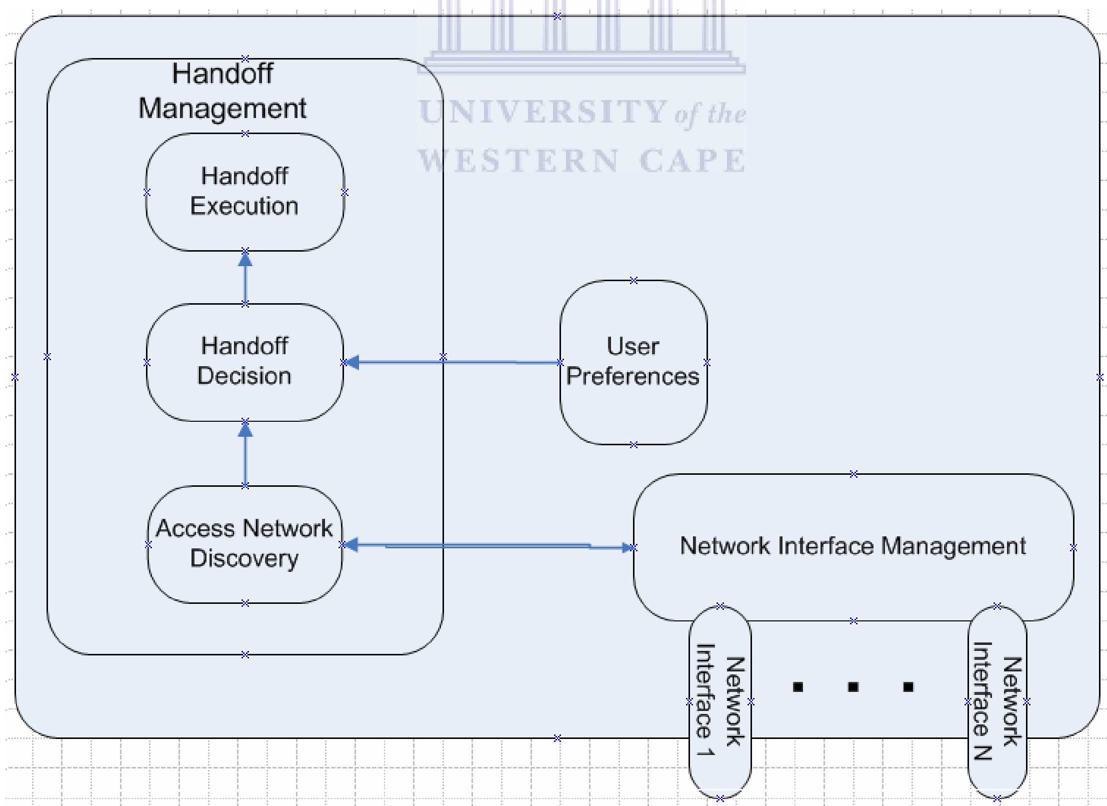


Figure 3.3: *Terminal-controlled Handoff Architecture*

3.4.3 Access Network Discovery

A mobile terminal (MT) equipped with multiple interfaces searches for reachable wireless access networks during the network discovery process. As the MT moves across the network, it has to discover other available access technologies in its surroundings which might be preferable to the currently used access network. For example, an MT using a UMTS access network in a NGWN system needs to discover when mobile WiMAX access networks become available and possibly handoff to a mobile WiMAX if this is more preferable to the operator and/or user, or if the radio signal from its serving UMTS cell starts to deteriorate significantly.

The MT can discover available cells with no assistance from the network by activating the interfaces to receive service advertisements broadcasted by different wireless technologies. Although this is very simple and does not require any modification in the network, this method causes problems. First, battery consumption can increase considerably. Second, information discovered about the available network is only limited.

To solve these problems and to minimize the modifications of legacy radio systems and facilitate the discovery of WiMAX cells, we propose that the neighbour cell information should be retrieved by the MT from the 3GPP standardized access network discovery and selection function (ANDSF) [63]. This function is a dynamic database that is queried by MTs whenever they need to discover neighbour cells of specific or any radio technology type. The ANDSF can also be used to provide operator policies (based on some dynamic operator preferences and rules) to the mobile devices, and additional information about neighbour cells such as QoS capabilities, service capabilities, charging rate, and a number of other attributes that cannot be continuously broadcast on radio channels due to the high radio capacity demand.

3.4.4 Vertical Handoff Execution

Once a mobile terminal decides to perform a vertical handoff, it should execute the vertical handoff process to be associated with the newly selected wireless access network.

We propose a transport layer mobility supporting scheme, which is based on mobile Stream Control Transmission Protocol (mSCTP). It utilizes the mSCTP's multi-homing capability and dynamic address reconfiguration feature. The mSCTP is targeted for the client-server model, where a mobile terminal (MT) initiates an SCTP session with a fixed correspondent node (CN). For supporting vertical handoff where the fixed CN initiates an SCTP session with the MT, the mSCTP must be used with a location management scheme such as the Session Initiation Protocol (SIP).

Scenario A: The MT communicates with a fixed Correspondent Node (CN)

This is the typical situation when a field worker needs to access vital information on the parent company's network server.

We use a client-server model where an MT communicates with a correspondent node (CN) fixed in a company network.

Using the multi-homing feature of SCTP, an MT can have two IP addresses during vertical handoff, one from the UMTS and the other from the WiMAX. However, the CN can be configured for:

- Single-homing: The CN provides only one IP address to support handoff. This scenario is likely to become the most common in today's heterogeneous wireless network environment since most Internet servers nowadays are configured with only one IP address.
- Dual-homing: The CN allows two IP addresses to support handoff.

The vertical handoff procedures of the single-homing and dual-homing configurations of the CN are shown in Figure 3.4 and Figure 3.5, respectively. For each of these configurations the handoff procedure has three basic steps:

- Add IP address
- Vertical handoff triggering
- Delete IP address.

Single-homing CN – In this case, a CN is configured with only one IP address, say, CN_IP. Suppose that an MT has been allocated with an IP address, UMTS_IP, in a UMTS cell and is using this IP address to communicate with the CN via mSCTP. The main phases in the handoff process are explained below:

1. When the MT moves into a WiMAX cell covered by a UMTS cell, it obtains a new IP address, WiMAX_IP, either via DHCP or IPv6 auto-configuration.
2. The new IP address is signaled to the mSCTP stack.
3. The mSCTP on the MT notifies the CN of the new IP by sending an ASCONF message to the CN with parameters set to "Add IP Address" and WiMAX_IP.
4. The CN confirms the incorporation of the new IP address into the association by sending an ASCONF_ACK message to the MT.
5. Then, at an appropriate moment the primary path change is triggered by the MT as soon as its handoff decision module indicates that the signal strength of the WiMAX access router exceeds the minimum signal strength that enables communication. The UMTS-to-WiMAX handoff is triggered by the MT sending an ASCONF message with parameters set to "Set Primary Address" and WiMAX_IP.
6. Once the MT receives an acknowledgment (ACK) from the CN, the WiMAX becomes the primary choice, and the traffic between the MT and the CN is routed through the WiMAX.
7. The WiMAX-to-UMTS handoff is triggered by the MT sending an ASCONF message to the CN with parameters set to "Set Primary Address" and UMTS_IP.
8. Once the MT receives an ACK from the CN, the UMTS becomes the primary choice, and the traffic between the MT and the CN is routed through the UMTS.
9. If the MT loses the signal from the WiMAX cell, it starts the delete IP address process. The MT sends an ASCONF message to the CN with parameters set to "Delete IP Address" and WiMAX_IP to request that the CN release the address WiMAX_IP from its host routing table.
10. After the MT receives an ACK from the CN, it deletes the WiMAX_IP from its address list, and WiMAX_IP is released from the association.

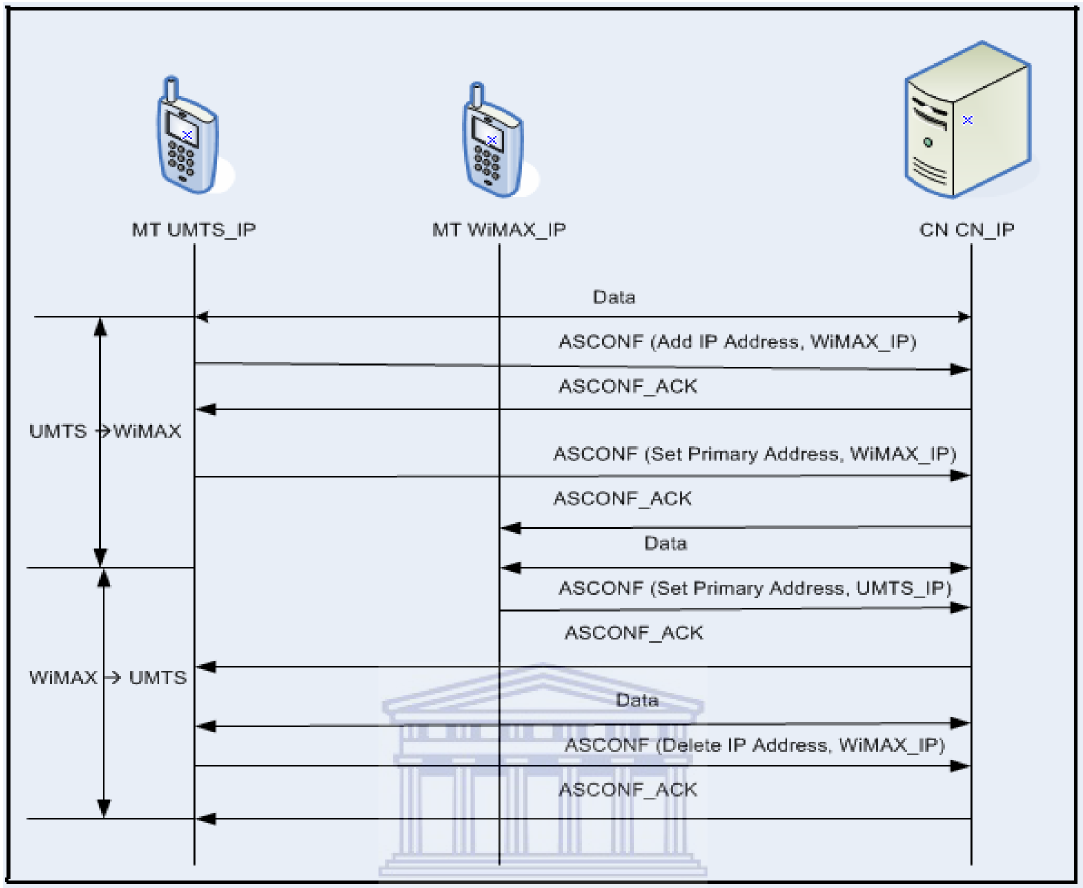


Figure 3.4: *The vertical handoff procedure for a single-homing CN*

Dual-homing CN – In this case, a CN is configured with two IP addresses, say, CN_IP_1 and CN_IP_2. The handoff procedure begins with WiMAX_IP and CN_IP_1 as the primary IP addresses of the MT and the CN, respectively. The dual-homing CN configuration vertical handoff scheme is similar to the single-homing CN vertical handoff and offers better flexibility as a consequence of having multiple paths between the endpoints. However, there are two differences in the add/delete IP process and the handoff triggering process. The main phases in the handoff process are explained below:

1. When the MT moves into a WiMAX cell covered by a UMTS cell, it obtains a new IP address, WiMAX_IP, either via DHCP or IPv6 auto-configuration.
2. The new IP address is signaled to the mSCTP stack.
3. The mSCTP on the MT notifies the CN of the new IP by sending an ASCONF message to the CN with parameters set to “Add IP Address” and WiMAX_IP.

4. The CN bundles an ASCONF to request the MT to add the CN's secondary IP address, CN_IP_2, into the association.
5. The MT then sends an ACK to confirm the completion of the add IP address process and the WiMAX becomes the primary choice, and the traffic between the MT and the CN is routed through the WiMAX.
6. If the MT loses the signal from the WiMAX cell, it starts the delete IP address process. The MT sends an ASCONF message to the CN with parameters set to "Delete IP Address" and WiMAX_IP to request that the CN release the address WiMAX_IP from its host routing table.
7. The CN bundles an ASCONF to request the MT to delete the CN's secondary IP address, CN_IP_2, from the association.
8. The CN sends an ACK to confirm the delete IP address process from the association.

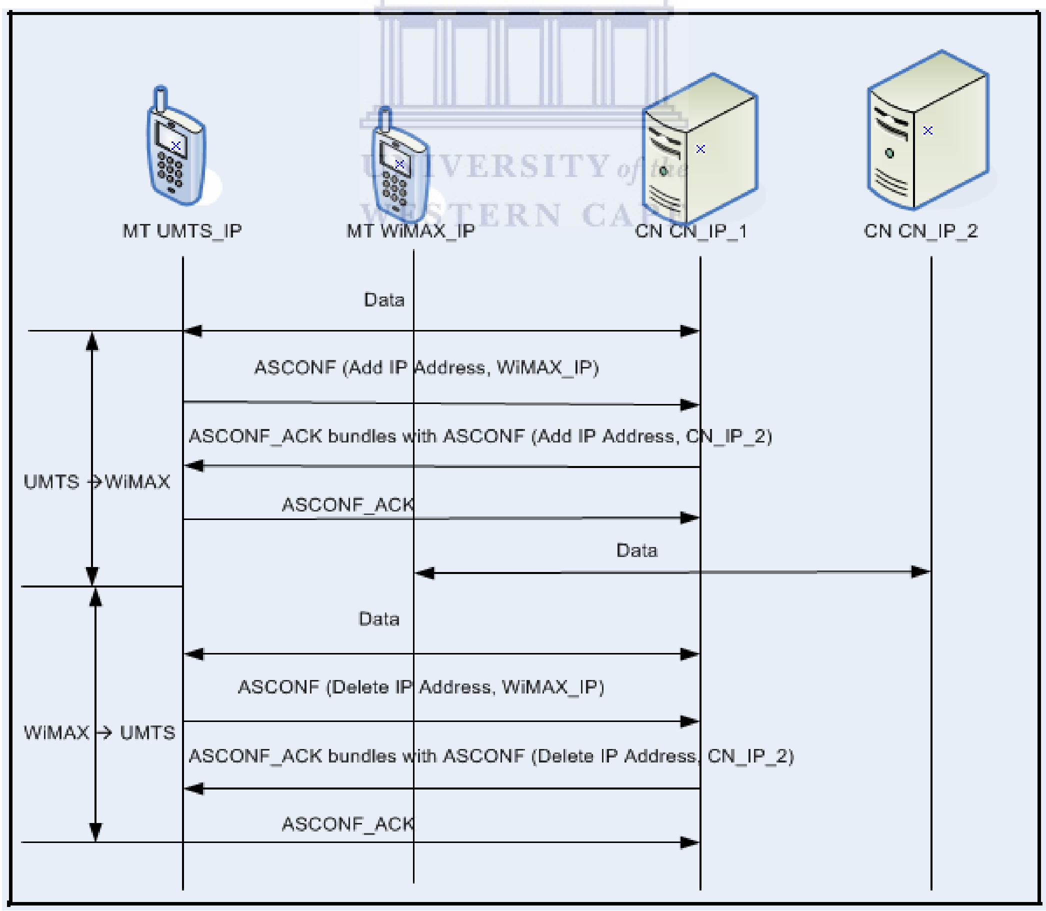


Figure 3.5: *The vertical handoff procedure for a dual-homing CN*

Scenario B: The fixed CN initiates an SCTP session with the MT.

This scenario applies to situations when a company manager needs to contact the field workforce through the fixed server (the correspondent node) in the company's virtual private network.

As indicated in Figure 3.6, the handoff scheme requires a location manager for the fixed CN to locate the current position of the MT when a new association is to be set up by the CN [64]. We are using SIP to provide the location management scheme. When the CN requests a call set up with the MT, the (home) SIP proxy server will interrogate the location database to locate the MT and then relay the SIP INVITE message to the (visiting) SIP Proxy server up to the MT. Once the SCTP association is established via the SIP signaling, the data transport between the two concerned hosts will be done according to the mSCTP handoff mechanisms.

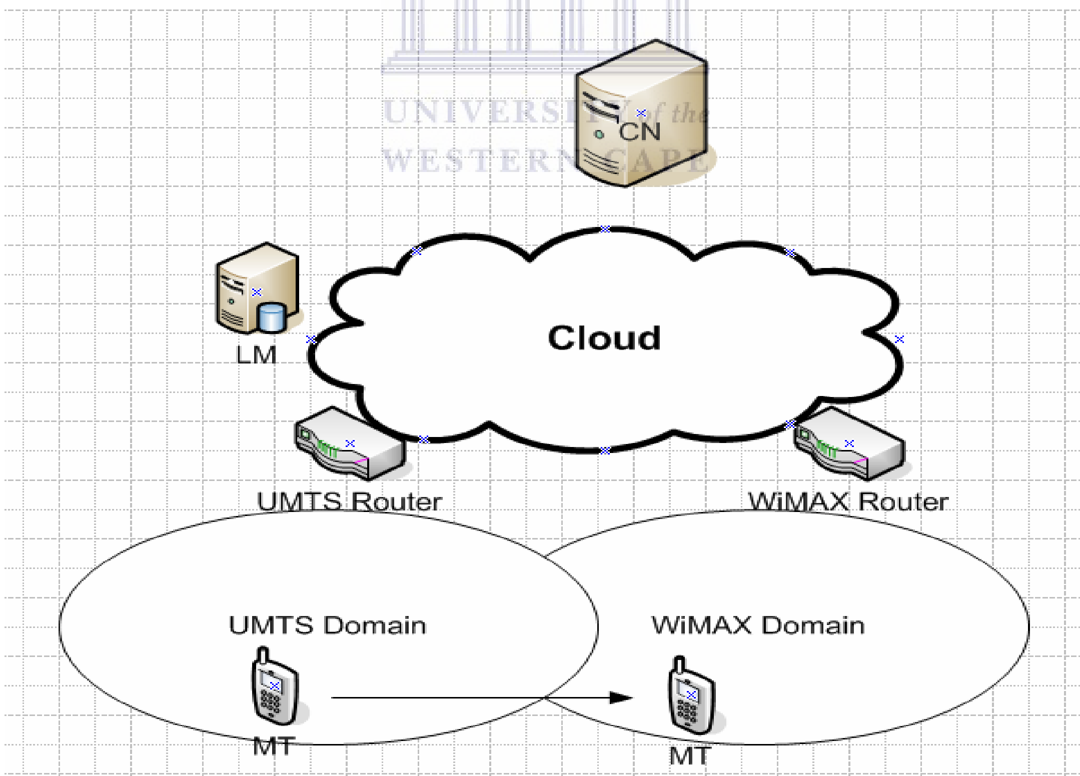


Figure 3.6: *The vertical handoff using mSCTP and SIP.*

3.5 An Intelligent and Optimal Vertical Handoff Decision Algorithm

We present our Vertical Handoff Decision Algorithm (VHDA) that is user-centric, context-aware, uses a fuzzy logic inference system (FIS) to process multiple vertical handoff initiation parameters (criteria), uses FMADM to choose an alternative from a set of alternative wireless access networks based on the classification of their imprecise attributes, and uses optimization techniques including the GA and the SA to maximize or minimize key parameters in order to select an optimum target access network for possible vertical handoff. As an incomplete knowledge representation technique, fuzzy logic is well suited for addressing imprecision issues. In the user-centric approach, users will have greater control and will be able to select the access network with which they are most satisfied. The VHDA is supposed to be context-aware in the sense that it should be aware of the MT conditions, possibilities offered by the access networks, and QoS requirements for a particular service. The VHDA is formulated as a FMADM problem since vertical handoff in an NGWN environment depends on multiple attributes.

3.5.1 Vertical Handoff Decision Criteria

Vertical handoff decision (VHD) in an NGWN environment is more complex and involves a tradeoff among many handoff metrics including quality of service (QoS) requirements (such as network conditions and system performance), mobile terminal conditions, power requirements, application types, user preferences, and a price model. Generally, the VHD criteria can be based on user preferences, operator preferences, or a combination of both. Using the metrics considered in Chapter 2 in a user-centric view involves the optimization of key criteria (attributes) including [21, 65, 66, 67]:

- *Signal strength*: Signal strength is used to detect the presence of and to indicate the availability of an access network, and an available network can be detected if its signal strength is good.

- *Network coverage*: Frequent handoffs incur delay and loss of packets. A network that provides a large coverage area enables mobile users to avoid frequent handoffs as they roam about. Coverage is measured in meters. For example, the network coverage areas for UMTS and Mobile WiMAX are 30 m to 20 Km and up to 50 Km respectively.
- *Data rate*: Different access network technologies offer different data rates. Data rate is usually measured in kbps or mbps. A network that can transfer signals at a high rate is preferred since a maximum data rate reduces service-delivery time for non-real-time services and enhances QoS for adaptive real-time services. As an illustration, a user watching a video clip encoded at 64 and 256 kbps, respectively, can see better pictures with better resolution at 256 kbps than at 64 kbps.
- *Service cost*: The cost of services offered and of using a particular access network is a major consideration to users and may affect the user's choice of access network and consequently handoff decision. A user may prefer to be connected through the cheapest available access network in order to reduce service cost incurred. Service cost is usually measured in unit price per second for real-time services and unit price per KB for non-real-time services.
- *Network reliability*: A reliable network is not error prone and so can be trusted to deliver a high level of performance.
- *Network security*: As strong security enhances information integrity, a network with high encryption is preferred when the information exchanged is confidential. The level of security is usually specified using linguistic terms such as very high, high, average, low, and very low.
- *MT velocity*: Handing off to an embedded network in an overlaid architecture of heterogeneous networks is discouraged when traveling at a high speed since a handoff back to the original network will occur very shortly afterward when the mobile terminal leaves the smaller embedded network. High mobile users are connected to the upper layers and benefit from a greater coverage area.
- *Battery power requirements*: Current multimode mobile devices consume energy while powered on. Power consumption should be minimized since mobile devices have limited battery power capabilities. When the battery level decreases, handing

off to a network with lower power requirements would be a better decision. Battery power consumption is usually specified using linguistic terms such as very low, low, average, high, and very high; and

- *Network latency*: Low network latency enhances quality of service since high network latency degrades applications and the transfer of information. Real-time services (e.g., voice call) are sensitive to delay. A handoff algorithm should be fast so that the mobile device does not experience service degradation or interruption. Network latency is usually measured in seconds.

These parameters may be of different levels of importance to the vertical handoff decision. They should be selected so that the handoff decision algorithm maximizes the QoS for the user and minimizes the use of system resources. Developing a handoff decision algorithm that takes into account and combines numerous parameters remains a research challenge. A next generation vertical handoff will have to incorporate many of the above criteria and others rather than only the signal strength in order to perform a handoff decision.

3.5.2 Components of the VHDA

Since a user in the heterogeneous NGWN may be offered connectivity from more than one access network at any time, we have to consider how the MT and/or the WON initiate a vertical handoff and select the most suitable wireless access network for serving the needs of the user. Thus a vertical handoff algorithm is needed to determine when to perform a handoff to a target network and to make an optimal choice of a wireless access network among all the available access networks.

A vertical handoff decision in an NGWN environment must solve the following problem: given a mobile user equipped with a contemporary multi-interfaced mobile device (with radio interfaces including UMTS, WLAN, WiMAX, and Digital Video Broadcasting-Handheld) connected to an access network, determine whether a vertical handoff should be initiated and dynamically select the optimum network connection from the available access network technologies to continue with an existing service. Consequently, our proposed VHDA consists of two parts [37]: a Handoff Initiation

Algorithm and an Access Network Selection Algorithm. To clarify the VHD idea, we focus on an interesting use case: vertical handoff between mobile WiMAX and UMTS.

The Handoff Initiation Algorithm uses a fuzzy logic inference system (FIS) to process multiple vertical handoff initiation parameters (criteria). We use a Mamdani FIS that is composed of the functional blocks [61]: a *fuzzifier*, a *fuzzy rule base*, a *database*, a *fuzzy inference engine*, and a *defuzzifier*. Since the inputs and outputs of the FIS are crisp in nature, the fuzzifier and defuzzifier are needed to transform them to and from fuzzy representation.

The Access Network Selection (ANS) algorithm involves decision making in a fuzzy environment. It can be solved using fuzzy multiple attribute decision making (FMADM) which deals with the problem of choosing an alternative from a set of alternatives based on the classification of their imprecise attributes. It applies a multiple attribute defined access network selection function (ANSF) to select the best access network that is optimized to the user's location, device conditions, service and application requirements, cost of service and throughput. This thesis proposes to use Genetic Algorithm (GA) and Simulated Annealing (SA) to optimize the ANSF with the goal of selecting the optimal access network.

The vertical handoff decision function is triggered when any of the following events occur: (a) when the availability of a new attachment point or the unavailability of an old one is detected, (b) when the user changes his/her profile, and thus altering the weights associated with the network selection attributes, and (c) when there is severe signal degradation or complete signal loss of the current radio link. Then the two-part algorithm is executed for the purpose of finding the optimum access network for the possible handoff of the already running services to the optimum target network.

The block diagram shown in Figure 3.7 describes the vertical handoff decision algorithm.

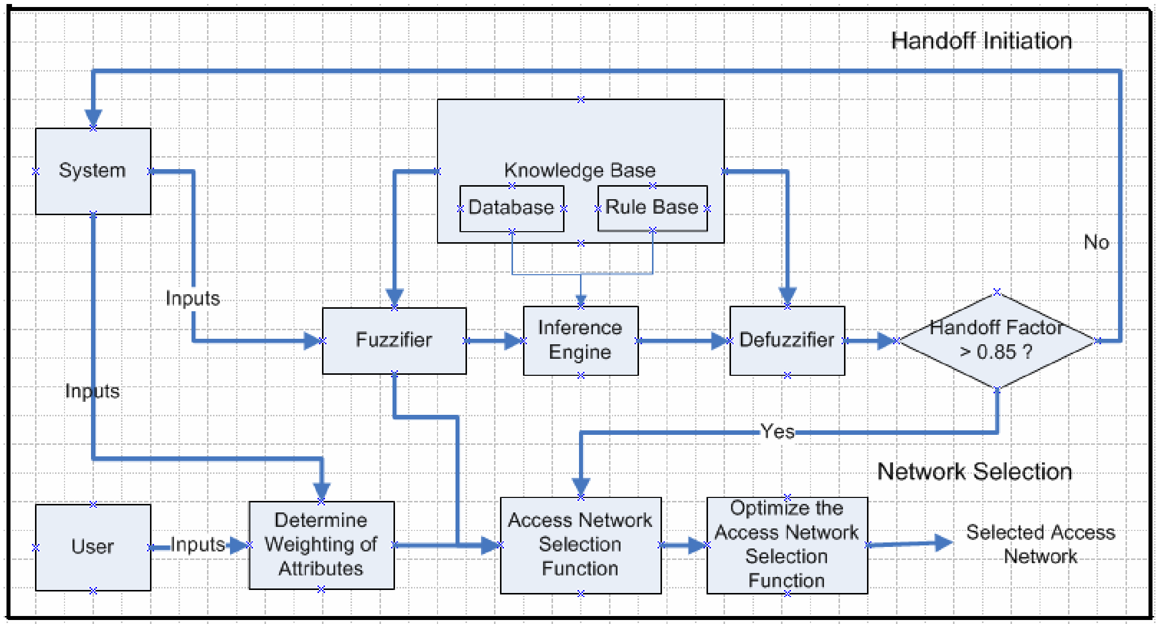


Figure 3.7: *Block diagram for Vertical Handoff Decision Algorithm*

3.6 Handoff Initiation Algorithm

Computing and choosing the correct time to initiate vertical handoff reduces subsequent handoffs, improves QoS, and limits the data signaling and rerouting that is inherent in the handoff process. To process vertical handoff-related parameters, we use fuzzy logic, which uses approximate modes of reasoning to tolerate vague and imprecise data that are affected by errors in precision and accuracy from measurements by the system. An FIS expresses mapping rules in terms of linguistic language. A Mamdani FIS can be used for computing accurately the handoff factor which determines whether a handoff initiation is necessary between an UMTS and WiMAX. We consider two handoff scenarios: handoff from UMTS to WiMAX, and handoff from WiMAX to UMTS. The proposed Handoff Initiation Algorithm (HIA) has a number of advantages including:

- **Fuzzy Logic Attributes**. Fuzzy logic is conceptually easy to understand and easy to apply. These advantages of fuzzy logic help in providing a simple solution to the HIA. Conflicting criteria can be resolved using fuzzy logic.
- **Use of Multiple Handoff Decision Criteria.** In a complex and uncertain heterogeneous NGNW environment, multiple criteria algorithms can reduce the uncertainty about the alternative access networks and allow a reasonable choice to

be made among them. The HIA is a multiple criteria problem in nature and a multiple criteria algorithm provides better performance than a single criterion algorithm due to the additional flexibility and complementary nature of the criteria.

3.6.1 Handoff from UMTS to WiMAX

Suppose that an MT that is connected to a UMTS network detects a new WiMAX network. A Mamdani FIS calculates the handoff factor which determines whether the MT should handoff to the WiMAX.

We use fuzzy modeling [68, 69] in the following steps to build the Mamdani FIS [70]:

Step 1: Identify and name the input linguistic variables and define their numerical ranges

Four input variables have been identified: the RSS indication (RSSI), data rate, network coverage area, and network latency of the target WiMAX network.

There are three ranges of the RSSI:

Table 3.1: RSSI (dBm)

Linguistic range	*Low*	*High*
Strong	-72	-66
Medium	-78	-66
Weak	-78	-72

There are three ranges of data rate:

Table 3.2: Data Rate (Mb/s)

Linguistic range	*Low*	*High*
High	30	60
Medium	0	60
Low	0	30

There are three ranges of network coverage area:

Table 3.3: Network Coverage (Km)

Linguistic range	*Low*	*High*
Good	25	50
Medium	0	50
Bad	0	25

There are three ranges of network latency:

Table 3.4: Network Latency (ms)

Linguistic range	*Low*	*High*
High	100	200
Medium	0	200
Low	0	100

Step 2: Identify and name the linguistic output variables and define their numerical ranges

There is one output variable identified: handoff factor. There are five ranges of handoff factor:

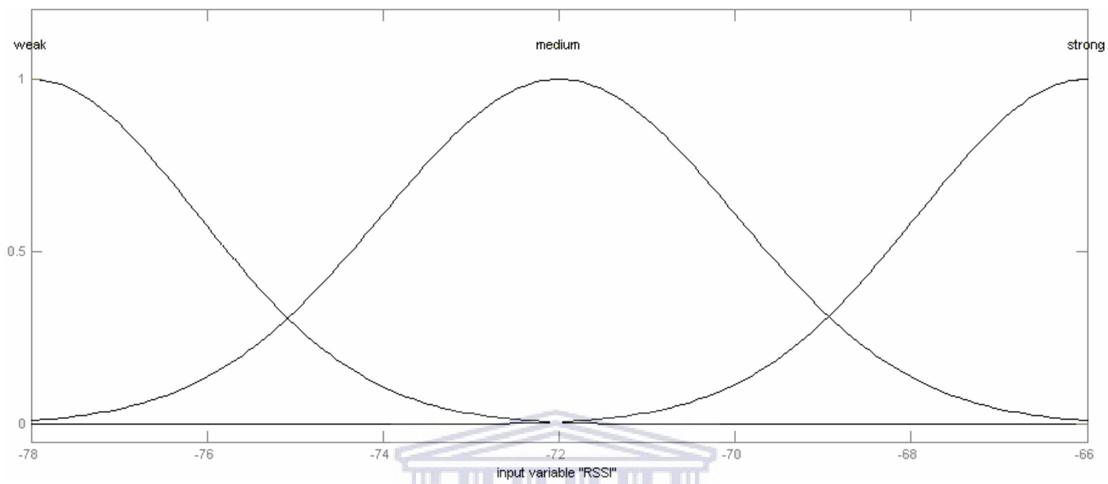
Table 3.5: Handoff Factor

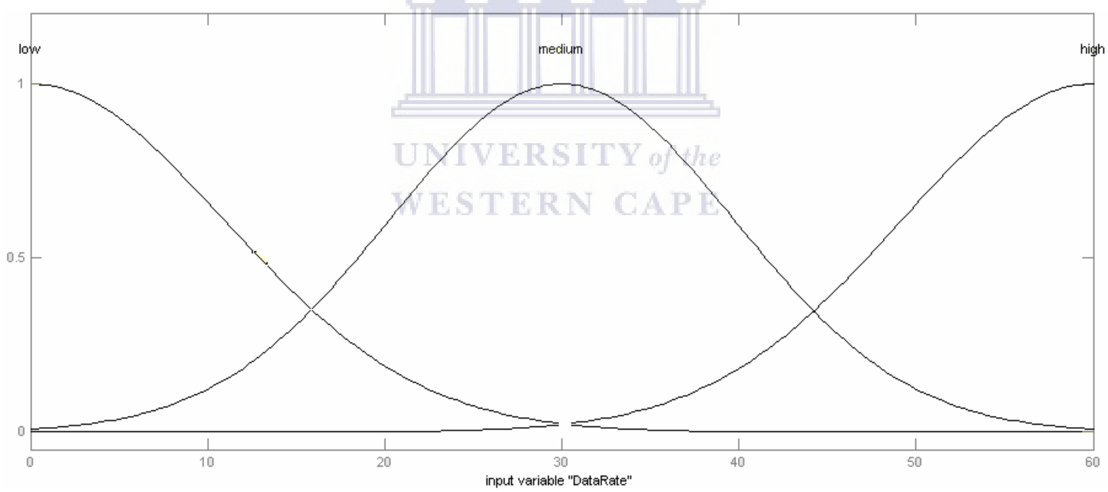
Linguistic range	*Low*	*High*
Higher	0.75	1
High	0.5	1
Medium	0.25	0.75
Low	0	0.5
Lower	0	0.25

Step 3: Define a set of fuzzy membership functions for each input variable and the output variable

We use Gaussian MFs because of the concise notation and smoothness of the Gaussian function. Each range of input and output variables is defined to associate with a fuzzy set that has the same name as the range. Therefore, there are three fuzzy sets defined for each input variable and five fuzzy sets for the output variable; for example, the fuzzy set values for the RSSI consist of the linguistic terms: Strong, Medium, and Weak. The low and high values of each range are used to define its associated set's Gaussian MFs. The universe of discourse for the fuzzy variable RSSI is defined from -78 dBm to -66 dBm. The fuzzy set "Strong" is defined from -72 dBm to -66 dBm with the maximum membership at -66 dBm. Similarly, the fuzzy set "Medium" for the RSSI is defined from -78 dBm to -66 dBm with the maximum membership at -72 dBm, and the fuzzy set "Weak" for the RSSI is defined from -78 dBm to -72 dBm with the maximum membership at -78 dBm. The universe of discourse for the variable Data Rate is defined from 0 Mbps to 60 Mbps, the universe of discourse for the variable Network Coverage is defined from 0 m to 50 Km, and the universe of discourse for the variable Network

Latency is defined from 0 ms to 200 ms. The fuzzy set values for the output decision variable Handoff Factor are Higher, High, Medium, Low, and Lower. The universe of discourse for the variable Handoff Factor is defined from 0 to 1, with the maximum membership of the sets "Lower" and "Higher" at 0 and 1, respectively.





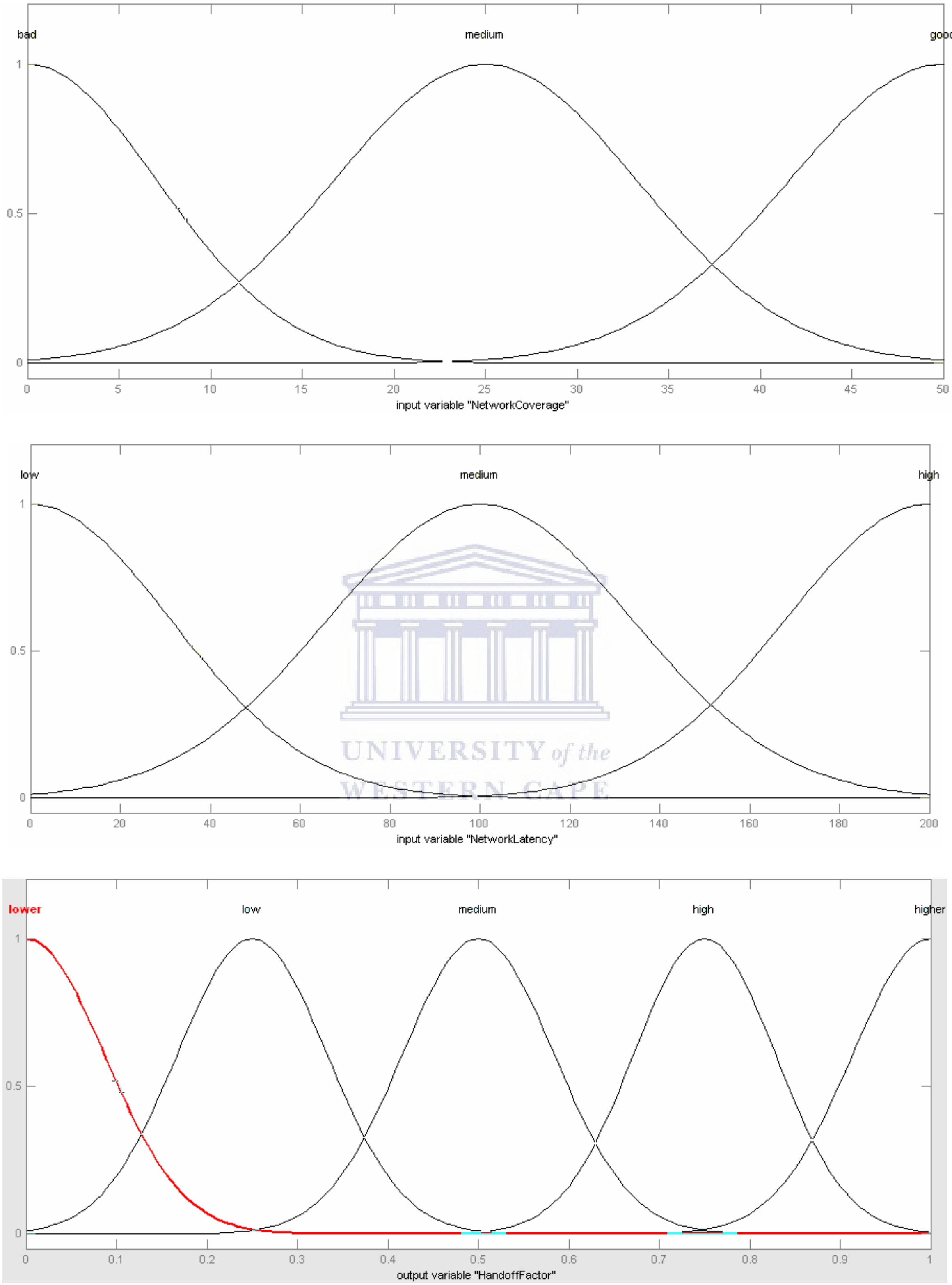


Figure 3.8: *Membership Functions*

Step 4: Construct the rule base that will govern the FIS's operation

Since there are four fuzzy input variables and three fuzzy sets for each fuzzy variable, the maximum possible number of rules in our rule base is $3^4 = 81$. The fuzzy rule base contains IF-THEN rules such as (see Table 3.6):

- IF RSSI is weak, and data rate is low, and network coverage area is bad, and network latency is high, THEN handoff factor is lower.
- IF RSSI is strong, and data rate is high, and network coverage area is good, and network latency is low, THEN handoff factor is higher.
- IF RSSI is strong, and data rate is medium, and network coverage area is good, and network latency is medium, THEN handoff factor is high.
- IF RSSI is medium, and data rate is medium, and network coverage area is good, and network latency is medium, THEN handoff factor is medium.

Table 3.6: Examples of Fuzzy Rules

Rule Number	RSSI	Data Rate	Network Coverage Area	Network Latency	Handoff Factor
1	Strong	High	Good	Low	Higher
2	Strong	High	Good	Medium	Higher
3	Strong	High	Good	High	High
4	Strong	High	Medium	Low	Higher
5	Strong	High	Medium	Medium	Higher
…	…	…	…	…	…
10	Strong	Medium	Good	Low	Higher
11	Strong	Medium	Good	Medium	High
…	…	…	…	…	…
14	Strong	Medium	Medium	Medium	Medium
15	Strong	Medium	Medium	High	Medium
…	…	…	…	…	…
23	Strong	Low	Medium	Medium	Medium
24	Strong	Low	Medium	High	Low
…	…	,,,	…	…	…
29	Medium	High	Good	Medium	High
30	Medium	High	Good	High	Medium
…	…	…	…	…	…
35	Medium	High	Bad	Medium	Low
36	Medium	High	Bad	High	Low
37	Medium	Medium	Good	Low	High
38	Medium	Medium	Good	Medium	Medium
39	Medium	Medium	Good	High	Medium
40	Medium	Medium	Medium	Low	High
…	…	…	…	…	…
50	Medium	Low	Medium	Medium	Low
51	Medium	Low	Medium	High	Low
52	Medium	Low	Bad	Low	Lower
53	Medium	Low	Bad	Medium	Low
54	Medium	Low	Bad	High	Low
55	Weak	High	Good	Low	Low
56	Weak	High	Good	Medium	Low
…	…	…	…	…	…
60	Weak	High	Medium	High	Lower
61	Weak	High	Bad	Low	Low
62	Weak	High	Bad	Medium	Lower
63	Weak	High	Bad	High	Lower
64	Weak	Medium	Good	Low	Low
…	…	…	…	…	…
79	Weak	Low	Bad	Low	Lower
80	Weak	Low	Bad	Medium	Lower
81	Weak	Low	Bad	High	Lower

In order to make our work easier we use the MATLAB Fuzzy Logic ToolBox [70] in the above stated Steps 1 through 4 and follow up with Steps 5 and through 8.

Step 5: Apply Fuzzy Operator

Step 6: Apply Implication Method

Step 7: Aggregate All Outputs

Step 8: Deffuzify

We deffuzify to obtain the output crisp handoff factor.

The crisp handoff factor computed after defuzzification is used to determine when a handoff is required as follows:

if *handoff factor* > 0.85, then initiate handoff;
otherwise do nothing.

3.6.2 Handoff from WiMAX to UMTS

The parameters that we are using in this directional handoff include the RSSI, data rate, network coverage area, and network latency of the current WiMAX network.

The design of the fuzzy inference system for this handoff scenario is similar to the design of the fuzzy inference system for the UMTS-to-WiMAX handoff. However, the fuzzy rule base contains IF-THEN rules which are the direct opposite of those in the UMTS-to-WiMAX handoff.

3.7 Case Study

The performance of the vertical handoff decision algorithm is tested within the framework of a scenario that simulates a typical day in the life of a telecommunication field technician, Mr. Alex Star. Mr. Star commutes from his home to his office and proceeds to carry out service requests in the residences of several clients of his company. Three cellular networks (GPRS_1, UMTS_1, and UMTS_2) cover the whole simulation area while three WMAN networks (WiMAX_1 (IEEE 802.16d), WiMAX_2 (IEEE 802.16e), and WiMAX_3 (IEEE 802.16e)) partly overlay the service area, and a number of WLANs are in hotspot areas.

3.7.1 Use Cases:

We present two use cases for the vertical handoff decision algorithm and examine the use cases for the handoff initiation component of the algorithm. The second component of the algorithm will be treated in Chapters 4, 5 and 6.

Case 1: As Mr. Star drives along the highway, he decides to download a large file containing a scientific report using the UMTS_2 network. After encountering traffic jam in the metro, he changes his user profile to complete downloading and start reading the report. In this case, the attributes are rated (on the Saaty Judgment Scale explained in Chapter 4) as: service cost (9), data rate (8), RSSI (7), security (6), reliability (5), network coverage area (4), mobile terminal velocity (3), network latency (2), and power requirement (1).

Case 2: Mr. Star arrives at Trendy Mall to perform service maintenance for City Computers. As the Company van is delayed, Mr. Star realizes that he will not be able to make it to the office on time for his afternoon meeting. Therefore, he initiates a video call, in order to participate in the meeting from the mall. The video call is allocated to WLAN_2 because WLAN technology is preferred as it offers high bit rates at low cost. As the driver starts driving the company van from the mall, Mr. Star is not completely satisfied with the current quality level of the video call service, and so he modifies his profile, specifying the maximum allowed quality level for this service. In this case, the attributes are rated as: data rate (9), service cost (8), network latency (7), RSSI (6), security (5), mobile terminal velocity (4), network coverage area (3), power requirement (2), and reliability (1).

3.7.2 Evaluation:

Case 1:

We first check to see whether a handoff should be initiated by calculating the handoff initiation factor.

Suppose that the MT records the data values of RSSI (dBm), Data Rate (Mbps), Network Coverage Area (Km), and Network Latency (ms) as {-66.2, 41.3, 36.1, 60.7}, {-67.3, 48.6, 42.3, 56.5} and {-67.01, 48.2, 42.5, 55.6} for WiMAX_1, WiMAX_2 and WiMAX_3 respectively.

We note that the data values for WiMAX_1 translate to the fuzzy IF-THEN rule

- IF RSSI is strong, and data rate is high, and network coverage area is good, and network latency is low, THEN handoff factor is higher (with actual value of 0.897).

The data values for WiMAX_2 and WiMAX_3 translate into similar fuzzy IF-THEN rules.

The three sets of values are fed into the FIS and we obtain the Handoff Factor values 0.897, 0.902 and 0.903 shown in Figure 3.9, Figure 3.10, and Figure 3.11, respectively. These values thus indicate the need to handoff to any of three WiMAX networks for the requested service.

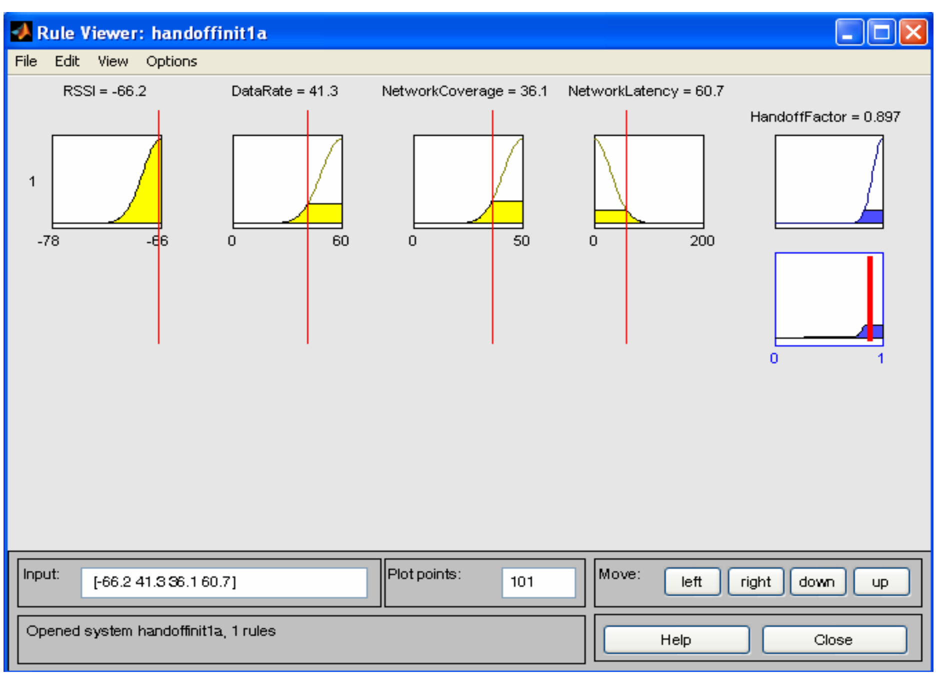


Figure 3.9: *Case 1 FIS Rule Viewer for WiMAX_1*

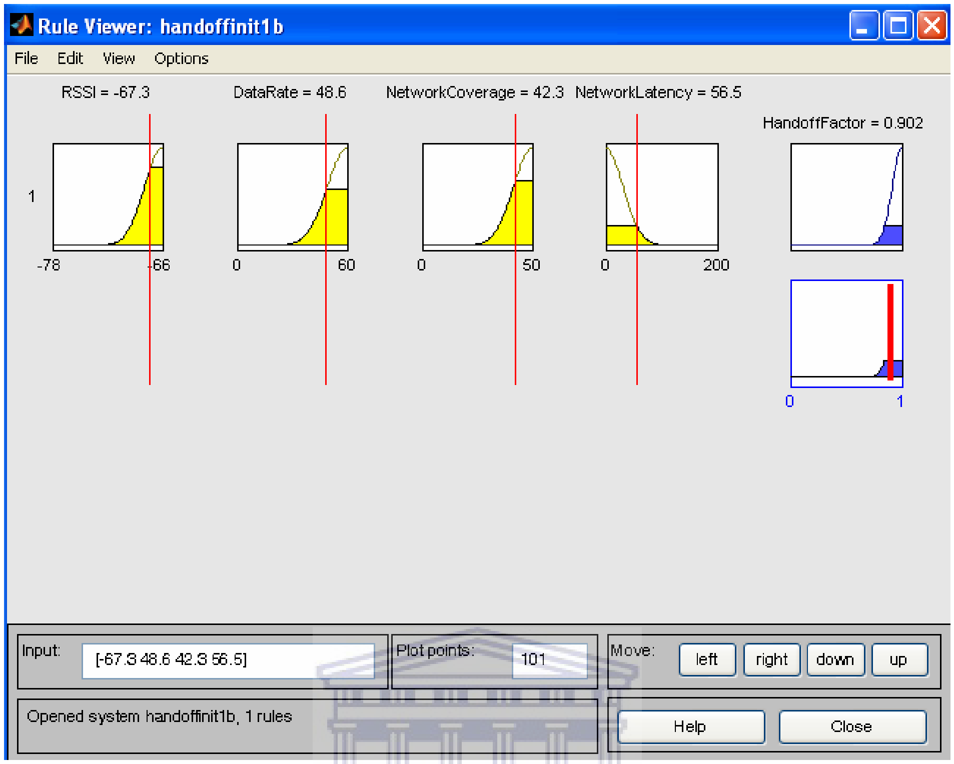


Figure 3.10: *Case 1 FIS Rule Viewer for WiMAX_2*

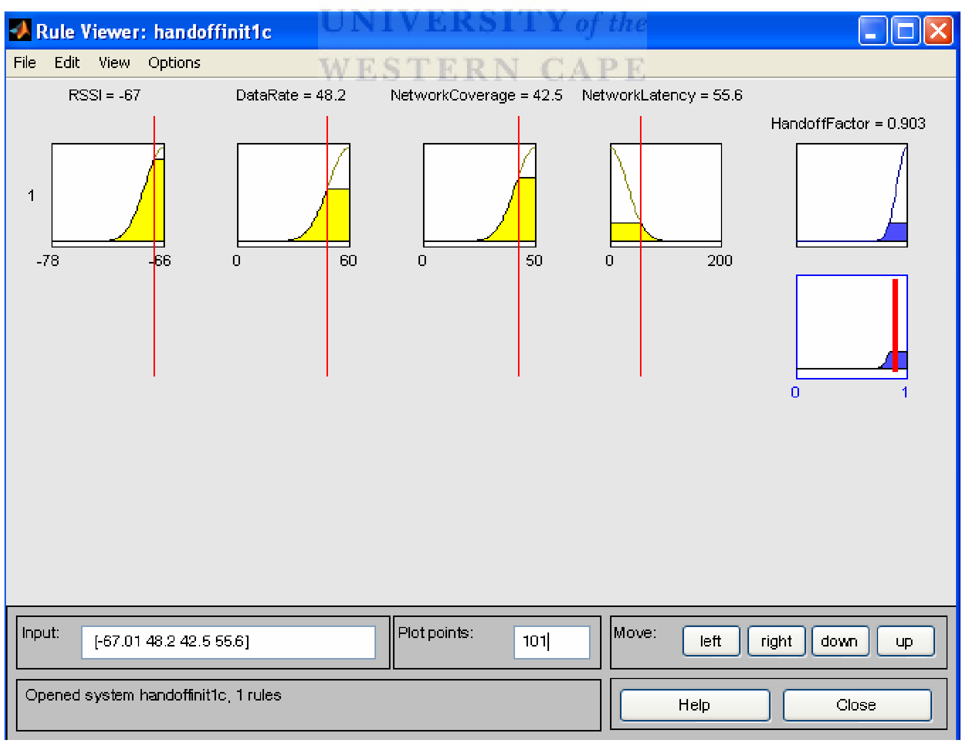


Figure 3.11: *Case 1 FIS Rule Viewer for WiMAX_3*

Case 2:

We first check to see whether a handoff should be initiated by calculating the handoff initiation factor.

Suppose that the MT records the data values of RSSI (dBm), Data Rate (Mbps), Network Coverage Area (Km), and Network Latency (ms) as {-69.4, 40.5, 35.7, 70.8}, {-68.6, 42.1, 38.6, 65.4} and {-68.1, 43.2, 40.3, 62.8} for WiMAX_1, WiMAX_2 and WiMAX_3 respectively. These set of values are fed into the FIS and we obtain the Handoff Factor values 0.886, 0.892 and 0.895 shown in Figure 3.12, Figure 3.13, and Figure 3.14, respectively. These Handoff Factor values indicate the need to handoff to any of the WiMAX networks for the requested service.

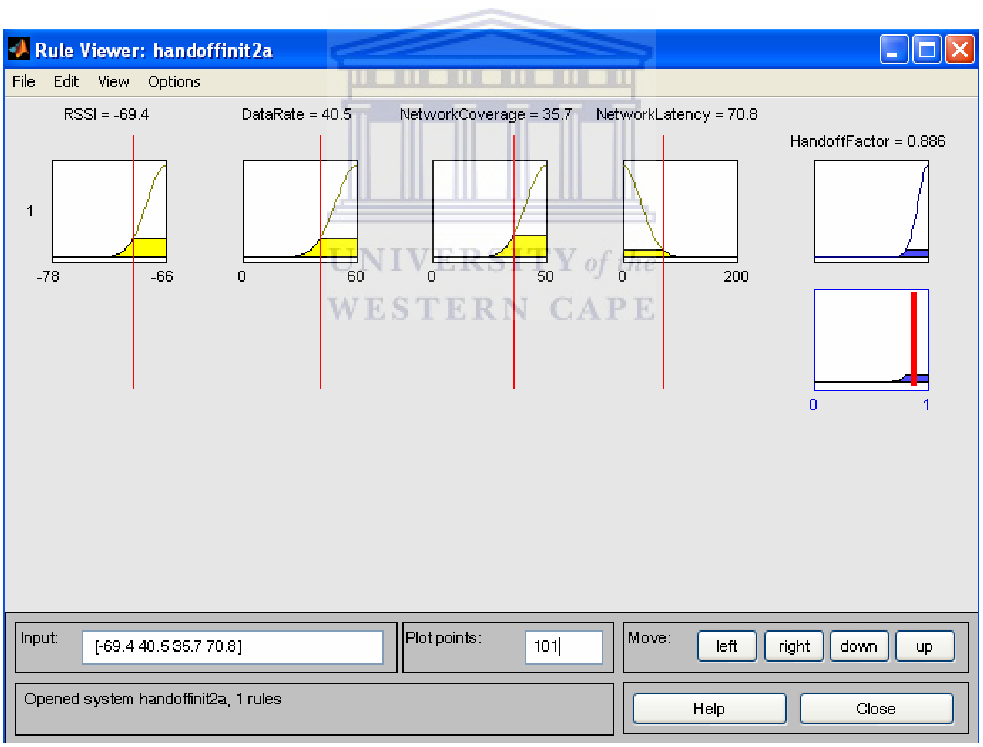


Figure 3.12: *Case 2 FIS Rule Viewer for WiMAX_1*

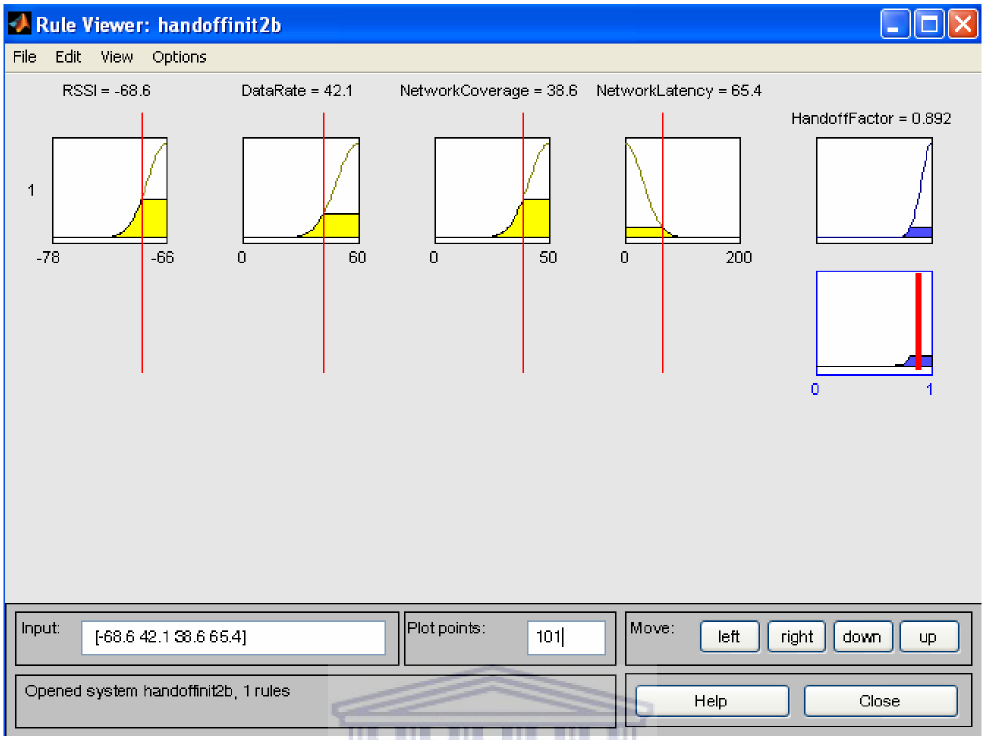


Figure 3.13: *Case 2 FIS Rule Viewer for WiMAX_2*

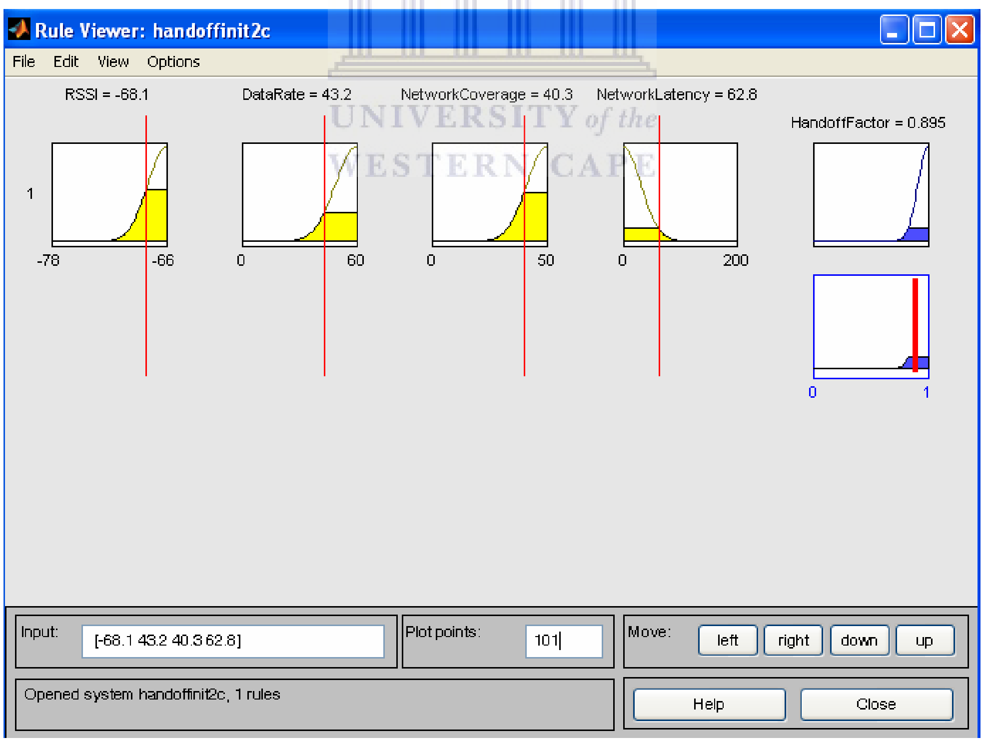


Figure 3.14: *Case 2 FIS Rule Viewer for WiMAX_3*

3.8 Summary

Fuzzy logic is proposed in this thesis as an effective means of meeting the challenges, such as imprecision and complexity constraints, characterizing the vertical handoff decision process as far as both knowledge representation and inferencing are concerned.

An introduction to the principal concepts and mathematical notations of fuzzy set theory and fuzzy inference systems was given in this chapter. The Chapter presented the design of a novel seamless vertical handoff management architecture that can achieve seamless mobility in the next generation heterogeneous wireless IP access networks for the mobile field workforce, and indeed for all mobile users, equipped with multimode mobile terminals. Our intelligent vertical handoff decision algorithm and its two components were then presented. The first component – vertical handoff initiation – was developed using a fuzzy inference system. The handoff initiation was evaluated for two use cases. As was discussed in section 3.6, the proposed Handoff Initiation Algorithm has a number of advantages.

CHAPTER 4

ACCESS NETWORK SELECTION USING FUZZY MULTIPLE ATTRIBUTES DECISION MAKING

The handoff decision problem is a typical multiple attributes decision making (MADM) problem since the handoff decision problem involves selecting a suitable access network among a number of candidate access networks with respect to several criteria (attributes). Formulating the vertical handoff decision as a fuzzy multiple attributes decision making (FMADM) problem allows the use of fuzzy logic to deal with imprecise and vague information that the decision attributes could contain as well as to combine and evaluate multiple attributes. We pose the access network selection problem as a mathematical optimization problem using an objective function that provides the means to quantify the weights or relative merits of satisfying the different parameters including the network, user and QoS requirements.

4.1 Introduction

To select the best available access network, a decision maker must consider several essential characteristics or attributes during the initial stage of the vertical handoff decision algorithm. Some of the data for these attributes might be tangible because they can be determined objectively; other data, however, are highly subjective as they represent linguistic information and are therefore intangible. In addition, the decision is often made under uncertainties, especially linguistic uncertainties. To rate and rank available access networks under uncertainties, we need a methodology that incorporates a mixture of both tangible and intangible attributes. The fuzzy multiple attributes decision making (FMADM) method offers an appropriate solution. Formulating the vertical handoff decision as a FMADM problem allows the use of fuzzy logic to deal with imprecise and vague information that the decision attributes could contain as well as to combine and evaluate multiple attributes. We pose the access network selection problem as a mathematical optimization problem using an objective function that provides the means to quantify the weights or relative merits of satisfying the different parameters including the network, user and QoS requirements. The judgment scale proposed by Saaty [72] is used to determine the attribute weights, and optimization methods are used to rank the available access networks.

4.2 Fuzzy Multiple Attributes Decision Making

4.2.1 Multiple Criteria Decision Making

Decision making with more than one criterion to be considered is an essential part of everyday life. Decision making may be characterized as a process of choosing or selecting 'sufficiently good' alternative(s) or course(s) of action, from a set of alternatives, to attain a goal or goals [29]. Multiple criteria decision making (MCDM) refers to making a decision in the presence of multiple and often conflicting criteria.
All MCDM problems share the following common characteristics [30]:

- Multiple criteria: Each MCDM problem has multiple criteria.
- Conflict among criteria: The multiple criteria in an MCDM problem usually conflict with each other.
- Incommensurable units: The multiple criteria may have different units of measurement.
- Design or selection: The goal of an MCDM problem is either to design the best alternative(s) or to select the best one(s) among a pre-specified finite set of alternatives.

The MCDM method is sometimes applied to decisions involving multiple attributes or multiple objectives, but generally when they both apply. Thus the MCDM can be divided into multiple attributes decision making, and multiple objective decision making (MODM).

Multiple attributes decision making (MADM) deals with the problem of choosing an alternative from a finite and countable set of alternatives that are characterized in terms of their multiple attributes.

A typical MADM problem is formulated as [29]:

$$\text{select } A_i \text{ from } A_1, A_2, \dots, A_m$$
$$\text{using } C_1, C_2, \dots, C_n \tag{4.1}$$

where $\{A_1, A_2, \dots, A_m\}$ denotes m alternatives, and $\{C_1, C_2, \dots, C_n\}$ represents n criteria or attributes. The selection is usually based on maximizing a multiple attributes utility function.

The decision about access network selection in a heterogeneous wireless environment can be solved using specific multiple attributes decision making (MADM) algorithms such as:

- Weighted Sum Model (WSM) or Simple Additive Weighting (SAW) – If there are m alternatives and n criteria then the best alternative is the one that satisfies the following expression:

$$A_{WSM} = \sum_{i=1}^{n} w_j \cdot a_{ij}, \quad \text{for } i = 1, 2, \dots, m. \tag{4.2}$$

  where A_{WSM} is the WSM score of the best alternative, a_{ij} is the actual value of the i-th alternative in terms of the j-th criterion, and w_j is the weight of importance of the j-th criterion. The overall score of a candidate network is determined by the weighted sum of all the attribute values;
- Weighted Product Model (WPM) or Multiplicative Exponent Weighting (MEW) – Each alternative is compared with the others by multiplying a number of ratios, one for each criterion. Each ratio is raised to the power equivalent to the relative weight of the corresponding criterion. In general, in order to compare the alternatives A_K and A_L, the following product is obtained:

$$R(A_K / A_L) = \prod_{j=1}^{n} (a_{Kj} / a_{Lj})^{wj}, \tag{4.3}$$

  where n is the number of criteria, a_{ij} is the actual value of the *i-th* alternative in terms of the *j-th* criterion, and w_j is the weight of importance of the j-th criterion.

If $R(A_K / A_L)$ is greater than or equal to one, then alternative A_K is more desirable than alternative A_L. The best alternative is the one that is better than or at least equal to all the other alternatives;

- Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) – It is a method for ranking a finite number of alternatives using a number of decision criteria. The chosen candidate network using this method is the one which is closest to the ideal solution and farthest from the worst case solution;
- Analytic Hierarchy Process (AHP) – The AHP method is used as a means of finding the optimal solution to a complex decision problem. It first decomposes the network selection problem into a system of hierarchies, that is, into several sub-problems, and then elicitates pairwise judgments of criteria and assigns a weight value for each sub-problem; and
- Grey Relational Analysis (GRA) – The GRA method builds grey relationships between elements of two series to compare each member quantitatively. It is used to rank the candidate networks and select the one with the highest ranking.

Multiple objective decision making (MODM) consists of a set of conflicting goals that cannot be achieved simultaneously [29]. Usually it concentrates on continuous decision spaces, has several objective functions, and can be solved with mathematical programming techniques [73].

An MODM problem can be formulated as [74]:

$$\begin{aligned} &\text{maximize:} \quad \mathbf{f}(\mathbf{x}) \\ &\text{subject to:} \quad \mathbf{x} \in \mathrm{R}\,,\ \mathbf{g}(\mathbf{x}) \leq \mathbf{b},\ \mathbf{x} \leq 0 \end{aligned} \tag{4.4}$$

where $\mathbf{f}(\mathbf{x})$ represents n conflicting objective functions, $\mathbf{g}(\mathbf{x}) \leq \mathbf{b}$ represents m constraints in continuous decision spaces, and $\mathbf{x}$ is an m-vector of decision variables.

4.2.2 Fuzzy Decision Making

In this thesis, we focus on multiple attributes decision making (MADM) under uncertainties, especially linguistic uncertainties.

Fuzzy logic can be combined with the MADM method to develop *fuzzy MADM (FMADM)* algorithms for applications with imprecise data that cannot be handled efficiently by using classical MADM methods. Formulating the vertical handoff decision as a FMADM problem allows the use of fuzzy logic to deal with imprecise and vague information that the decision attributes could contain as well as to combine and evaluate multiple attributes. The principal elements of the FMADM problem are [69]:

- **Alternatives** – the aim of the decision process is the selection of an alternative A_i from a set $A = \{A_1, A_2, \dots, A_m\}$ of possible alternatives.
- **Decision criteria (attributes)** – The decision is taken based upon a number of decision criteria, say, a set $C = \{C_1, C_2, \dots, C_n\}$ that defines each alternative. The best alternative is selected according to the degree to which the decision criteria are satisfied. In fuzzy decision making, the criteria consist of the decision goals and the constraints.
- **Membership values** – In order to be able to evaluate the alternatives based on the criteria, it should be known how the preference of the decision maker varies with the variable form of the criterion. This information leads to a judgment for each alternative for each criterion which indicates the desirability of the alternative in the corresponding criterion. In fuzzy decision making, this judgment information is usually provided in the form of the degree of membership of alternative A_j in the criterion C_i, denoted $\mu_{Ci}(A_j)$, which is the degree to which alternative A_j satisfies the criterion C_i.
- **Weight factors** – Usually the decision criteria are not equally important for a particular decision problem. The difference in the importance of the criteria can be modeled by the assignment of different weight factors to different criteria in the decision function.
- **Performance function** – A performance function orders the set of alternatives according to the criteria.
- **Ranking of alternatives** – an ordering mechanism is needed for the ranking of alternatives.

The most common solution procedures for the rating of alternatives based on the relative merit of the criteria (attributes) in a FMADM problem are weighting average methods. Approaches to solving the rating of alternatives in the FMADM include: the conjunction implication method of Bellman and Zadeh [75], the conjunction implication method of Yager [72, 76], and weighting average methods.

The Conjunction Implication Method of Bellman and Zadeh: Bellman and Zadeh [75] proposed the first decision making model where goals and constraints are treated as fuzzy sets. A fuzzy decision D can be defined as the choice that satisfies simultaneously the goals G and constraints C.

Formally, let $A = \{A_1, A_2, \ldots , A_m\}$ be a set of m alternatives and $C = \{C_1, C_2, \ldots , C_n\}$ be a set of n decision criteria (attributes) that can be expressed as fuzzy sets in the space of alternatives. The degree of membership of alternative A_j in the criterion C_i, denoted $\mu_{Ci}(A_j)$, is the degree to which alternative A_j satisfies this criterion. Then, the decision function with the relative importance of the criteria is

$$\mu_D (A_i) = \min_{A_i \in A} \{\mathrm{w}_1 \cdot \mu_{C1}(A_i), \mathrm{w}_2 \cdot \mu_{C2}(A_i), \ldots, \mathrm{w}_n \cdot \mu_{Cn}(A_i)\}, \tag{4.5}$$

with

$$\sum_{j=1}^{n} w_j = 1 \tag{4.6}$$

where $\mu_D(A_i)$ can be interpreted as the degree to which the alternative A_i satisfies the criteria. The optimal decision, A_i*, is the alternative that satisfies

$$\mu_D (A_i^*) = \max_{A_i \in A} (\mu_D (A_i)) \tag{4.7}$$

The Conjunction Implication Method of Yager: Yager's method assumes the max-min principle of Bellman and Zadeh but represents the importance of criteria as exponential scalars. Hence, the fuzzy set decision is the intersection of all criteria [29, 72, 76, 77]:

$$\mu_D (A_i) = \min_{A_i \in A} \{(\mu_{C1}(A_i))^{\mathrm{w}1}, (\mu_{C2}(A_i))^{\mathrm{w}2}, \ldots, (\mu_{Cn}(A_i))^{\mathrm{w}n}\}, \tag{4.8}$$

where $\mu_D(A_i)$ can be interpreted as the degree to which the alternative A_i satisfies the criteria. The optimal decision, A_i*, is the alternative that satisfies

$$\mu_D (A_i^*) = \max_{A_i \in A} (\mu_D (A_i))$$

A Weighted Sum Method: A common method for a fuzzy decision D is a weighted combination of n criteria such as

$$\mu_D(x) = \sum_{j=1}^{n} w_j \cdot \mu_{Cj}(x). \tag{4.9}$$

All the three methods were used in the research work. However, a weighted sum method is used in the thesis report as it is suitable to design an objective function that can be optimized by optimization algorithms.

4.3 Access Network Selection Algorithm

4.3.1 Problem Definition and Modeling

A suitable access network has to be selected once the handoff initiation algorithm indicates the need to handoff from the current access network to a target network. This section is concerned with the access network selection (ANS) problem: how to choose the best or optimal access network(s) among the available access networks for a particular service. The network selection decision process is formulated as a FMADM problem that deals with the evaluation of a set of alternative access networks using a multiple attribute access network selection function (ANSF) defined on a set of attributes. Fuzzy logic is used to deal with imprecise data that the decision attributes might contain, and to combine and evaluate multiple attributes simultaneously.

When designing the access network selection algorithm, an objective function or optimization criterion has to be identified, as well as a set of design constraints. The access network selection algorithm is modeled as an objective function that includes the rules for selecting the best candidate access network. A desirable objective function is the ANSF that measures the efficiency in utilising radio resources and the improvement in QoS to mobile users gained by handing off to a particular network. It is defined for all alternative target access networks that cover the service area of a user. The network that provides the highest ANSF value is selected as the best network to handoff from the current access network according to the mobile terminal conditions, network conditions, service and application requirements, cost of service, and user preferences.

The ANSF is triggered when any of the following events occur: (a) a new service request is made; (b) a user changes his/her preferences; (c) the MT detects the availability of a new network; (d) there is severe signal degradation or complete signal loss of the current radio link.

Parameters (attributes) used for the ANSF include the signal strength (*S*), network coverage area (*A*), data rate (*D*), service cost (*C*), reliability (*R*), security (*E*), battery power (*P*), mobile terminal velocity (*V*), and network latency (*L*). The parameter values are obtained as the MT collects information including characteristics of serving and neighbouring access networks, available access interfaces, terminal movement speed, and service cost.

The main objective of the network selection algorithm is to determine and select an optimum cellular/wireless access network for a particular high-quality service that can satisfy the following goals:

- *Good signal strength*: An available network can be detected if its signal strength is good;
- *Good network coverage*: The widest network coverage reduces handoff frequency for highly mobile users;
- *Optimum data rate*: A maximum data rate enhances the quality of service for adaptive real-time applications, and reduces the service delivery time for non real-time applications;
- *Low service cost*: A low service cost reduces the total service cost incurred by a subscriber;
- *High network reliability*; A network with high reliability is not error prone and so can be trusted to deliver a high level of performance.
- *Strong network security*: A strong network security enhances information integrity;
- *Good MT velocity*: A good mobile terminal velocity would be suitable to users on the move including occupants in moving vehicles;
- *Low battery power requirements*: Low battery power consumption increases battery lifetime and reduces recharge frequency; and
- *Low network latency*: Low network latency enhances quality of service since high network latency degrades applications and the transfer of information.

Selecting an optimum cellular/wireless access network for a particular high-quality service that can satisfy the above goals may prove difficult as some parameters may conflict. The complexity of the access network selection based on multiple attributes could be overcome through user preferences. We pose the access network selection problem as a mathematical optimization problem using an objective function that provides the means to quantify the relative merits of satisfying the different parameters including the network, user and QoS requirements.

The optimum wireless access network in the ANS problem must satisfy:

$$\underset{\boldsymbol{u}}{\text{maximize}}\, f_i(\boldsymbol{u}),$$

where $f_i(\boldsymbol{u})$ is the objective function evaluated for the network i and $\boldsymbol{u}$ is the vector of input parameters. The function f_i can be mathematically expressed as:

$$\begin{aligned} f_i(\boldsymbol{u}) &= f\,(S_i, A_i, D_i, 1/C_i, R_i, E_i, V_i, 1/P_i, 1/L_i) \\ &= \sum_{i=1}^{6} w_X \cdot N_f\,(X_i) + \sum_{i=1}^{3} w_Y \cdot N_f\,(1/Y_i), \end{aligned} \tag{4.10}$$

where $N_f(X)$ is the normalized function of the parameter X and w_X is the weight which indicates the importance of the parameter X, with $X_i = S_i, A_i, D_i, R_i, E_i, V_i$, and $Y_i = C_i, P_i, L_i$. Normalization is needed to ensure that the sum of the values in different units is meaningful.

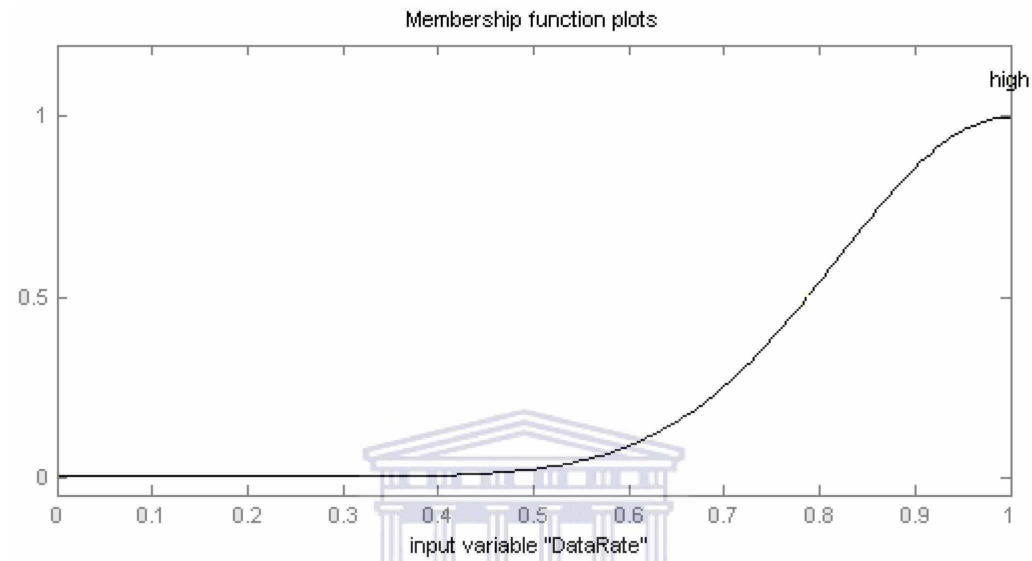
Each weight is proportional to the significance of the associated parameter to the decision maker in the VHDA. The more important a parameter is to the user, the larger the weight of that specific parameter and vice versa. The values of these weights range from 0 to 1 and add up to 1.

A simple way to obtain $N_f\,(X)$ is normalization with respect to the maximum or minimum values of the real-valued parameters. Therefore, we have

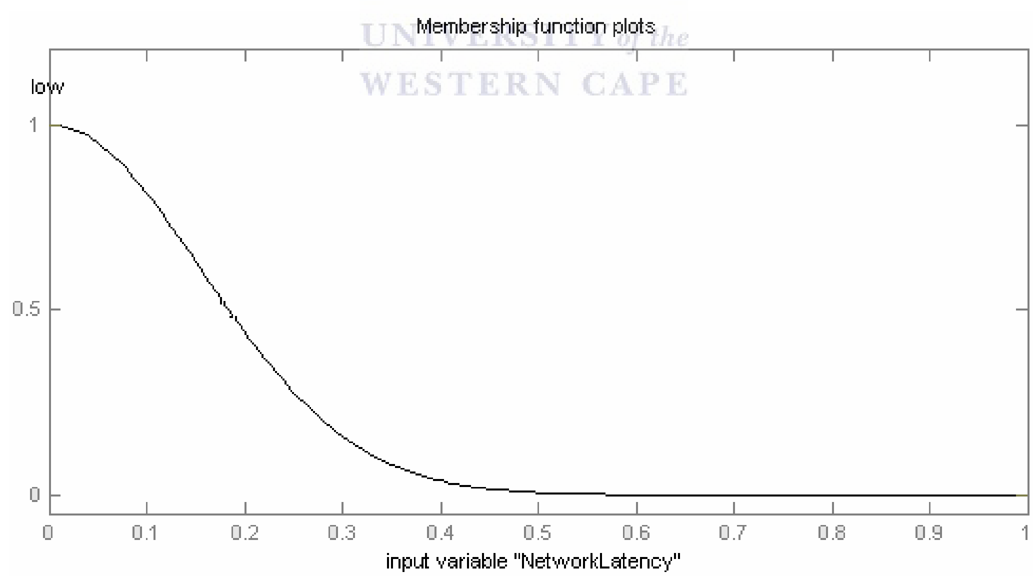
$$f_i(\boldsymbol{x}) = \sum_{i=1}^{6} w_X \cdot (X_i/X_{max}) + \sum_{i=1}^{3} w_Y \cdot (Y_{min}/Y_i) \tag{4.11}$$

Since we are using FMADM, a suitable normalized function of the parameter X is the fuzzy membership function μ_X. In order to develop this function, data from the system are fed into a fuzzifier to be converted into fuzzy sets. The values of the parameters are normalized between 0 and 1. Then a single membership function is defined such that

$\mu_{Cj}(0) = 0$ and $\mu_{Cj}(1) = 1$ if the goal is to select a network with a high parameter X value; and such that $\mu_{Cj}(0) = 1$ and $\mu_{Cj}(1) = 0$ if the goal is to select a network with a low parameter X value. For example, Figure 4.1a and Figure 4.1b show the membership function plots for high data rate and low network latency, respectively.



(a) Objective: High Data Rate



(b) Objective: Low Network Latency

Figure 4.1: *Membership Function plots for Data Rate and Network Latency*

In general, the fitness value for the network i is given by the final form of the weighting linear function f_i defined as

$$f_i(\boldsymbol{x}) = \sum_{j=1}^{n} w_j \cdot \mu_{Cj}(A_i), \qquad (4.12)$$

where $\boldsymbol{x}$ is the vector of membership function values and w_j are the attribute weights which signify the importance of the attributes.

Having described the attributes and the ANSF, the next issue is how to determine the attribute weights.

Determination of Attribute Weights: Data from the system are fed into a fuzzifier to be converted into fuzzy sets. Suppose that $A = \{A_1, A_2, \ldots, A_m\}$ is a set of m alternatives and $C = \{C_1, C_2, \ldots, C_n\}$ is a set of n handoff decision criteria (attributes) that can be expressed as fuzzy sets in the space of alternatives. The criteria are rated on a scale of 0 to 1. The degree of membership of alternative A_j in the criterion C_i, denoted $\mu_{Ci}(A_j)$, is the degree to which alternative A_j satisfies this criterion. A decision maker (such as a field worker or a mobile user) translates the available data into paired comparisons by answering a basic question: Given a specific criterion and two alternatives, A_i and A_j, which one is preferred and to what level of intensity? The decision maker makes this comparison linguistically and assigns the values a_{ij} using the judgment scale proposed by Saaty [72], with the less important element used as the unit and the more important element assigned a value from this scale as a multiple of that unit:

- 1 – equal importance (A_i and A_j are equally important);
- 3 – weak importance over one another (A_i is weakly more important than A_j);
- 5 – strong importance over one another (A_i is strongly more important than A_j);
- 7 – demonstrably more or very strong importance over one another (A_i is demonstrably more important than A_j);
- 9 – absolutely or extremely more important over one another (A_i is absolutely more important than A_j);
- 2, 4, 6, and 8 – represent compromise judgments, that is, intermediate values between two adjacent judgments.

Table 4.1 summarises the judgment scale of Saaty mentioned above.

Table 4.1: *The Judgment Scale of Saaty*

Intensity of importance	Definition	Explanation
1	Equal importance	Two elements contribute equally to the objective
2	Slightly weak importance	Between Equal and Weak
3	Weak importance	Experience and judgment slightly favour one element over the other
4	Less strongly importance	Between Weak and Strong
5	Strong importance	Experience and judgment strongly favour one element over another
6	Stronger importance	Between Strong and Very Strong
7	Very strong importance	One element is favoured very strongly over another; its dominance is demonstrated in practice
8	Very, very strong importance	Between Very Strong and Absolutely
9	Absolutely important	The evidence favouring one element over the other is of the highest possible order of affirmation

Possible values of a_{ij} are 1, 2, 3, 4, 5, 6, 7, 8, 9, 1/2, 1/3, 1/4, 1/5, 1/6, 1/7, 1/8, 1/9. Depending on the criteria that are being compared the following cases exist:

1. $a_{ii} = 1$, since an attribute is compared with itself.
2. $a_{ij} > 1$ when attribute i is considered to be more important than attribute j.
3. $a_{ij} < 1$ when attribute j is considered to be more important than attribute i.

An n x n matrix $\boldsymbol{B}$ of pairwise comparison judgments about the relative importance or preference between any two criteria is constructed so that:

$$\begin{cases} b_{ii} = 1 \\ b_{ij} = a_{ij},\ i \neq \mathrm{j} \\ b_{ji} = 1/b_{ij} \end{cases} \tag{4.13}$$

The value b_{ij} of this matrix represents the relative importance of the ith criterion over the jth criterion.

Using this matrix, the unit eigenvector, $\boldsymbol{V}$, corresponding to the maximum eigenvalue, λ_{max}, of $\boldsymbol{B}$ is then determined by solving the equation:

$$\boldsymbol{B} \cdot \boldsymbol{V} = \lambda_{max} \cdot \boldsymbol{V} \tag{4.14}$$

Finding the unit eigenvector, $\boldsymbol{V}$, corresponding to the maximum eigenvalue of $\boldsymbol{B}$ produces the cardinal ratio scale of the compared attributes. The eigenvectors are then normalized to ensure consistency. In other words, the values of $\boldsymbol{V}$ are scaled for use as factors in weighting the membership values of each attribute by a scalar division of $\boldsymbol{V}$ by the sum of values of $\boldsymbol{V}$ to obtain a weighting vector or vector of priorities

$$\boldsymbol{W} = \{w_j \mid 0 \le w_j \le 1, \sum_{j=1}^{n} w_j = 1\}. \tag{4.15}$$

The optimum wireless network is given by the optimization problem:

$$\max f_i(\boldsymbol{x}) = \max\{\sum_{j=1}^{n} w_j \cdot \mu_{Cj}(A_i)\} \tag{4.16}$$

such that

$$0 \le w_j \le 1, \text{ and } \sum_{j=1}^{n} w_j = 1, \tag{4.17}$$

and

$$\{\mu_{Cj}(A_i)\}_{\min} \le \mu_{Cj}(A_i) \le \{\mu_{Cj}(A_i)\}_{\max}. \tag{4.18}$$

The last inequality constraint arises from the fact that every attribute is bounded either by an acceptable minimum threshold value and the possible maximum value, or by the possible minimum value and an acceptable maximum threshold value.

The MT calculates the handoff initiation factor in the handoff initiation algorithm when the MT detects a new network or the user changes his/her preferences or the current radio link is about to drop. If the handoff initiation algorithm indicates the need for a handoff of the already running services from the current network to a target network, the mobile terminal then calculates the ANSF f_i for the current network and target networks. Vertical handoff takes place if the target network receives a higher f_i.

4.3.2 Solution Approach

We shall first solve the access network selection problem by using a mathematical optimization algorithm. A number of standard algorithms can be used for solving problems such as unconstrained optimization, least squares optimization, constrained optimization, and multi-objective optimization. We evaluated the MATLAB Optimization Toolbox functions and settled on the solver "fmincon" which finds the minimum of a constrained nonlinear multivariable function for a problem specified by

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$

such that

$$\begin{cases} c(\boldsymbol{x}) \leq 0 \\ ceq(\boldsymbol{x}) = 0 \\ \boldsymbol{A} \cdot \boldsymbol{x} \leq \boldsymbol{b} \\ \boldsymbol{Aeq} \cdot \boldsymbol{x} = \boldsymbol{beq} \\ \boldsymbol{lb} \leq \boldsymbol{x} \leq \boldsymbol{ub}, \end{cases} \tag{4.19}$$

where $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ is a vector of decision variables, $\boldsymbol{b}$ is a vector of numbers, $\boldsymbol{beq}$ is a vector for linear equality constraints, $\boldsymbol{lb}$ is a vector of lower bounds, $\boldsymbol{ub}$ is a vector of upper bounds, $\boldsymbol{A}$ is a matrix, $\boldsymbol{Aeq}$ is a matrix for linear equality constraints, $c(\boldsymbol{x})$ and $ceq(\boldsymbol{x})$ are functions that return vectors, and $f(\boldsymbol{x})$ is a function that returns a scalar value. The functions $f(\boldsymbol{x})$, $c(\boldsymbol{x})$, and $ceq(\boldsymbol{x})$ can be nonlinear.

4.4 Performance Evaluation of the ANS Algorithm

In order to test the performance of the vertical handoff decision algorithm, we shall continue to use the case study described in section 3.7.

4.4.1 Use Cases:

We shall use the two use cases presented in subsection 3.7.1.

4.4.2 Evaluation using Mathematical Optimization:

Case 1:

We first check to see whether a handoff should be initiated by calculating the handoff initiation factor.

Suppose that the MT records the data values of RSSI (dBm), Data Rate (Mbps), Network Coverage Area (Km), and Network Latency (ms) as {-66.2, 41.3, 36.1, 60.7}, {-67.3, 48.6, 42.3, 56.5} and {-67.01, 48.2, 42.5, 55.6} for WiMAX_1, WiMAX_2 and WiMAX_3 respectively.

As shown in section 3.7.2, the three sets of values were fed into the FIS and we obtained the Handoff Factor values 0.897, 0.902, and 0.903, respectively, thus indicating the need to hand off to any of the WMANs for the requested service.

The second stage of the VHDA is to compute the ANSF for all the available networks. The mobile terminal proceeds to gather data on all required parameters. The system converts the data into the corresponding membership function values which are used in computing the ANSF.

We use the notation

$$\{S, D, A, L, R, E, P, V, C\} = \{C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9\} \tag{4.20}$$

The matrix $\boldsymbol{B}$, the unit eigenvector $\boldsymbol{V}$, and weighting matrix $\boldsymbol{W}$ are indicated below:

$$\boldsymbol{B} = \begin{array}{c} \\ C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \\ C_9 \end{array} \begin{array}{c} \begin{array}{ccccccccc} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 & C_9 \end{array} \\ \begin{pmatrix} 1 & 1/8 & 7 & 7 & 7 & 7 & 7 & 7 & 1/9 \\ 8 & 1 & 8 & 8 & 8 & 9 & 8 & 8 & 1/9 \\ 1/7 & 1/8 & 1 & 4 & 1/5 & 1/6 & 4 & 4 & 1/9 \\ 1/7 & 1/8 & 1/4 & 1 & 1/5 & 1/6 & 2 & 1/3 & 1/9 \\ 1/7 & 1/8 & 5 & 5 & 1 & 1/6 & 5 & 5 & 1/9 \\ 1/7 & 1/8 & 6 & 6 & 6 & 1 & 6 & 6 & 1/9 \\ 1/7 & 1/8 & 1/4 & 1/2 & 1/5 & 1/6 & 1 & 1/3 & 1/9 \\ 1/7 & 1/8 & 1/4 & 3 & 1/5 & 1/6 & 3 & 1 & 1/7 \\ 9 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 1 \end{pmatrix} \end{array} \Rightarrow \boldsymbol{V} = \begin{pmatrix} -0.2393 \\ -0.4408 \\ -0.0514 \\ -0.0268 \\ -0.0812 \\ -0.1359 \\ -0.0237 \\ -0.0352 \\ -0.8474 \end{pmatrix} \Rightarrow \boldsymbol{W} = \begin{pmatrix} 0.1272 \\ 0.2343 \\ 0.0273 \\ 0.0142 \\ 0.0431 \\ 0.0722 \\ 0.0126 \\ 0.0187 \\ 0.4503 \end{pmatrix} \tag{4.21}$$

The attribute weights and the membership values (lower bound, upper bound) obtained from the characteristics of the four available networks for the attributes are summarised in Table 4.2.

Table 4.2: *Parameters for Performance Evaluation of Case 1*

Criteria		w_j	Membership Values (lb, ub)			
			UMTS_2	WiMAX_1	WiMAX_2	WiMAX_3
RSSI	C_1	0.1272	0.5, 0.9	0.5, 0.9	0.5, 0.9	0.5, 0.9
Data Rate	C_2	0.2343	0.05, 0.1	0.2, 0.94	0.2, 0.9	0.2, 0.9
Network Coverage Area	C_3	0.0273	0.1, 0.45	0.2, 0.85	0.2, 0.95	0.2, 0.92
Network Latency	C_4	0.0142	0.4, 0.6	0.5, 0.88	0.5, 0.9	0.5, 0.9
Reliability	C_5	0.0431	0.7, 0.9	0.7, 0.89	0.7, 0.9	0.7, 0.9
Security	C_6	0.0722	0.8, 0.9	0.8, 0.9	0.8, 0.9	0.8, 0.9
Power Requirement	C_7	0.0126	0.7, 0.8	0.6, 0.75	0.6, 0.75	0.6, 0.75
Mobile Velocity	C_8	0.0187	0.01, 0.9	0.005, 0.1	0.01, 0.5	0.01, 0.5
Service Cost	C_9	0.4503	0.5, 0.6	0.6, 0.81	0.6, 0.85	0.6, 0.87

We performed mathematical optimization experiments by using the MATLAB Optimization Toolbox. The solutions obtained are summarized in Table 4.3.

Based on the results of the optimal values of the ANSF for the access networks in Table 4.3, the WiMAX_3 provides the optimal positive result and it will be suitable to handoff from the UMTS_2 to the WiMAX_3 to complete downloading the large file.

Table 4.3: *The Mathematical Optimization Values for Case 1*

Criteria		w_j	Optimal Membership Values			
			UMTS_2	WiMAX_1	WiMAX_2	WiMAX_3
RSSI	C_1	0.1272	0.9000	0.9000	0.9000	0.9000
Data Rate	C_2	0.2343	0.1000	0.9400	0.9000	0.9000
Network Coverage Area	C_3	0.0273	0.4500	0.8500	0.9500	0.9200
Network Latency	C_4	0.0142	0.6000	0.8800	0.9000	0.9000
Reliability	C_5	0.0431	0.9000	0.8900	0.9000	0.9000
Security	C_6	0.0722	0.9000	0.9000	0.9000	0.9000
Power Requirement	C_7	0.0126	0.8000	0.7500	0.7500	0.7500
Mobile Velocity	C_8	0.0187	0.9000	0.1000	0.5000	0.5000
Service Cost	C_9	0.4503	0.6000	0.8100	0.8500	0.8700
Optimal ANSF Value			-0.5596	-0.8498	-0.8694	-0.8776

Case 2:

We first check to see whether a handoff should be initiated by calculating the handoff initiation factor.

Suppose that the MT records the data values of RSSI (dBm), Data Rate (Mbps), Network Coverage Area (Km), and Network Latency (ms) as {-69.4, 40.5, 35.7, 70.8}, {-68.6, 42.1, 38.6, 65.4} and {-68.1, 43.2, 40.3, 62.8} for WiMAX_1, WiMAX_2 and WiMAX_3 respectively.

As shown in section 3.7.2, these set of values were fed into the FIS and we obtained the Handoff Factor values 0.886, 0.892 and 0.895, respectively, and thus indicating the need to handoff to any of the three WiMAX networks for the requested service.

The second stage of the VHDA is to compute the ANSF for all the available networks. The mobile terminal proceeds to gather data on all required parameters.

The mobile terminal proceeds to gather data on all required parameters. The matrix ***B*** and weighting matrix ***W*** are indicated below:

$$
\boldsymbol{B} = \begin{array}{c} \\ C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \\ C_9 \end{array}
\begin{array}{c}
\begin{array}{ccccccccc} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 & C_9 \end{array} \\
\begin{pmatrix}
1 & 1/9 & 6 & 1/7 & 6 & 6 & 6 & 6 & 1/8 \\
9 & 1 & 9 & 9 & 9 & 9 & 9 & 9 & 9 \\
1/6 & 1/9 & 1 & 1/7 & 3 & 1/5 & 3 & 1/4 & 1/8 \\
7 & 1/9 & 7 & 1 & 7 & 7 & 7 & 7 & 1/8 \\
1/6 & 1/9 & 1/3 & 1/7 & 1 & 1/5 & 1/2 & 1/4 & 1/8 \\
1/6 & 1/9 & 5 & 1/7 & 5 & 1 & 5 & 5 & 1/8 \\
1/6 & 1/9 & 1/3 & 1/7 & 2 & 1/5 & 1 & 1/4 & 1/8 \\
1/6 & 1/9 & 4 & 1/7 & 4 & 1/5 & 4 & 1 & 1/8 \\
8 & 1/9 & 8 & 8 & 8 & 8 & 8 & 8 & 1
\end{pmatrix}
\end{array}
\Rightarrow \boldsymbol{V} = \begin{pmatrix} -0.1359 \\ -0.8474 \\ -0.0352 \\ -0.2393 \\ -0.0237 \\ -0.0812 \\ -0.0268 \\ -0.0514 \\ -0.4408 \end{pmatrix}
\Rightarrow \boldsymbol{W} = \begin{pmatrix} 0.0722 \\ 0.4503 \\ 0.0187 \\ 0.1272 \\ 0.0126 \\ 0.0431 \\ 0.0142 \\ 0.0273 \\ 0.4503 \end{pmatrix} \qquad (4.22)
$$

The attribute weights and the membership values (lower bound, upper bound) obtained from the characteristics of the four available networks for the attributes are summarised in Table 4.4.

Table 4.4: *Parameters for Performance Evaluation for Case 2*

Criteria		w_j	Membership Values (lb, ub)			
			UMTS_2	WiMAX_1	WiMAX_2	WiMAX_3
RSSI	C_1	0.0722	0.5, 0.9	0.5, 0.9	0.5, 0.9	0.5, 0.9
Data Rate	C_2	0.4503	0.05, 0.1	0.2, 0.94	0.2, 0.9	0.2, 0.9
Network Coverage Area	C_3	0.0187	0.1, 0.45	0.2, 0.85	0.2, 0.95	0.2, 0.92
Network Latency	C_4	0.1272	0.4, 0.6	0.5, 0.88	0.5, 0.9	0.5, 0.9
Reliability	C_5	0.0126	0.7, 0.9	0.7, 0.89	0.7, 0.9	0.7, 0.9
Security	C_6	0.0431	0.8, 0.9	0.8, 0.9	0.8, 0.9	0.8, 0.9
Power Requirement	C_7	0.0142	0.7, 0.8	0.6, 0.75	0.6, 0.75	0.6, 0.75
Mobile Velocity	C_8	0.0273	0.01, 0.9	0.005, 0.1	0.01, 0.5	0.01, 0.5
Service Cost	C_9	0.2343	0.5, 0.6	0.6, 0.81	0.6, 0.85	0.6, 0.87

The solutions obtained using the MATLAB Optimization Toolbox are summarized in Table 4.5. From the results of the optimal values of the ANSF for the access networks in Table 4.5, the WiMAX_3 provides the optimal positive result and it will be suitable to handoff from the WLAN to the WiMAX_3 to continue with the video call.

Table 4.5: *The Mathematical Optimization Values for Case 2*

Criteria		w_j	Optimal Membership Values			
			UMTS_2	WiMAX_1	WiMAX_2	WiMAX_3
RSSI	C_1	0.0722	0.9000	0.9000	0.9000	0.9000
Data Rate	C_2	0.4503	0.1000	0.9400	0.9000	0.9000
Network Coverage Area	C_3	0.0187	0.4500	0.8500	0.9500	0.9200
Network Latency	C_4	0.1272	0.6000	0.8800	0.9000	0.9000
Reliability	C_5	0.0126	0.9000	0.8900	0.9000	0.9000
Security	C_6	0.0431	0.9000	0.9000	0.9000	0.9000
Power Requirement	C_7	0.0142	0.8000	0.7500	0.7500	0.7500
Mobile Velocity	C_8	0.0273	0.9000	0.1000	0.5000	0.5000
Service Cost	C_9	0.2343	0.6000	0.8100	0.8500	0.8700
Optimal ANSF Value			-0.5510	-1.0442	-1.0597	-1.0681

We observe that the Mathematical Optimization algorithm always picks up the upper bound values of the membership values as the solution values for the attributes.

4.5 Summary

The handoff decision problem is a typical multiple attribute decision making (MADM) problem since it involves selecting a suitable access network among a number of candidate access networks with respect to several criteria (attributes). In this chapter the access network selection problem was posed as a mathematical optimization problem using an objective function that provides the means to quantify the weights or relative merits of satisfying the different parameters including the network, user and QoS requirements.

We observe from the results in Table 4.3 and Table 4.5 that Mathematical Optimization always selects the upper bounds of a solution vector $\boldsymbol{x}$ for calculating the optimum value of an objective function. Consequently, Mathematical Optimization is not suitable for solving problems of the type posed in the Use Cases 1 and 2 of section 4.4.

CHAPTER 5

OPTIMIZATION OF ACCESS NETWORK SELECTION USING A GENETIC ALGORITHM

This chapter aims at optimizing the access network selection function introduced in the last chapter by using a natural optimization method, the genetic algorithm, which generates new points in the search space by applying operators to current points and statistically moving toward more optimal places in the search space. Before doing that, it is important to define what is meant by optimization in general, why it makes sense to differentiate between discrete, continuous and mixed-variable optimization, and why genetic algorithms may be used for tackling these types of problems.

5.1 Introduction

In this chapter we define what is meant by optimization in general, why it makes sense to differentiate between discrete, continuous and mixed-variable optimization. Then we present a natural optimization method, the genetic algorithm (GA), which generates new points in the search space by applying operators to current points and statistically moving toward more optimal places in the search space. The GA models natural selection and evolution and relies upon an intelligent search of a large but finite solution space using statistical methods. It does not require taking objective function derivatives and can therefore deal with discrete parameters and noncontinuous objective functions. The GA has been formalized into a meta-heuristic for combinatorial optimization problems. The goal of heuristic methods is to quickly produce good approximate solutions, without necessarily providing any guarantee of solution quality. Typically, the core of heuristic methods is an iterative principle that includes stochastic elements in generating new candidate solutions and/or in deciding whether these replace their predecessors – while still incorporating some mechanism that prefers and encourages improvements. They have several advantages over traditional deterministic methods:

- The randomness allows escaping local optima;
- They are typically for a general purpose and can therefore deal with a vast array of different functional forms and constraints; and
- They are reasonably efficient compared to other widely used methods when traditional optimization fails.

The access network selection problem, described in Chapter 4, is not solvable realistically using mathematical optimization algorithms. Therefore, we explore in this chapter the use of the GA for solving the optimization problem of maximizing the access network selection function of Chapter 4.

5.2 Optimization Problems and Algorithms

5.2.1 Optimization Problems

Optimization is the process of finding the best of all possible solutions to a problem. It refers to the study of problems in which one seeks to find the minimum or maximum value of an objective function, or cost function, by adjusting the variables or parameters from within an allowed set.

An optimization problem begins with a set of independent variables (parameters) and a scalar measure of "goodness" termed the objective function which depends in some way on the variables, and often includes conditions or restrictions (termed the constraints of the problem) that define acceptable values of the variables [78]. The solution of an optimization problem is the set of allowed values of the variables for which the objective function assumes an "optimal" value.

A single objective optimization problem can be defined as follows [79]:

Definition 5.1: Given a function $f : \boldsymbol{S} \rightarrow \mathbf{R}$, find $\mathbf{x}^* \in \boldsymbol{S} \subset \mathbf{R}^n$: $\forall\, \mathbf{x}^* \in \boldsymbol{S}$ $f(\mathbf{x}^*) \leq f(\mathbf{x})$ (minimization) or $f(\mathbf{x}^*) \geq f(\mathbf{x})$ (maximization) subject to a set Ω of constraints $b_1, \ldots, b_m$ defined by $f_i : \boldsymbol{S} \rightarrow \mathbf{R}$ and $f_i \leq b_i$, for $i = 1, 2, \ldots, m$.

The n-dimensional real-valued function f is called an *objective function*, *energy function*, or *cost function*. Its domain $\boldsymbol{S}$ is called the *search* (or *solution*) *space*, and the elements of $\boldsymbol{S}$ are called *feasible solutions*. A feasible solution $\mathbf{x} \in \boldsymbol{S}$ is a vector of optimization variables $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ that satisfy all the constraints. If the set Ω of constraints is empty, the problem is called an unconstrained problem, otherwise it is said to be constrained.

The generic variable x_i takes values in the domain $D_i = \{v_i^1, v_i^2, \ldots, v_i^m$, where $m = |D_i|\}$.

In global optimization, it is a convention that optimization problems are most often defined as minimizations and if an objective function f is subject to maximization, we simply minimize its negation $-f$.

The optimization algorithm searches for a solution in a *search space*, $\boldsymbol{S}$, of candidate solutions. In the case of constrained problems, a solution is found in the feasible space, $\boldsymbol{F} \subseteq \boldsymbol{S}$.

Definition 5.2: The solution, $\mathbf{x}^* \in \boldsymbol{N} \subseteq \boldsymbol{F}$, is a strong local minimum of f if

$$f(\mathbf{x}^*) < f(\mathbf{x}),\ \forall\, \mathbf{x} \in \boldsymbol{N} \qquad (5.1)$$

where $\boldsymbol{N} \subseteq \boldsymbol{F}$ is a set of feasible points in the neighbourhood of $\mathbf{x}^*$.

Definition 5.3: The solution, $\mathbf{x}^* \in \boldsymbol{F}$, is a strict global optimum of the objective function, f, if

$$f(\mathbf{x}^*) < f(\mathbf{x}),\ \forall\, \mathbf{x} \in \boldsymbol{F} \qquad (5.2)$$

where $\boldsymbol{F} \subseteq \boldsymbol{S}$. That is, the global optimum is the best possible solution to a problem.

The ultimate goal of an optimization method is to find the globally optimal solution $\mathbf{x}^*$ and the corresponding optimal value of the objective function, $f^* = f(\mathbf{x}^*)$. Although there can be an infinite number of strong local minima, there is at most one strict global minimum.

Optimization problems can be classified into various families based on a number of characteristics such as the type of variables, the number of variables, the number of optimization criteria. In this thesis, we use the classification based on the type of variables, that is, on how the optimization problems differ in the definition of their search space. Optimization problems can be classified into three types based on the type of variables:

- *Discrete optimization problems* – problems in which all the optimization variables x_i, $i = 1, 2, \ldots, n$ are discrete or integers, that is, x_i belong to a countable set D_i, $i = 1, 2, \ldots, n$.
- *Continuous optimization problems* – problems in which all the optimization variables are continuous, that is, $x_i \in \Re$ for each $i = 1, 2, \ldots, n$.
- *Mixed-variable optimization problems* – problems that have both discrete-valued and continuous-valued variables, that is, p out of $n = p + q$ variables are discrete, $x_i \in D_i$ for $i = 1, \ldots, p$, and q are continuous $x_i \in \boldsymbol{R}$ for each $i = p + 1, \ldots, p + q$.

Solving these three different classes of problems poses different difficulties and often requires different methods. While some methods work well on discrete optimization problems, they may not be used for continuous optimization problems, and vice versa. Discrete variables have only a finite number of possible values, whereas continuous variables have an infinite number of possible values. If we are deciding in what order to attack a series of tasks on a list, discrete optimization is employed. However, if we are trying to find the optimum value of $f(x)$ on a number line, it is more appropriate to view the problem as continuous.

A *combinatorial optimization problem* has its solution as a permutation or a combination of discrete-valued variables. It is therefore characterized by a finite set of possible solutions. In the thesis we shall focus on combinatorial optimization problems since there are many problems of this type, and several algorithms have been proposed to deal exclusively with such combinatorial problems.

Definition 5.4: A model $P = (\boldsymbol{S}, \Omega, f)$ of a combinatorial optimization problem consists of:

- a *search* (or *solution*) *space*, $\boldsymbol{S}$, of discrete decision variables x_i, $i = 1, 2, \ldots, n$;
- a set Ω of constraints among the variables; and
- an objective function f to be minimized, where $f : \boldsymbol{S} \rightarrow \boldsymbol{R}$.

The generic variable x_i takes possible values v_i^j in $\boldsymbol{D}_i = \{v_i^1, v_i^2, \ldots, v_i^m$, where $m = |\boldsymbol{D}_i|\}$. A feasible solution $\boldsymbol{x} \in \boldsymbol{S}$ is a complete assignment of values to variables that satisfy all the constraints in the set Ω.

Definition 5.5: A model $P = (\boldsymbol{S}, \Omega, f)$ of a continuous optimization problem consists of:

- a *search* (or *solution*) *space*, $\boldsymbol{S}$, of a finite set of continuous decision variables x_i, $i = 1, 2, \ldots, n$;
- a set Ω of constraints among the variables; and
- an objective function f to be minimized, where $f : \boldsymbol{S} \rightarrow \boldsymbol{R}$.

The generic continuous variable x_i takes possible values $v_i \in \boldsymbol{D}_i \subseteq \boldsymbol{R}$. The search space of a continuous optimization problem is not finite as each of the continuous decision variables may assume an infinite number of values.

5.2.2 Optimization Algorithms

Optimization algorithms are search methods, where the goal is to find a solution to an optimization problem, such that a given quantity is optimized, possibly subject to a set of constraints.

Two classes of algorithms are available for the solution of combinatorial optimization problems: *exact* and *approximate algorithms*.

Exact algorithms are guaranteed to find the optimal solution and to prove its optimality for every finite size instance of a combinatorial optimization problem within an instance-dependent run time.

Approximate methods are not guaranteed to find an optimal solution but, in practice, they are often able to find good solutions, possibly near-optimal, in a relatively short time.

Heuristic methods are approximate methods that produce solutions close to the optimum but the majority of them were conceived specifically for a given problem.

A new kind of approximate algorithm, called a *metaheuristic*, has emerged which basically tries to combine basic heuristic methods in higher level frameworks aimed at efficiently and effectively exploring a search space [80]. A metaheuristic is a set of concepts that can be used to define heuristic methods applicable to a wide set of different problems. In other words, a metaheuristic can be seen as a general algorithmic framework which can be applied to different optimization problems with relatively few modifications to make them adapted to a specific problem [81]. Examples of metaheuristics are evolutionary algorithms [82] (including the GA and genetic programming [83]), simulated annealing (SA), tabu search [84, 85], particle swarm optimization [86, 87], and ant colony optimization [88].

The algorithms available for the solution of continuous optimization problems may also be divided into the *exact* and *approximate algorithms*.

The exact algorithms for the solution of continuous optimization problems include analytical approach, numerical methods based on gradient descent, and direct search methods.

The approximate algorithms for the solution of continuous optimization problems are metaheuristics. While some metaheuristics were developed with the continuous optimization in mind, most of them have been adapted to continuous optimization based on their counterparts initially developed for combinatorial optimization. Metaheuristics for continuous optimization are often hybridized with local search methods to allow them to focus on global optimization, while the local search methods, such as direct search or gradient-based methods, help them in finding local optimums.

In this thesis, we solve the access network selection problem which is characterized by a search space that increases exponentially with the input size. Therefore, an optimal solution to this problem cannot be found in polynomial time using simple deterministic algorithms. Stochastic heuristics such as the GA and SA offer more appropriate alternatives. These methods, which aim at finding near-optimal solutions in polynomial time, have proved to be very successful when applied to problems in various fields such

as industrial management, financial services, graph theory, and Very-Large-Scale-Integration. Their application to the field of mobile communications is still in its infancy.
Natural optimization algorithms, including the GA and the SA generate new points in the search space by applying operators to current points and statistically moving towards more optimal places in the search space. They rely upon an intelligent search of a large but finite solution space using statistical methods. They represent processes in nature that are remarkably successful at optimizing natural phenomena.

5.3 The Genetic Algorithm

5.3.1 Introduction to Genetic Algorithm

The Genetic Algorithm (GA) is a search method for solving both constrained and unconstrained optimization problems that is based on the mechanics of natural selection and genetics [89, 90, 91]. It is applicable to a variety of optimization problems that are not well suited for standard optimization algorithms, including problems in which the objective function is discontinuous, non-differentiable, stochastic, or highly nonlinear. Genetic Algorithms manipulate a population of potential solutions to an optimization (or search) problem. Specifically, they operate on encoded representations of the solutions and not directly on the solutions themselves. The GA encodes the decision variables of a search problem into finite-length strings of alphabets of certain cardinality [92]. The strings which are candidate solutions to the search problem are referred to as chromosomes, the alphabets are referred to as genes and the values of genes are referred to as alleles. In contrast to traditional optimization techniques, the GA works with coding of parameters rather than the parameters themselves.
Each solution is associated with a fitness measure that reflects how good it is, compared with other solutions in the population. The measure could be an objective function (either a single objective function in single-objective optimization or a multi-objective function in multi-objective optimization) that is a mathematical model or a computer simulation, or it can be a subjective function where people intuitively choose better solutions over worse ones.

Unlike traditional search methods, the GA relies on a population of candidate solutions. The population size, which is usually a user-specified parameter, is an important factor that affects the scalability and performance of genetic algorithms [90]. Small population sizes may lead to premature convergence and yield substandard solutions, while large population sizes lead to unnecessary expenditure of valuable computational time.

The GA begins with and repeatedly modifies a randomly generated population of individual solutions modeled on gene combinations in biological reproduction and not a single trajectory evolving from a unique initial solution. At each step, the GA selects individuals at random according to their fitness from the current population to be parents and uses them to produce the children for the next generation. The algorithm repeatedly applies selection, crossover, mutation and replacement operators at each step to create the next generation from the current population until a satisfactory solution is found or a maximum number of iterations is reached. The population evolves toward an optimal solution over successive generations.

In the following, we assume a function minimization problem. Hence, a good solution is one that has low relative fitness.

5.3.2 The Genetic Algorithm Scheme

The basic steps in a GA system are shown in Table 5.1. Having encoded a problem in a chromosomal manner and having devised a means of discriminating good solutions from bad ones, we can use a GA to evolve solutions to the problem by the following steps [89, 90, 92]:

Step 1: *Initialization.* The algorithm begins by creating a random initial population of individuals.

Step 2: The algorithm then creates a sequence of new populations. At each step, the algorithm uses the individuals in the current generation to create the next population. To create the new population, the algorithm iteratively performs the following steps:

a) *Evaluation.* The objective function values of the candidate solutions in the current population are computed.

b) *Fitness Assignment*. The algorithm uses the objective function values to determine fitness values of the candidate solutions.

c) *Selection*. The algorithm selects members, called parents, based on their fitness. The main idea of selection is to prefer better solutions to worse ones, and many selection procedures have been proposed to accomplish this including roulette-wheel selection, ranking selection, stochastic universal selection, and tournament selection.

d) *Elitism.* Some of the individuals in the current population that have lower fitness are chosen as *elite* individuals and are passed to the next population as children.

e) *Crossover (Recombination).* Crossover combines the vector entries or genes of two parents to form potentially better solutions (offspring) for the next generation. The crossover is controlled by the crossover probability p_c which is typically in the range [0.7 – 0.95]. That is, a uniform random number, r, is generated and if $r \leq p_c$, the two randomly selected parents undergo recombination. Common crossover operators are *k-point crossover* and *uniform crossover*.

f) *Mutation.* Mutation applies random changes to one or more genes of an individual parent to form children. Mutation adds to the diversity of a population. It is performed with a low probability p_m typically in the range [0.01 – 0.2].

g) *Replacement.* The current population is replaced with the children created by selection, crossover, and mutation to form the next generation.

Step 3: The algorithm stops when one of the stopping criteria is met.

Table 5.1: Genetic Algorithm

Let $t = 0$ be the generation counter;
Create and initialize an n_x-dimensional population: $P(0)$;
repeat
  Compute the objective function value $f(\mathbf{x}_i)$ of each individual $\mathbf{x}_i$ in the population $P(t)$;
  Determine the fitness values by using the objective function values $f(\mathbf{x}_i)$;
  Select parents in $P(t)$ based on their fitness;
  Perform crossover to produce offspring;
  Perform mutation on offspring;
  Replace the current population with offspring to form the new generation $P(t + 1)$;
  Advance to the new generation, i.e., $t = t + 1$;
until *stopping condition is true*;

5.3.3 Main Features of Genetic Algorithms

A GA for a particular problem must have the following five components:

- A genetic representation for potential solutions to the problem,
- A way to create an initial population of potential solutions,
- An evaluation function rating solutions in terms of their fitness,
- Genetic operators that alter the composition of children, and
- Values for various parameters that the GA uses (population size, probabilities of applying genetic operators, etc.).

The main features of the GAs are discussed below.

Representation

Fundamental to the GA structure is the encoding mechanism to represent the decision variables of the optimization problem into finite-length strings of alphabets of certain cardinality. The strings which are candidate solutions to the search problem are referred to as chromosomes, the alphabets are referred to as genes and the values of genes are referred to as alleles. The encoding scheme depends on the nature of the problem variables. The classical representation scheme for GAs is binary vectors of fixed length for binary-valued, nominal-valued, and continuous-valued variables. GAs have also been developed that use integer or real-valued representations and order-based representations where the order of variables in a chromosome plays an important role.

Initial Population

There are several strategies to generate the initial population. The standard way of generating an initial population is to assign a random value from the allowed domain to each of the genes of each chromosome. The goal of random selection is to ensure that the initial population is a uniform representation of the entire search space.

The size of the initial population has consequences for scalability and performance in terms of computational complexity and exploration abilities. A large initial population of individuals increases the diversity, thereby improving the exploration abilities of the population. However, there is a high computational complexity per generation for a large population. Smaller population sizes may lead to premature convergence and yield substandard solutions.

Selection Methods

Selection models nature's survival-of-the-fittest mechanism by allocating more copies of those solutions with higher fitness values. The main idea of selection is to prefer better solutions to worse ones, and many selection procedures have been proposed to accomplish this including roulette-wheel selection, ranking selection, stochastic universal selection, and tournament selection. Selection operators are characterized by their *selection pressure*, which is defined as the speed at which the best solution will occupy the entire population by repeated application of the selection operator alone. An operator with a high selection pressure decreases diversity in the population more rapidly than operators with a low selection pressure. The most frequently used selection methods [89, 93] are fitness proportionate selection methods such as roulette-wheel selection, ordinal selection methods such as tournament selection and truncation selection, rank-based selection, stochastic remainder technique, and elitism.

Crossover Operators

Crossover (or recombination) is a genetic operator that combines the traits of two or more parental solutions to create new, hopefully better offspring. The offspring under crossover will not be identical to any particular parent and will instead combine parental traits in a novel manner [90]. The two main attributes of crossover that can be varied are the type of crossover that is implemented and the probability of its occurrence. In most crossover operators, two individuals are randomly selected and are recombined with a crossover probability p_c which is typically in the range [0.7 – 0.95]. That is, a uniform random number, r, is generated and if $r \leq p_c$, the two randomly selected parents undergo recombination. Common crossover operators are *k-point crossover* and *uniform crossover* [94].

Mutation Operators

Mutation is the process of randomly changing the values of genes in a chromosome. Mutation stochastically flips bits in each generation. It applies random changes to one or more genes of an individual parent to form children. The aim of mutation is to introduce new genetic material into an existing individual, that is, to add to the diversity of the genetic material of the population. It thereby enlarges the search space and keeps the GA from converging too fast. The mutation rate greatly affects the performance of the GA. Too much mutation badly affects the results. It is performed with a low probability p_m typically in the range [0.01 – 0.2].

Replacement

The current population is replaced with the children created by selection, crossover, and mutation to form the next generation. Some of the common replacement techniques are [92]:

- Delete-all. This technique deletes all the members of the current population and replaces them with the same number of chromosomes that have just been created.
- Steady-state. This technique deletes *n* old members and replaces them with *n* new members.
- Steady-state-no-duplicates. This is the same as the steady-state technique but the algorithm checks that no duplicate chromosomes are added to the population.

5.4 Genetic Algorithm Optimization of the Access Network Selection Function

The access network selection problem, described in Chapter 4, is not solvable realistically using simple deterministic algorithms.

In this section we explore the use of GAs for solving the optimization problem of maximizing the ANSF in equation (4.16) of Chapter 4.

The major steps for using the GA as recalled from section 5.3.2 are as follows:

(1) *Initialization*: generate a random initial population of candidate solutions.

(2) *Evaluation*: the objective function values of the candidate solutions in the current population are evaluated.

(3) *Fitness Assignment*: use the objective function values to determine the fitness values of the candidate solutions in the current population.

(4) *Selection*: select members, called parents, based on their fitness.

(5) *Reproduction*: reproduce the children created by *selection*, *crossover*, and *mutation* to form the next generation.

(6) Go to step 2, until *stopping criteria* are met.

To tackle the optimization problem of maximizing the ANSF in equation (4.16) under the weights w_j by using a GA, we assume a function minimization problem. Hence, a good solution is one that has low relative fitness. Since our GA algorithm performs minimization of an objective function $f(\boldsymbol{x})$, maximization of the objective function in equation (4.16) is achieved by supplying the routine with minus $f_i(\boldsymbol{x})$ because the point at which the minimum of $-f_i(\boldsymbol{x})$ occurs is the same as the point at which the maximum of $f_i(\boldsymbol{x})$ occurs.

Therefore, we define the equivalent minimization problem:

$$\min\ (-f_i(\boldsymbol{x})) = \min\{-\sum_{j=1}^{n} w_j \cdot \mu_{Cj}(A_i)\} \tag{5.3}$$

such that

$$0 \le w_j \le 1, \text{ and } \sum_{j=1}^{n} w_j = 1, \tag{5.4}$$

and

$$\{\mu_{Cj}(A_i)\}_{\min} \le \mu_{Cj}(A_i) \le \{\mu_{Cj}(A_i)\}_{\max}. \tag{5.5}$$

We add the linear constraint:

$$\sum_{i=1}^{9} \mu_{Cj}(A_i) \le 9. \tag{5.6}$$

In order to achieve good results in using a GA, one may have to experiment with different GA operators (selection, crossover, and mutation operators) and different GA parameters (population size, crossover probability, and mutation probability). In [95], the authors made comparisons of the GA performance as a function of the different GA operators and the different GA parameters and concluded that the population size and the mutation rate have the most impact on the ability of the GA to find a better minimum value for the objective function.

5.5 Performance Evaluation of the ANS Algorithm with Genetic Algorithm

The performance of the vertical handoff decision algorithm is tested within the framework of a scenario that simulates a typical day in the life of a telecommunication field technician, Mr. Alex Star as presented in section 3.7.

5.5.1 Use Cases:

We shall continue with the use *Case 1* and *Case 2* initially presented in section 3.7.1 for the performance evaluation.

5.5.2 Evaluation using Genetic Algorithm:

Case 1:

We first check to see whether a handoff should be initiated by calculating the handoff initiation factor.

As shown in section 3.7.2, the three sets of values were fed into the FIS and we obtained the Handoff Factor values 0.897, 0.902, and 0.903 for WiMAX_1, WiMAX_2, and WiMAX_3 respectively, thus indicating the need to hand off to any of the WMANs for the requested service.

The second stage of the VHDA is to compute the ANSF for all the available networks. The mobile terminal proceeds to gather data on all required parameters. The matrix ***B*** and weighting matrix ***W*** were calculated and indicated in subsection 4.4.2 as equation 4.21.

The attribute weights and the membership values (lower bound, upper bound) obtained from the characteristics of the four available networks for the attributes were summarised in Table 4.2 of Chapter 4 and reproduced below for convenience.

Table 4.2: *Parameters for Performance Evaluation of Case 1*

Criteria		w_j	Membership Values (lb, ub)			
			UMTS_2	WiMAX_1	WiMAX_2	WiMAX_3
RSSI	C_1	0.1272	0.5, 0.9	0.5, 0.9	0.5, 0.9	0.5, 0.9
Data Rate	C_2	0.2343	0.05, 0.1	0.2, 0.94	0.2, 0.9	0.2, 0.9
Network Coverage Area	C_3	0.0273	0.1, 0.45	0.2, 0.85	0.2, 0.95	0.2, 0.92
Network Latency	C_4	0.0142	0.4, 0.6	0.5, 0.88	0.5, 0.9	0.5, 0.9
Reliability	C_5	0.0431	0.7, 0.9	0.7, 0.89	0.7, 0.9	0.7, 0.9
Security	C_6	0.0722	0.8, 0.9	0.8, 0.9	0.8, 0.9	0.8, 0.9
Power Requirement	C_7	0.0126	0.7, 0.8	0.6, 0.75	0.6, 0.75	0.6, 0.75
Mobile Velocity	C_8	0.0187	0.01, 0.9	0.005, 0.1	0.01, 0.5	0.01, 0.5
Service Cost	C_9	0.4503	0.5, 0.6	0.6, 0.81	0.6, 0.85	0.6, 0.87

The GA-related results: We performed the GA optimization experiments by using the stochastic universal selection rule that lays out a line in which each parent corresponds to a section of the line of length proportional to its scaled value, and used the control parameter options: population size p_s = 20, elite count = 2, single-point crossover with p_c = 0.8, and mutation probability p_m = 0.01. These parameters are listed in Table 5.2.

Table 5.2: *Parameters Used for the GA*

	Case 1	**Case 2**
Population size	20	20
Elite count	2	2
Crossover type	Single point	Single point
Crossover probability	0.8	0.8
Mutation probability	0.01	0.01
Selection method	Stochastic universal	Stochastic universal

The initial population starts off with a set of feasible solutions and infeasible solutions are dropped during reproduction.

The solutions (a list of the optimum ANSF values and optimum membership function values) obtained using the GA with the MATLAB GA Toolbox are summarised in Table 5.3.

Table 5.3: *The GA Optimization Values for Case 1*

Criteria		w_j	Optimal Membership Values			
			UMTS_2	WiMAX_1	WiMAX_2	WiMAX_3
RSSI	C_1	0.1272	0.9000	0.9000	0.9000	0.8831
Data Rate	C_2	0.2343	0.0919	0.9400	0.9000	0.9000
Network Coverage Area	C_3	0.0273	0.4486	0.8469	0.9496	0.9050
Network Latency	C_4	0.0142	0.5766	0.8418	0.6660	0.8550
Reliability	C_5	0.0431	0.8911	0.8900	0.8344	0.8902
Security	C_6	0.0722	0.8968	0.8938	0.9000	0.8920
Power Requirement	C_7	0.0126	0.8000	0.6810	0.7500	0.7504
Mobile Velocity	C_8	0.0187	0.6019	0.1000	0.5000	0.4980
Service Cost	C_9	0.4503	0.6000	0.8080	0.8500	0.8700
Optimal ANSF Value			-0.5511	-0.8470	-0.8632	-0.8733

Based on the results of the optimal values of the ANSF for the access networks in Table 5.3, the WiMAX_3 provides the optimal positive result. Therefore, it is more preferable to handoff from the UMTS_2 to the WiMAX_3 to complete downloading the large file.

In order to compare the performances of the GA and the SA used in our experiments, the best fitness value and the best individual given by the GA for the WiMAX_3 network are plotted in Figure 5.1. The upper plot displays the corresponding changes in the best and mean fitness values in each generation. The lower plot displays the coordinates of the point with the best fitness value in the current generation.

The GA converges quickly to the best solution for the control parameter settings indicated above. The GA terminates if the maximum number of the GA generations reaches 63 generations. After 63 generations, the objective function value of the best solution is -0.87327, which is the same value as the Optimal ANSF Value for WiMAX_3 indicated in Table 5.3.

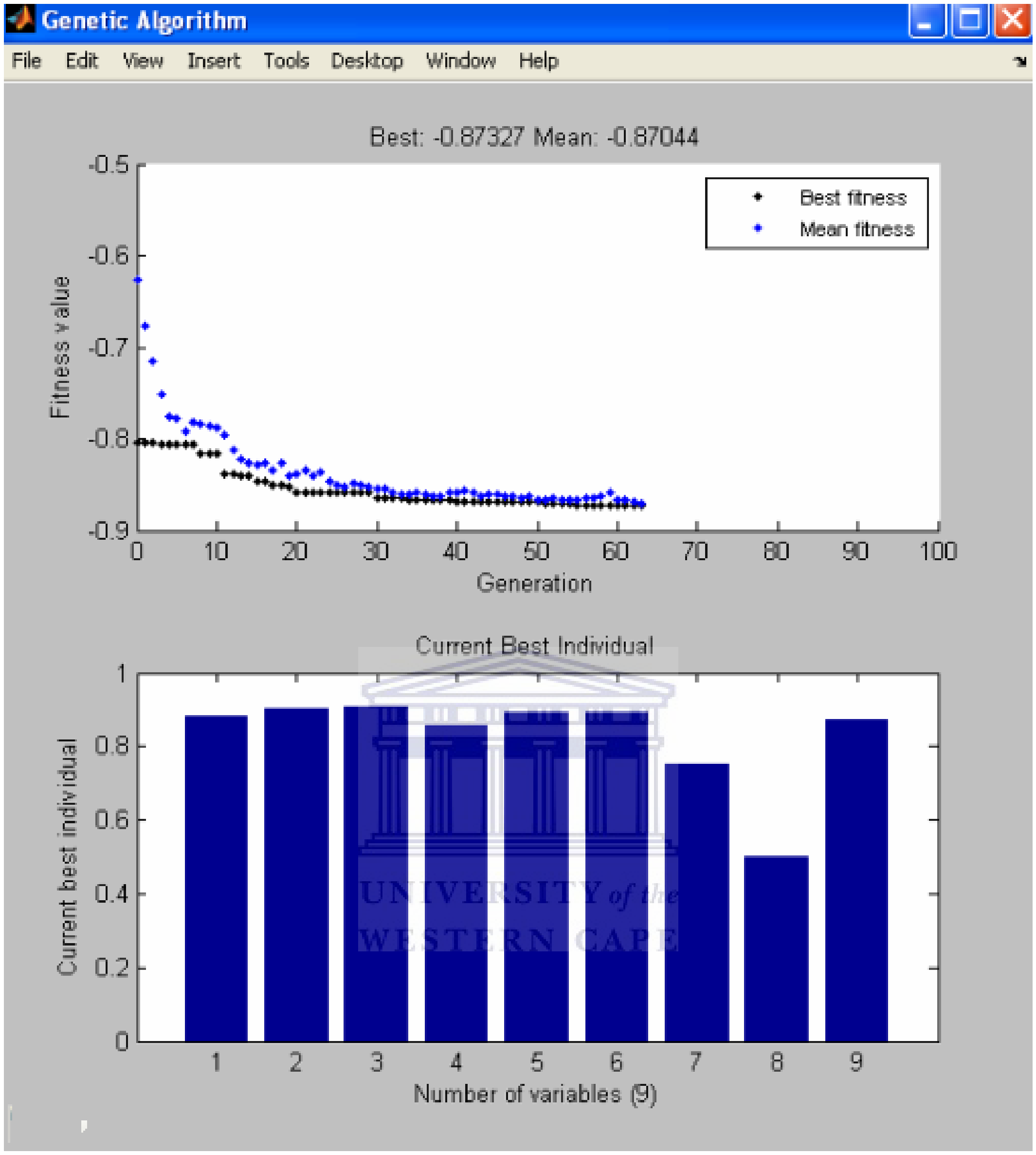


Figure 5.1: *The Case 1 GA Best Fitness Value and Best Individual for WiMAX_3*

Case 2:

We first check to see whether a handoff should be initiated by calculating the handoff initiation factor.

As shown in Chapter 3, we obtained the Handoff Factor values 0.886, 0.892 and 0.895 for WiMAX_1, WiMAX_2, and WiMAX_3, respectively, and thus indicating the need to handoff to any of the three WiMAX networks for the requested service.

The second stage of the VHDA is to compute the ANSF for all the available networks. The mobile terminal proceeds to gather data on all required parameters. The matrix ***B*** and weighting matrix ***W*** were calculated and indicated in subsection 4.4.2 as equation 4.22.

The attribute weights and the membership values (lower bound, upper bound) obtained from the characteristics of the four available networks for the attributes were summarised in Table 4.4 of Chapter 4 and reproduced below for convenience.

Table 4.4: *Parameters for Performance Evaluation for Case 2*

Criteria		w_j	Membership Values (lb, ub)			
			UMTS_2	WiMAX_1	WiMAX_2	WiMAX_3
RSSI	C_1	0.0722	0.5, 0.9	0.5, 0.9	0.5, 0.9	0.5, 0.9
Data Rate	C_2	0.4503	0.05, 0.1	0.2, 0.94	0.2, 0.9	0.2, 0.9
Network Coverage Area	C_3	0.0187	0.1, 0.45	0.2, 0.85	0.2, 0.95	0.2, 0.92
Network Latency	C_4	0.1272	0.4, 0.6	0.5, 0.88	0.5, 0.9	0.5, 0.9
Reliability	C_5	0.0126	0.7, 0.9	0.7, 0.89	0.7, 0.9	0.7, 0.9
Security	C_6	0.0431	0.8, 0.9	0.8, 0.9	0.8, 0.9	0.8, 0.9
Power Requirement	C_7	0.0142	0.7, 0.8	0.6, 0.75	0.6, 0.75	0.6, 0.75
Mobile Velocity	C_8	0.0273	0.01, 0.9	0.005, 0.1	0.01, 0.5	0.01, 0.5
Service Cost	C_9	0.2343	0.5, 0.6	0.6, 0.81	0.6, 0.85	0.6, 0.87

The GA-related results: We once again performed the GA optimization experiments by using the stochastic universal selection rule that lays out a line in which each parent corresponds to a section of the line of length proportional to its scaled value, and used the control parameters: population size p_s = 20, elite count = 2, single-point crossover with p_c = 0.8, and mutation probability p_m = 0.01.

The solutions (a list of the optimum ANSF values and optimum membership function values) obtained using the GA with the MATLAB Genetic Algorithm Toolbox are summarised in Table 5.4.

From the results of the optimal values of the ANSF for the access networks in Table 5.4, the WiMAX_3 provides the optimal positive result and it will be suitable to handoff from the WLAN to the WiMAX_3 to continue with the video call.

Table 5.4: *The GA Optimization Values for Case 2*

Criteria		w_j	Optimal Membership Values			
			UMTS_2	WiMAX_1	WiMAX_2	WiMAX_3
RSSI	C_1	0.0722	0.9000	0.9000	0.8988	0.8981
Data Rate	C_2	0.4503	0.1000	0.9400	0.9000	0.8900
Network Coverage Area	C_3	0.0187	0.4384	0.8211	0.6842	0.9201
Network Latency	C_4	0.1272	0.6000	0.8789	0.8990	0.8973
Reliability	C_5	0.0126	0.9000	0.7552	0.7979	0.9000
Security	C_6	0.0431	0.8832	0.8946	0.8896	0.8960
Power Requirement	C_7	0.0142	0.7937	0.7473	0.7365	0.7234
Mobile Velocity	C_8	0.0273	0.9000	0.1000	0.5000	0.4940
Service Cost	C_9	0.2343	0.6000	0.8031	0.8500	0.8681
Optimal ANSF Value			-0.4203	-0.8650	-0.8690	-0.8783

The best fitness value and the best individual given by the GA for the WiMAX_3 network are plotted in Figure 5.2.

The GA converges quickly to the best solution for the control parameter settings indicated above. The GA terminates after 69 generations, and the objective function value of the best solution is -0.87831.

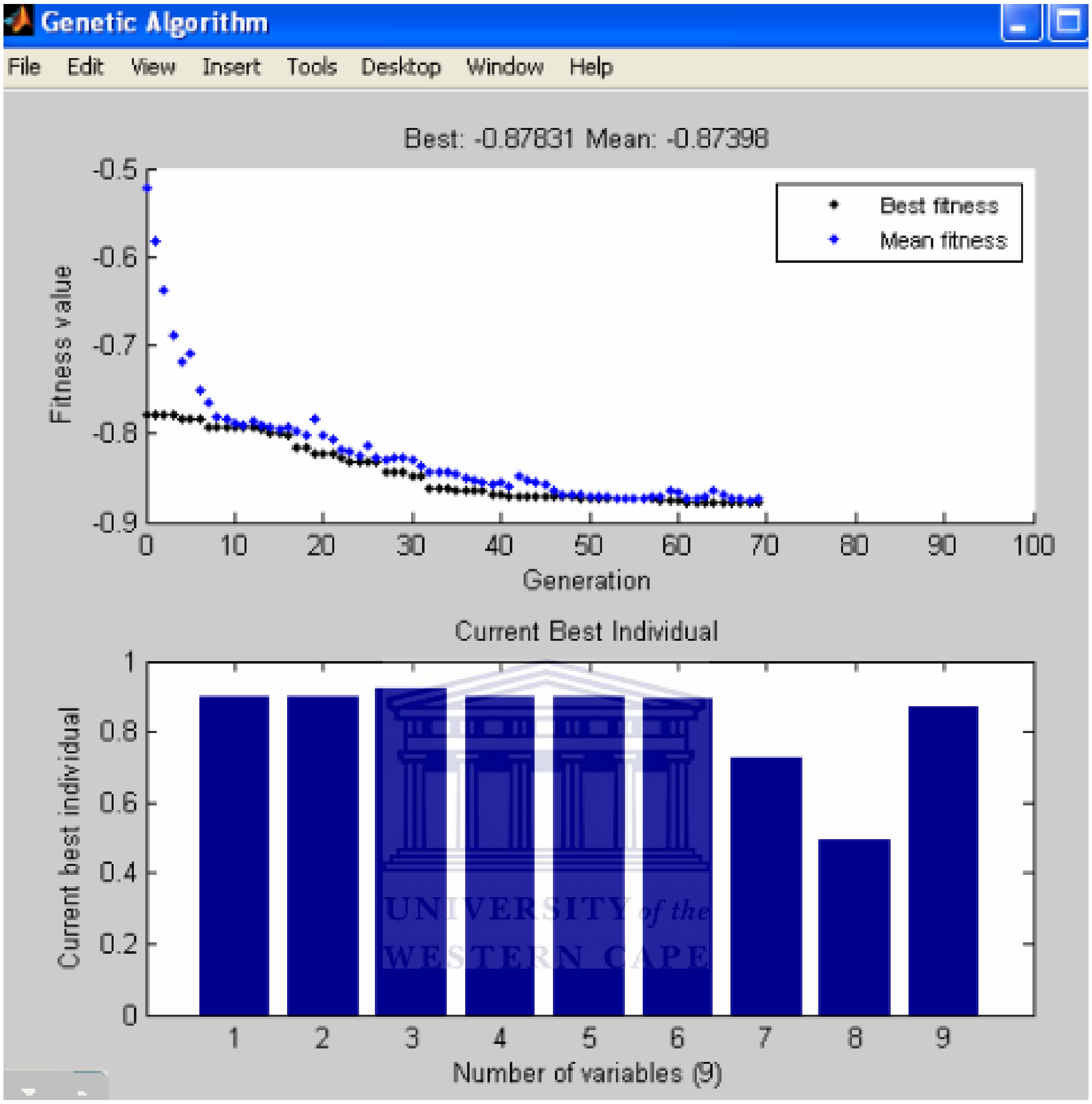


Figure 5.2: *The Case 2 GA Best Fitness Value and Best Individual for WiMAX_3*

5.6 Summary

This chapter defined the notion of optimization in general, explained why it makes sense to differentiate between discrete, continuous and mixed-variable optimization, and why genetic algorithms may be used for tackling these types of problems. An overview of genetic algorithms was given and then the chapter explored using the genetic algorithm to optimize the access network selection function defined in Chapter 4 with the view of selecting the optimal available access network for handoff.

We observe from the results in Table 4.3 and Table 4.5 that Mathematical Optimization always selects the upper bounds of a solution vector $\boldsymbol{x}$ for calculating the optimum value of an objective function, while the results in Table 5.3 and Table 5.4 show that the GA provides a list of the optimum objective function values and the optimum membership function values of the parameters. Consequently, the GA is better suited for solving problems of the type posed in Section 5.5.

CHAPTER 6

OPTIMIZATION OF ACCESS NETWORK SELECTION USING SIMULATED ANNEALING

In this chapter we present another natural meta-heuristic optimization method, the simulated annealing (SA), to solve the access network selection problem discussed in Chapter 4. The SA also generates new points in the search space by applying operators to current points and statistically moving toward more optimal places in the search space. An overview of the SA is first given before using the SA to optimize the access network selection function introduced in Chapter 4. Finally, a comparison of the performance of the GA and the SA is made.

6.1 Introduction

In our daily life and in various research domains, we encounter optimization problems some of which are so hard that we can at best approximate the best solutions with (meta-) heuristic methods. Simulated Annealing (SA) is a probabilistic meta-heuristic algorithm for solving unconstrained and bound-constrained optimization problems. It is a method to obtain an optimal solution of a single objective optimization problem and to obtain a Pareto-optimal set of solutions for a multiobjective optimization problem. The problem may have continuous or discrete variables. The SA is an iterative search method that models the physical process of annealing. It probabilistically generates a sequence of states based on an annealing schedule to ultimately converge to the global optimum. In the early 1980s Kirkpatrick *et al.* [96] and Cerny [97] showed that a model simulating the annealing of solids, proposed by Metropolis *et al.* [98], could be used for optimization of problems, where the objective function to be minimized corresponds to the energy of states of the metal. The SA has become one of the many metaheuristic approaches designed to give a good, not necessarily optimal solution.

The access network selection problem, described in Chapter 4, is not solvable realistically using mathematical optimization algorithms. Therefore, we explore the use of the SA for solving the optimization problem of maximizing the access network selection function of Chapter 4. Finally, we compare the performance of the GA and the SA for solving the access network selection problem.

6.2 Simulated Annealing

Simulated Annealing (SA) is a meta-heuristic algorithm for solving unconstrained and bound-constrained single objective and multiobjective optimization problems using a probabilistic search strategy.

6.2.1 The Physical Process of Annealing

The SA is based on an analogy of thermodynamics with the physical process of annealing, in which a material is heated to impart high energy to it and then its temperature slowly lowered to cool and anneal in the case of a metal or freeze and crystallize in the case of a liquid in order to reach a minimum energy state. The process of heating a liquid above its melting temperature and cooling it gradually produces the crystalline lattice which minimizes its energy probability distribution. However, if the cooling proceeds too quickly, that is, the liquid is quenched, the crystal never forms and the substance becomes an amorphous mass with a higher than optimum energy state. The key to reaching a frozen, steady state, corresponding to an absolute minimum of energy, is carefully controlling the rate of temperature change.

Annealing can be regarded as a natural physical optimization process [98]. In the context of mathematical optimization, the minimum of an objective function represents the minimum energy of the system. The SA is an algorithmic implementation of the annealing process to find the optimum of an objective function [87, 96, 97, 99]. In the SA algorithm, each point vector $\boldsymbol{x}$ of the search (or solution) space $\boldsymbol{S}$ is analogous to a state of some physical system, the objective function f to be minimized is analogous to the energy equation of the thermodynamic system in that state, the objective function value $f(\boldsymbol{x})$ to be minimized is analogous to the internal energy of the system in that state, and the global minimum is analogous to the ground state of the system. The effect of annealing is to perform the following unconstrained optimization:

minimize: $f(\boldsymbol{x})$

subject to: $\boldsymbol{x} \in \boldsymbol{S}$

The goal is to bring the system to an optimal state, that is, a state with the minimum possible energy, from an arbitrary initial state.

6.2.2 The Metropolis Algorithm

The SA algorithm is based on two theoretical results of statistical physics.

The SA consists in introducing a control parameter T in optimization, which plays the role of the temperature in physical annealing in accordance with the concept of the Boltzmann probability distribution. When thermal equilibrium is reached at a given temperature, the probability for a physical system to have a given energy E, is proportional to the Boltzmann factor: $\exp(- E / k_B T)$, where T is the temperature measured in Kelvin and k_B is the Boltzmann constant $k_B = 1.380650524 * 10^{-23}$ J/K.

In addition, to simulate the evolution of a physical system towards its thermal equilibrium at a given temperature, one can utilize the Metropolis algorithm, a Monte Carlo method used to generate sample states of a thermodynamic system, developed by Metropolis *et al.* in 1953 [98]. The Metropolis algorithm, which is presented in Table 6.1 [100], can be stated briefly as follows. Given the current state, s_i, of the solid with corresponding energy E_i a new random trial state, s_j, with energy E_j is generated but accepted only if it fulfills the Metropolis criterion:

$$P(\text{accept } s_j) = \begin{cases} 1, & \text{if } E_j < E_i \\ \exp((E_i - E_j) / k_B T), & \text{if } E_j > E_i \end{cases} \quad (6.1)$$

Table 6.1: *The Metropolis Algorithm*

```
Pick initial solution x_i randomly;
Set surrounding temperature T;
repeat
  Generate new solution s_j at random;
  if E_j < E_i then
     s_i = s_j;
  else
     s_i = s_j with probability exp ((E_i – E_j) / k_B T);
  end if
until Thermal equilibrium is reached;
```

6.2.3 The Simulated Annealing Algorithm for Single Objective Optimization

The SA algorithm has been introduced in the area of combinatorial optimization by Kirkpatrick *et al*. [96]. The SA is fundamentally the original Metropolis algorithm used with the purpose of obtaining a solution to a single optimization problem, close to the global optimum solution. The SA uses a random search strategy which not only accepts new positions that decrease the objective function (assuming a minimization problem), but also accepts positions that increase objective function values (uphill moves) in order to escape from local minima. The probability of doing an uphill move is decreased during the search.

The SA algorithm starts by generating a non-optimal solution $\boldsymbol{x_i}$ of the search (or solution) space $\boldsymbol{S}$, calculates its objective function value $f(\boldsymbol{x_i})$, and initializes the computational temperature parameter T. Then, it proceeds to improve the solution by randomly generating at each iteration, a new point $\boldsymbol{x_j} \in N(\boldsymbol{x_i})$ (the neighbourhood of $\boldsymbol{x_i}$) that is accepted as the new current solution if $\boldsymbol{x_j}$ has a better objective function value than $\boldsymbol{x_i}$, that is, if $f(\boldsymbol{x_j}) < f(\boldsymbol{x_i})$. If, on the other hand, $\boldsymbol{x_j}$ has a worse objective function value than $\boldsymbol{x_i}$, the solution is only accepted with a certain probability P_{ij} which is a function of (i) the temperature parameter T, and (ii) the difference $f(\boldsymbol{x_j}) - f(\boldsymbol{x_i})$ of the objective function values in $\boldsymbol{x_j}$ and $\boldsymbol{x_i}$:

$$P_{ij} = \exp(-(f(\boldsymbol{x_j}) - f(\boldsymbol{x_i})) / k_B T) \qquad (6.2)$$

This process is then repeated in an iterative manner, by keeping the constant temperature, until thermodynamic balance is reached, concretely at the end of a sufficient number of modifications. Then the temperature is lowered, before implementing a new series of transformations. Table 6.2 summarises the SA algorithm.

The SA works much like a bouncing ball that can bounce over mountains based on the ball's temperature until the highest tip is found. The higher the temperature, the more likely the ball will bounce over mountains and local valleys. As the temperature cools down, the ball bounces less strongly, until it stops at the bottom of the lowest valley encountered.

The initial temperature, cooling rate, and the number of iterations performed at a particular temperature are the three important parameters which govern the successful working of the SA procedure.

Table 6.2: *Simulated Annealing Algorithm*

```
Pick initial solution xi randomly;
Set initial value of control parameter T;
repeat
  repeat
    Generate new solution xj;
    Determine quality f(xj);
    if f(xj) < f(xi) then
      xi = xj;
    else if acceptance probability exp(−(f(xj) − f(xi)) / kB T) ≥ random(0, 1) then
      xi = xj;
    end
  until Equilibrium is established for current value of control parameter T;
  Decrease T according with the annealing schedule;
until stopping condition is true;
Return xj as the solution;
```


6.2.4 Key Components of the Simulated Annealing Algorithm

In order to obtain a working solution using the SA, we need a suitable objective function, a means of generating proposed states, an acceptance criterion for the proposed states, an annealing schedule, and a stopping criterion.

Objective function

The objective function is closely related to the problem formulation. It is the function you want to optimize and it provides a method to evaluate solutions.

Acceptance criterion

The acceptance probability P_{ij} of moving from point x_i to point x_j depends on the difference between the corresponding function values and a global parameter T, and is given by

$$P_{ij} = \begin{cases} 1, & \text{if } f(\boldsymbol{x}_j) < f(\boldsymbol{x}_i) \\ \exp\left(\left(f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j)\right) / k_B T\right), & \text{otherwise,} \end{cases} \tag{6.3}$$

where T is a control parameter, which corresponds to the temperature in analogy with the physical annealing. At a fixed temperature, the higher the difference $f(\boldsymbol{x}_j) - f(\boldsymbol{x}_i)$, the lower the probability to accept a move from x_i to x_j. On the other hand, the probability of uphill moves is higher for higher values of T. Therefore, the SA starts with a high temperature to avoid being trapped in a local minimum.

Annealing schedule

The *annealing schedule* consists of an initial value, T_0, of the temperature control parameter T and the rate by which it is decreased as the algorithm proceeds. The choice of an appropriate annealing schedule is crucial for the performance of the algorithm. An initial high temperature is selected, which is then incrementally reduced. If T is lowered too rapidly, then a global minimum will not be found because of inadequate exploration of the local minima. The slower the rate of decrease, the better the chances are of finding an optimal solution, but the longer the run time. The annealing schedule defines the value of T at each iteration k, $T_{k+1} = Q(T_k, k)$, where $Q(T_k, k)$ is a function of the temperature and the iteration number. Convergence of the SA to the set of optimal solutions is guaranteed if and only if the cooling schedule is sufficiently gradual [101], but experience has shown the SA to be a very effective optimization technique even with relatively rapid cooling schedules [102], [103]. Annealing schedules which guarantee convergence to a global optimum are not feasible in applications, because they are too slow for practical purposes. Therefore, faster annealing schedules are adopted in applications. An initial high temperature is selected, which is then incrementally reduced according to commonly used temperature update functions such as:

- Exponential annealing update: $T_{k+1} = \alpha\, T_k$, where $\alpha \in (0, 1)$.
- Logarithmic temperature update: $T_{k+1} = \Gamma / (\log(k + k_0)$, where $\Gamma \in \boldsymbol{R}$, and k_0 is a constant.

Stopping criterion

The algorithm stops when any of the following criteria are met:

- The average change in the objective function is very small;
- The objective function goes below objective limit;
- Only a small fraction of the proposed states are accepted for the current value of the control parameter.

6.3 The Simulated Annealing Optimization of the Access Network Selection Function

To tackle the optimization problem of maximizing the ANSF in equation (4.16) of Chapter 4 under the weights w_j by using a SA, we assume a function minimization problem. Hence, a good solution is one that has low relative fitness. Since our SA algorithm performs minimization of an objective function $f(\boldsymbol{x})$, maximization of the objective function in equation (4.16) is achieved by supplying the routine with minus $f_i(\boldsymbol{x})$ because the point at which the minimum of $-f_i(\boldsymbol{x})$ occurs is the same as the point at which the maximum of $f_i(\boldsymbol{x})$ occurs. We note that the optimization problem in this section is exactly the same problem of section 5.4.

Therefore, we define the equivalent minimization problem according to equations (5.3)-(5.6):

$$\min\,(-f_i(\boldsymbol{x})) = \min\{-\sum_{j=1}^{n} w_j \cdot \mu_{Cj}(A_i)\} \tag{5.3}$$

such that

$$0 \le w_j \le 1, \text{ and } \sum_{j=1}^{n} w_j = 1, \tag{5.4}$$

and

$$\{\mu_{Cj}(A_i)\}_{\min} \leq \mu_{Cj}(A_i) \leq \{\mu_{Cj}(A_i)\}_{\max}. \quad (5.5)$$

$$\sum_{i=1}^{9} \mu_{Cj}(A_i) \leq 9. \quad (5.6)$$

In order to achieve good results in using a SA, one may have to experiment with different SA parameters.

6.4 Performance Evaluation of the ANS Algorithm with Simulated Annealing

The performance of the vertical handoff decision algorithm is tested within the framework of a scenario that simulates a typical day in the life of a telecommunication field technician, Mr. Alex Star as presented in section 3.7.

6.4.1 Use Cases:

We continue with the two use cases presented in section 3.7.1.

6.4.2 Evaluation using Simulated Annealing:

Case 1:

We first check to see whether a handoff should be initiated by calculating the handoff initiation factor.

As shown in section 3.7.2, the three sets of values were fed into the FIS and we obtained the Handoff Factor values 0.897, 0.902, and 0.903 for WiMAX_1, WiMAX_2, and WiMAX_3 respectively, thus indicating the need to hand off to any of the WMANs for the requested service.

The second stage of the VHDA is to compute the ANSF for all the available networks. The mobile terminal proceeds to gather data on all required parameters. The matrix ***B*** and weighting matrix ***W*** were computed and indicated as equation 4.21 in section 4.4.

The attribute weights and the membership values (lower bound, upper bound) obtained from the characteristics of the four available networks for the attributes were summarised in Table 4.2 of Chapter 4 and reproduced below.

Table 4.2: *Parameters for Performance Evaluation of Case 1*

Criteria		w_j	Membership Values (lb, ub)			
			UMTS_2	WiMAX_1	WiMAX_2	WiMAX_3
RSSI	C_1	0.1272	0.5, 0.9	0.5, 0.9	0.5, 0.9	0.5, 0.9
Data Rate	C_2	0.2343	0.05, 0.1	0.2, 0.94	0.2, 0.9	0.2, 0.9
Network Coverage Area	C_3	0.0273	0.1, 0.45	0.2, 0.85	0.2, 0.95	0.2, 0.92
Network Latency	C_4	0.0142	0.4, 0.6	0.5, 0.88	0.5, 0.9	0.5, 0.9
Reliability	C_5	0.0431	0.7, 0.9	0.7, 0.89	0.7, 0.9	0.7, 0.9
Security	C_6	0.0722	0.8, 0.9	0.8, 0.9	0.8, 0.9	0.8, 0.9
Power Requirement	C_7	0.0126	0.7, 0.8	0.6, 0.75	0.6, 0.75	0.6, 0.75
Mobile Velocity	C_8	0.0187	0.01, 0.9	0.005, 0.1	0.01, 0.5	0.01, 0.5
Service Cost	C_9	0.4503	0.5, 0.6	0.6, 0.81	0.6, 0.85	0.6, 0.87

The SA-related results: We performed the SA optimization experiments by using Fast Annealing as the annealing function, Exponential Temperature update as the temperature update function, and the initial temperature as 100.

Table 6.3: *Parameters Used for the SA*

	Case 1	**Case 2**
Annealing function	Fast Annealing	Fast Annealing
Temperature update function	Exponential Temperature	Exponential Temperature
Initial temperature	100	100

The solutions (a list of the optimum ANSF values and optimum membership function values) obtained using the SA (MATLAB GA toolbox) are summarized in Table 6.4.

The results in Table 6.4 show again that the WiMAX_3 provides the optimal positive result and it will be suitable to handoff from the UMTS_2 to the WiMAX_3 to complete downloading the large file.

Table 6.4: *The SA Optimization Values for Case 1*

Criteria		w_j	Optimal Membership Values			
			UMTS_2	WiMAX_1	WiMAX_2	WiMAX_3
RSSI	C_1	0.1272	0.8992	0.8999	0.8998	0.8950
Data Rate	C_2	0.2343	0.0997	0.9399	0.8999	0.8981
Network Coverage Area	C_3	0.0273	0.4497	0.8499	0.9496	0.9140
Network Latency	C_4	0.0142	0.4001	0.8799	0.5764	0.8924
Reliability	C_5	0.0431	0.8986	0.8893	0.8730	0.8963
Security	C_6	0.0722	0.8999	0.8996	0.8993	0.8984
Power Requirement	C_7	0.0126	0.7993	0.6006	0.7500	0.7478
Mobile Velocity	C_8	0.0187	0.8986	0.0999	0.4885	0.4932
Service Cost	C_9	0.4503	0.5999	0.8098	0.8482	0.8704
Optimal ANSF Value			-0.5564	-0.8477	-0.8625	-0.8758

The best fitness value and the best individual given by the SA for the WiMAX_3 network are plotted in Figure 6.1. The upper plot displays the best and mean fitness values in each iteration. The lower plot displays the coordinates of the point with the best fitness value in the current iteration. The SA converges slowly to the best solution for the parameter settings indicated above and terminates if the maximum number of the SA iterations reaches 9670 iterations. After 9670 iterations, the objective function value of the best solution is -0.8733.

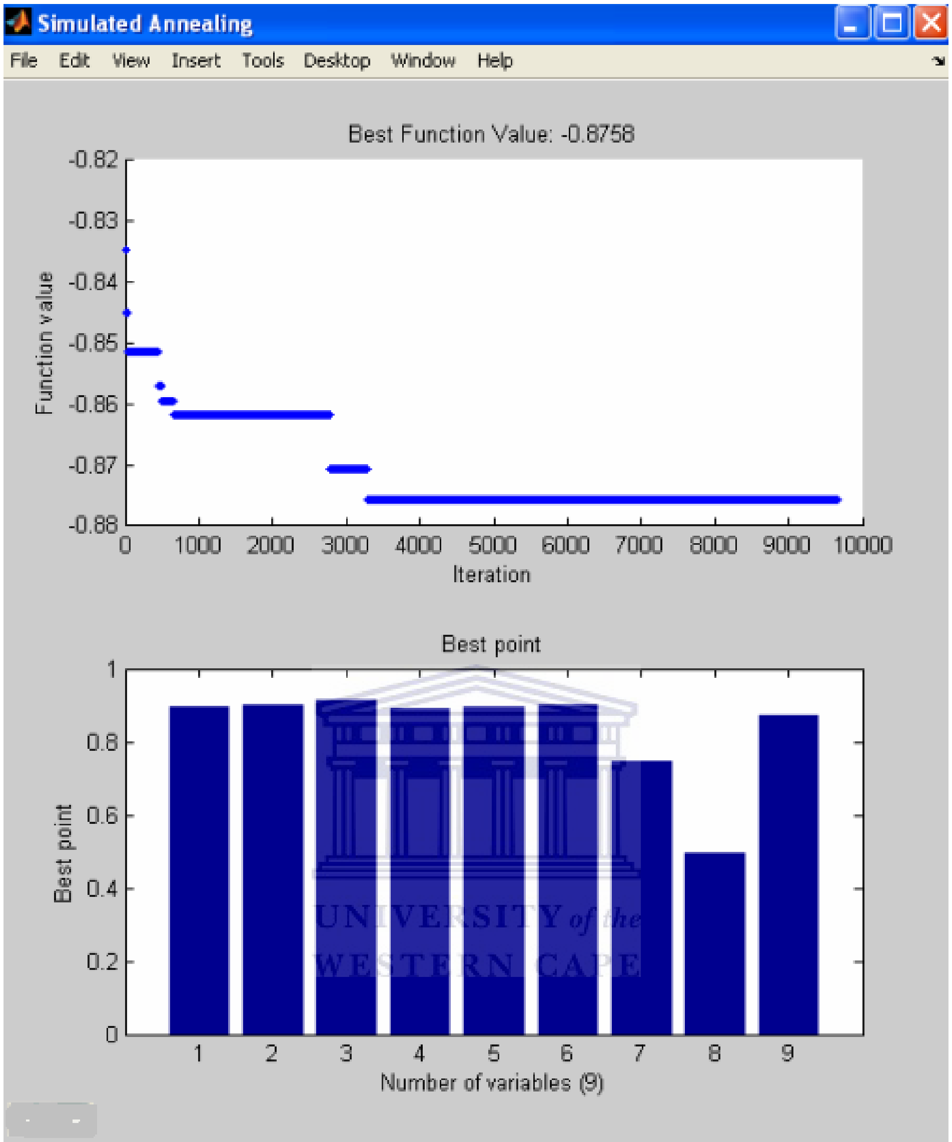


Figure 6.1: *The Case 1 SA Best Fitness Value and Best Individual for WiMAX_3*

Case 2:

We first check to see whether a handoff should be initiated by calculating the handoff initiation factor.

As shown in Chapter 3, we obtained the Handoff Factor values 0.886, 0.892 and 0.895 for WiMAX_1, WiMAX_2, and WiMAX_3, respectively, and thus indicating the need to handoff to any of the three WiMAX networks for the requested service.

The second stage of the VHDA is to compute the ANSF for all the available networks. The mobile terminal proceeds to gather data on all required parameters. The matrix ***B*** and weighting matrix ***W*** were indicated in subsection 4.4.2 as equation 4.22.

We reproduce below the attribute weights and the membership values (lower bound, upper bound) obtained from the characteristics of the four available networks for the attributes and summarised in Table 4.4 of Chapter 4.

Table 4.4: *Parameters for Performance Evaluation for Case 2*

Criteria		w_j	Membership Values (lb, ub)			
			UMTS_2	WiMAX_1	WiMAX_2	WiMAX_3
RSSI	C_1	0.0722	0.5, 0.9	0.5, 0.9	0.5, 0.9	0.5, 0.9
Data Rate	C_2	0.4503	0.05, 0.1	0.2, 0.94	0.2, 0.9	0.2, 0.9
Network Coverage Area	C_3	0.0187	0.1, 0.45	0.2, 0.85	0.2, 0.95	0.2, 0.92
Network Latency	C_4	0.1272	0.4, 0.6	0.5, 0.88	0.5, 0.9	0.5, 0.9
Reliability	C_5	0.0126	0.7, 0.9	0.7, 0.89	0.7, 0.9	0.7, 0.9
Security	C_6	0.0431	0.8, 0.9	0.8, 0.9	0.8, 0.9	0.8, 0.9
Power Requirement	C_7	0.0142	0.7, 0.8	0.6, 0.75	0.6, 0.75	0.6, 0.75
Mobile Velocity	C_8	0.0273	0.01, 0.9	0.005, 0.1	0.01, 0.5	0.01, 0.5
Service Cost	C_9	0.2343	0.5, 0.6	0.6, 0.81	0.6, 0.85	0.6, 0.87

The SA-related results: We once again performed the SA optimization experiments by using Fast Annealing as the annealing function, Exponential Temperature update as the temperature update function, and the initial temperature as 100.

The solutions (a list of the optimum ANSF values and optimum membership function values) obtained using the SA (MATLAB GA toolbox) are summarized in Table 6.5.

The results in Table 6.5 show again that the WiMAX_3 provides the optimal positive result and it will be suitable to handoff from the WLAN to the WiMAX_3 in order for Mr. Star to continue with the video call.

Table 6.5: *The SA Optimization Values for Case 2*

Criteria		w_j	Optimal Membership Values			
			UMTS_2	WiMAX_1	WiMAX_2	WiMAX_3
RSSI	C_1	0.0722	0.8989	0.8989	0.8999	0.8874
Data Rate	C_2	0.4503	0.0998	0.9392	0.8998	0.8970
Network Coverage Area	C_3	0.0187	0.4444	0.8499	0.9494	0.9170
Network Latency	C_4	0.1272	0.5962	0.8796	0.8997	0.8941
Reliability	C_5	0.0126	0.8962	0.7031	0.8998	0.8900
Security	C_6	0.0431	0.9000	0.8984	0.8999	0.8980
Power Requirement	C_7	0.0142	0.7997	0.7481	0.6000	0.7440
Mobile Velocity	C_8	0.0273	0.8857	0.0999	0.4846	0.4863
Service Cost	C_9	0.2343	0.5981	0.8094	0.8422	0.8664
Optimal ANSF Value			-0.4198	-0.8662	-0.8716	-0.8753

The best fitness value and the best individual given by the SA for the WiMAX_3 network are plotted in Figure 6.2. The SA converges slowly to the best solution for the parameter settings indicated above and terminates if the maximum number of the SA iterations reaches 8813 iterations. After 8813 iterations, the objective function value of the best solution is -0.87531.

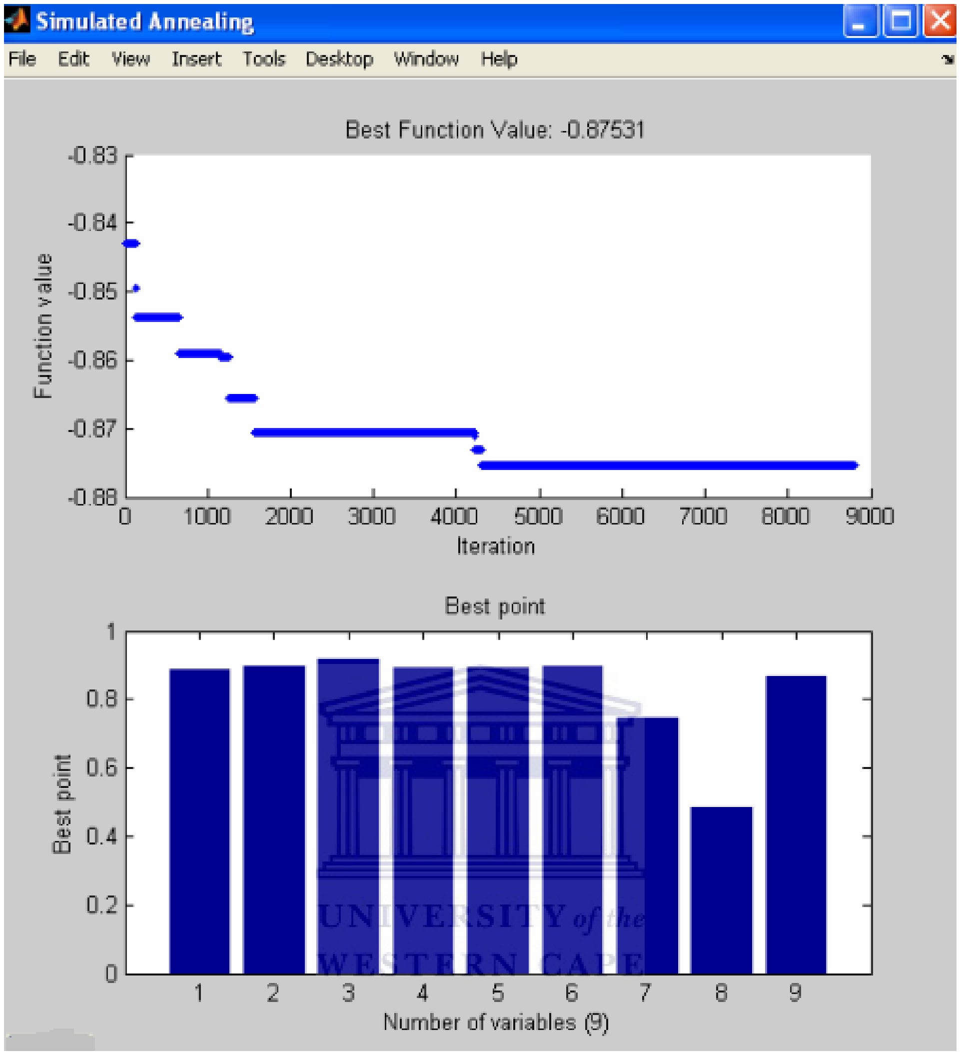


Figure 6.2: *The Case 2 SA Best Fitness Value and Best Individual for WiMAX_3*

6.5 Comparison of the Performances of the Genetic Algorithm and the Simulated Annealing Optimizations

The access network selection function described in Chapters 4, 5 and 6 was used to provide an optimal solution for the access network selection problem of the vertical handoff algorithm. The optimization package MATLAB Genetic Algorithm and Direct Search Toolbox 2 running on a 1.8 GHz machine with 0.98 GB of RAM was used to generate the optimal solution for different problem settings.

Two metaheuristics were used to solve the ANSF problem. The first heuristic implements a Genetic Algorithm (GA) approach in Chapter 5 while the second heuristic implements a

Simulated Annealing (SA) algorithm in Chapter 6. Two different sets of experiments involving two use cases were conducted for the GA and another two different sets of experiments were conducted for the SA.

6.5.1 Comparison of Figure 5.1 and Figure 6.1

We first compare Figure 5.1 with Figure 6.1. They show plots of the best fitness value and the best individual given by the GA and the SA, respectively, for the WiMAX_3 network.

In Figure 5.1, the upper plot displays the corresponding changes in the best and mean fitness values in each generation while the lower plot displays the coordinates of the point with the best fitness value in the current generation. Figure 5.1 shows that the GA converges quickly to the best solution for the control parameter settings indicated in Table 5.2. The GA terminates if the maximum number of the GA generations reaches 63 generations. After 63 generations, the objective function value of the best solution for the WiMAX_3 network is -0.87327, which is the same value as the Optimal ANSF Value for WiMAX_3 indicated in Table 5.3.

In Figure 6.1, the upper plot displays the best and mean fitness values in each iteration. The lower plot displays the coordinates of the point with the best fitness value in the current iteration. The SA converges slowly to the best solution for the parameter settings indicated in Table 6.3 and terminates if the maximum number of the SA iterations reaches 9670 iterations. After 9670 iterations, the objective function value of the best solution is -0.8733.

The results in Figure 5.1 and Figure 6.1 indicate that both the GA and the SA yield high-quality solutions in terms of objective function and running time. However, the running time of the SA was higher in the considered experimental setting.

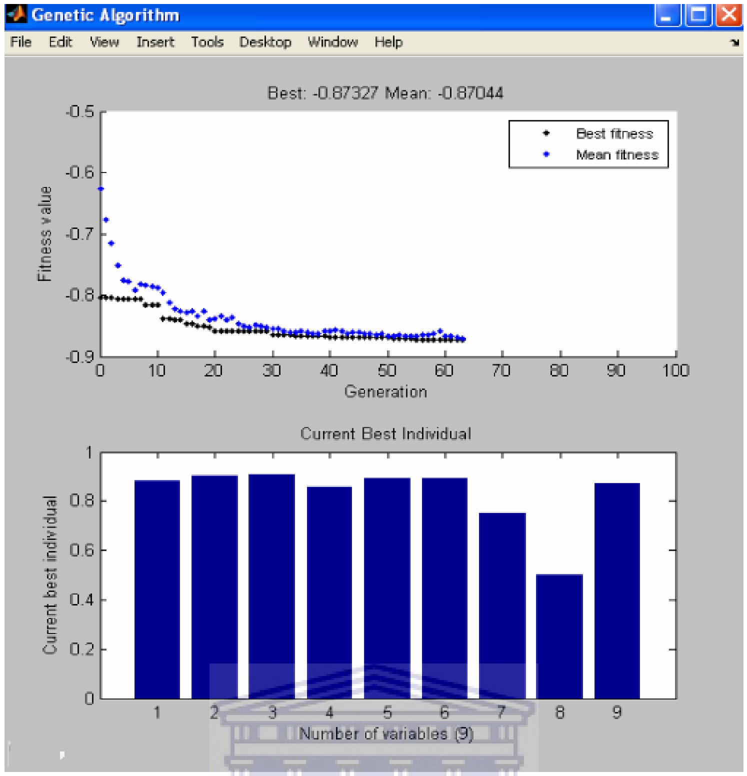


Figure 5.1: *The Case 1 GA Best Fitness Value and Best Individual for WiMAX_3*

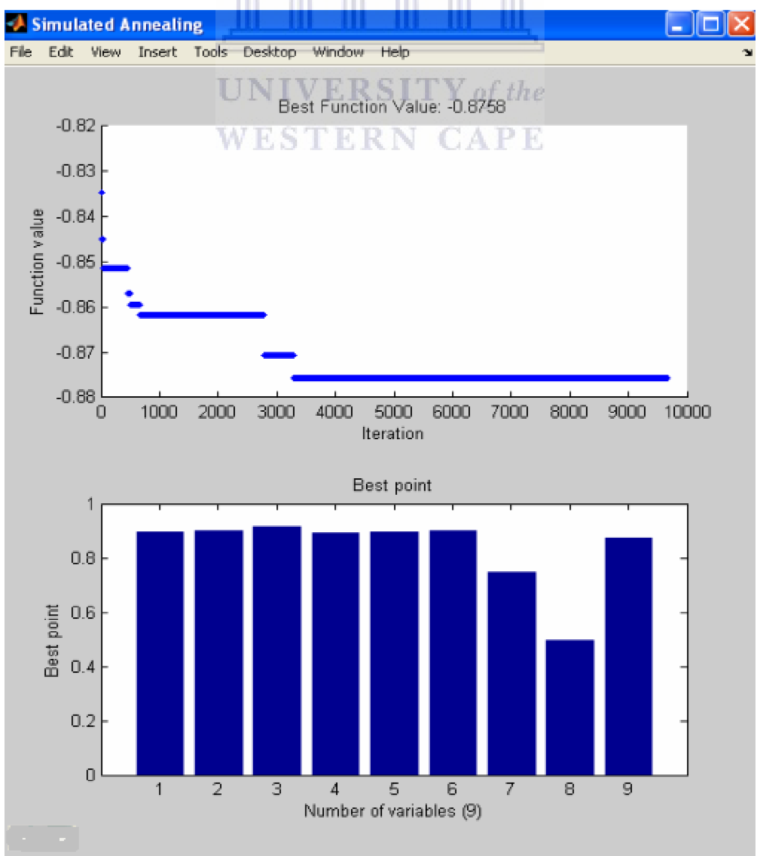


Figure 6.1: *The Case 1 SA Best Fitness Value and Best Individual for WiMAX_3*

6.5.2 Comparison of Figure 5.2 and Figure 6.2

We next compare Figure 5.2 with Figure 6.2. They show plots of the best fitness value and the best individual given by the GA and the SA, respectively, for the WiMAX_3 network for the Use Case 2.

As shown in Figure 5.2, the GA converges quickly to the best solution for the control parameter settings indicated in Table 5.2. The GA terminates after 69 generations, and the objective function value of the best solution for the WiMAX_3 network is -0.87831.

The SA converges slowly in Figure 6.2 to the best solution for the parameter settings indicated Table 6.3 and terminates if the maximum number of the SA iterations reaches 8813 iterations. After 8813 iterations, the objective function value of the best solution for the WiMAX_3 network is -0.87531.

Again, the results in Figure 5.2 and Figure 6.2 indicate that both the GA and the SA yield high-quality solutions in terms of objective function and running time. However, the running time of the SA was higher in the considered experimental setting.

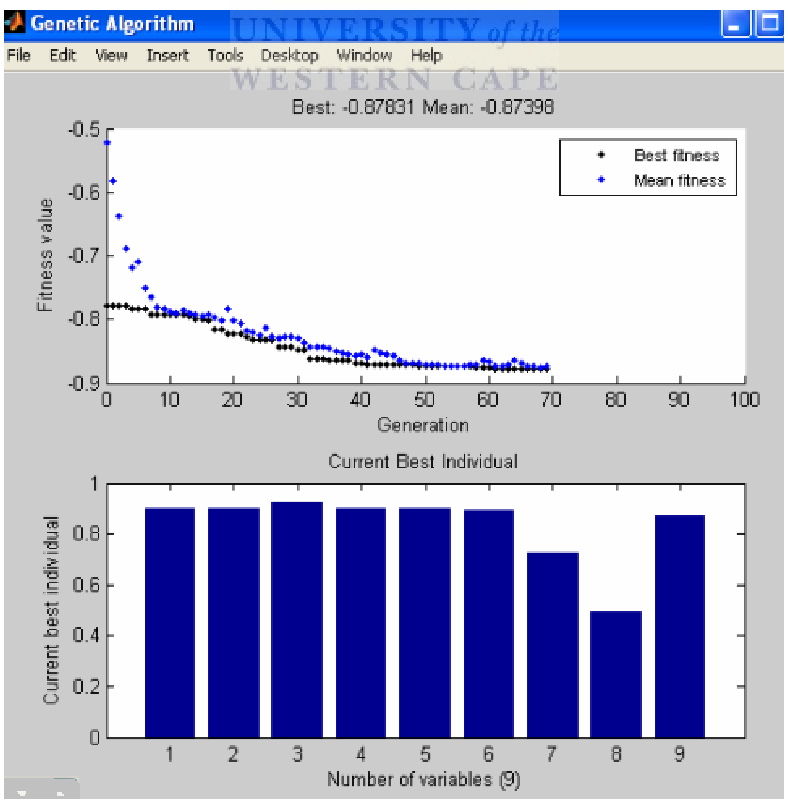


Figure 5.2: *The Case 2 GA Best Fitness Value and Best Individual for WiMAX_3*

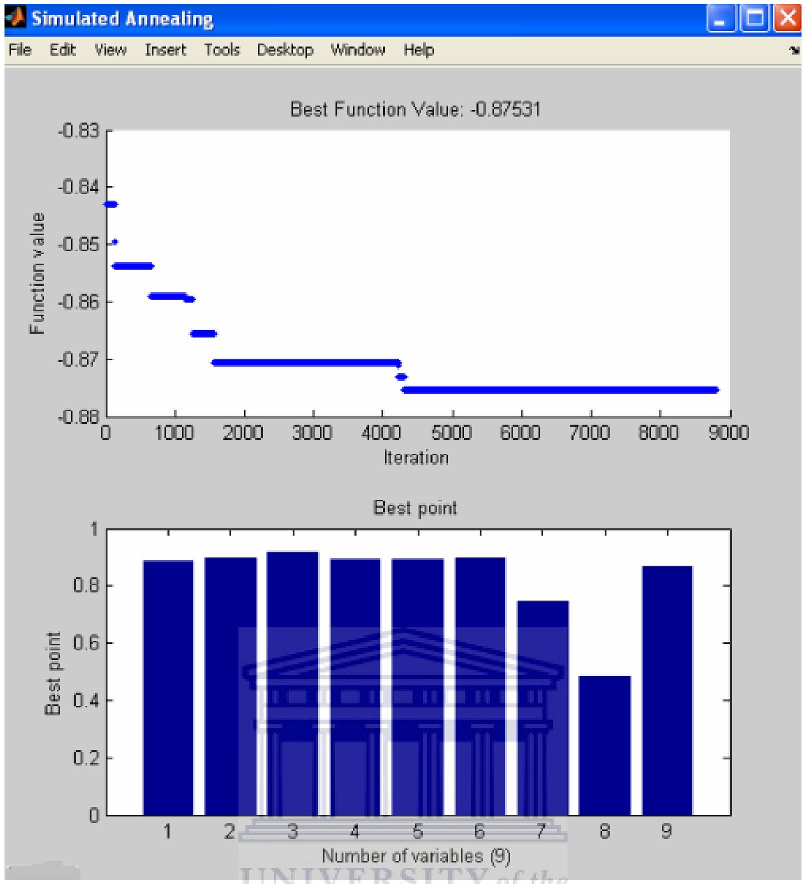


Figure 6.2: *The Case 2 SA Best Fitness Value and Best Individual for WiMAX_3*

6.5.3 Conclusion

The results indicate that both the GA and the SA yield high-quality solutions in terms of objective function and running time. However, the running time of the SA was higher in all considered experimental settings by comparing Figure 5.1 with Figure 6.1 and Figure 5.2 with Figure 6.2 (and additional plots in Figure A.1 with Figure A.2 and Figure A.3 with Figure A.4 in Appendix A) in the performance evaluation. Therefore, the GA will be the first option for optimizing the ANSF problem because of its simpler computational complexity, even though the SA is also suitable.

However, the SA takes less Central Processing Unit (CPU) time than the GA when used to solve optimization problems, because the SA finds the optimal solution using point-by-point iteration rather than a search over a population of individuals. Therefore, if the CPU time were to be used then the SA may outperform the GA.

6.6 Summary

This chapter gave an overview of the simulated annealing (SA) and then used the SA to optimize the access network selection function defined in Chapter 4 with the view of selecting the optimal available access network for handoff.

We conclude that the SA (just like the GA) is a better approach for solving problems of the type posed in section 6.4 since the results as presented in Table 6.4 and Table 6.5 are far more improved than the mathematical optimization results in Table 4.3 and Table 4.5 respectively.

CHAPTER 7

CONCLUSIONS

In this chapter a final summary of the thesis report is given. Contributions made to the research field and future plans are also presented.

7.1 Summary

1. First the background and motivation for automating the field workforce was given. It was stipulated that provision of seamless mobility for the user in the next generation wireless networks is the necessary missing ingredient to make a field workforce automation software a complete automation solution and that seamless vertical handoff was the required function to realize the seamless mobility experience.

2. Next there was an exhaustive presentation on vertical handoffs in next generation wireless networks. Related vertical handoff research literature was reviewed.

3. The design of a novel terminal-controlled handoff architecture that ensures seamless vertical handoff was presented. The vertical handoff decision module of the handoff architecture was presented as consisting of two components: handoff initiation and access network selection. The next step was to design the vertical handoff initiation algorithm using a fuzzy inference system.

4. The access network selection algorithm was designed using fuzzy multiple attribute decision making for a single objective optimization access network selection function. The access network selection function was optimized using a typical mathematical optimization function which always selected the upper bounds of a solution vector $\boldsymbol{x}$ for calculating the optimum value of an objective function.

5. The access network selection function was optimized using a genetic algorithm (GA). The performance evaluation of the GA for two use cases showed that the optimized access network selection function can select the optimal available access network for handoff. The results showed that the GA provides a list of the optimum objective function values and the optimum membership function values of the parameters.

6. The access network selection function was optimized using simulated annealing (SA). The performance evaluation of the SA for two use cases showed that the SA provides a list of the optimum objective function values and the optimum membership function values of the parameters and therefore the optimized access network selection function can select the optimal available access network for handoff.

7. A comparison of the performances of the genetic algorithm and the simulated annealing showed that both the GA and the SA are suitable as optimization algorithms for the access network selection component of the vertical handoff decision algorithm even though the GA seemed to be better than the SA in terms of their running times.

7.2 Contributions

1. The first contribution is the importance of providing a seamless mobility experience to a user in the next generation wireless networks. Seamless mobility is needed to make the field workforce automation software become a full field workforce automation solution that delivers business benefits. Seamless mobility allows field workforces to communicate seamlessly with the central company management.

2. The second contribution is the proposal of a novel terminal-controlled handoff management architecture that provides context-awareness and seamless handoff in next generation heterogeneous wireless IP access networks under the

motivation of increasing user satisfaction while maintaining a satisfactory level of QoS and seamless connectivity. The architecture allows users to configure the operational mode of their MTs and prioritise their preferences through changing weight factors and constraints of a single-objective access network selection optimization function; uses fuzzy multiple attributes decision making (FMADM) to select a suitable access network among a number of candidate networks with respect to different criteria; uses fuzzy logic to deal with imprecise data that the decision attributes might contain, and to combine and evaluate multiple attributes simultaneously; uses optimization techniques including the Genetic Algorithm (GA) and Simulated Annealing (SA) to maximize or minimize key parameters in order to select an optimum target access network for possible vertical handoff; and does not require significant changes to legacy systems since service providers do not want their existing systems to be modified due to cost and complexity.

3. The third contribution is the design of an optimal and intelligent vertical handoff decision algorithm that allows seamless mobility to be achieved. The vertical handoff decision algorithm consists of two parts: a handoff initiation algorithm and an access network selection algorithm.

4. The fourth contribution is the full design of the vertical handoff initiation algorithm using a fuzzy inference system to process multiple vertical handoff initiation parameters (criteria). It uses fuzzy logic to deal with imprecise data that the decision attributes might contain, and to combine and evaluate multiple attributes simultaneously.

5. The fifth contribution is the modeling and design of the access network selection algorithm (ANSA) using fuzzy multiple attributes decision making (FMADM) which deals with the problem of choosing an alternative from a set of alternatives based on the classification of their imprecise attributes. The ANSA applies a multiple attributes defined access network selection function (ANSF) to select the best access network that is optimized to the user's location, device conditions, service and application requirements, cost of service and throughput.

6. The sixth contribution is the optimisation of the ANSF using a genetic algorithm to select an available optimal access network for handing off.

7. The seventh contribution is the optimisation of the ANSF using simulated annealing to select an available optimal access network for handing off.

7.3 Future Work

This thesis involved the design of an optimal and intelligent vertical handoff decision algorithm as part of a seamless vertical handoff algorithm that is the enabling function for achieving the seamless mobility experience which is the missing ingredient needed to make the field workforce automation software a complete field workforce automation solution.

Traditional type-1 fuzzy sets (T1 FSs) and type-1 fuzzy inference systems (T1 FISs) have been used in the thesis. Although T1 FSs have been successfully used in many applications, however, such FSs have limited capabilities to directly handle lots of data uncertainties which can be handled by type-2 fuzzy sets (T2 FSs). Therefore, in future we plan to use T2 FSs and T2 FISs. Hopefully, designing and implementing the vertical handoff initiation algorithm using the T2 FIS will completely tackle all data uncertainties.

The FMADM was used in designing a single objective optimization access network selection function from the perspective of the mobile user for the decision making phase of the access network selection algorithm.

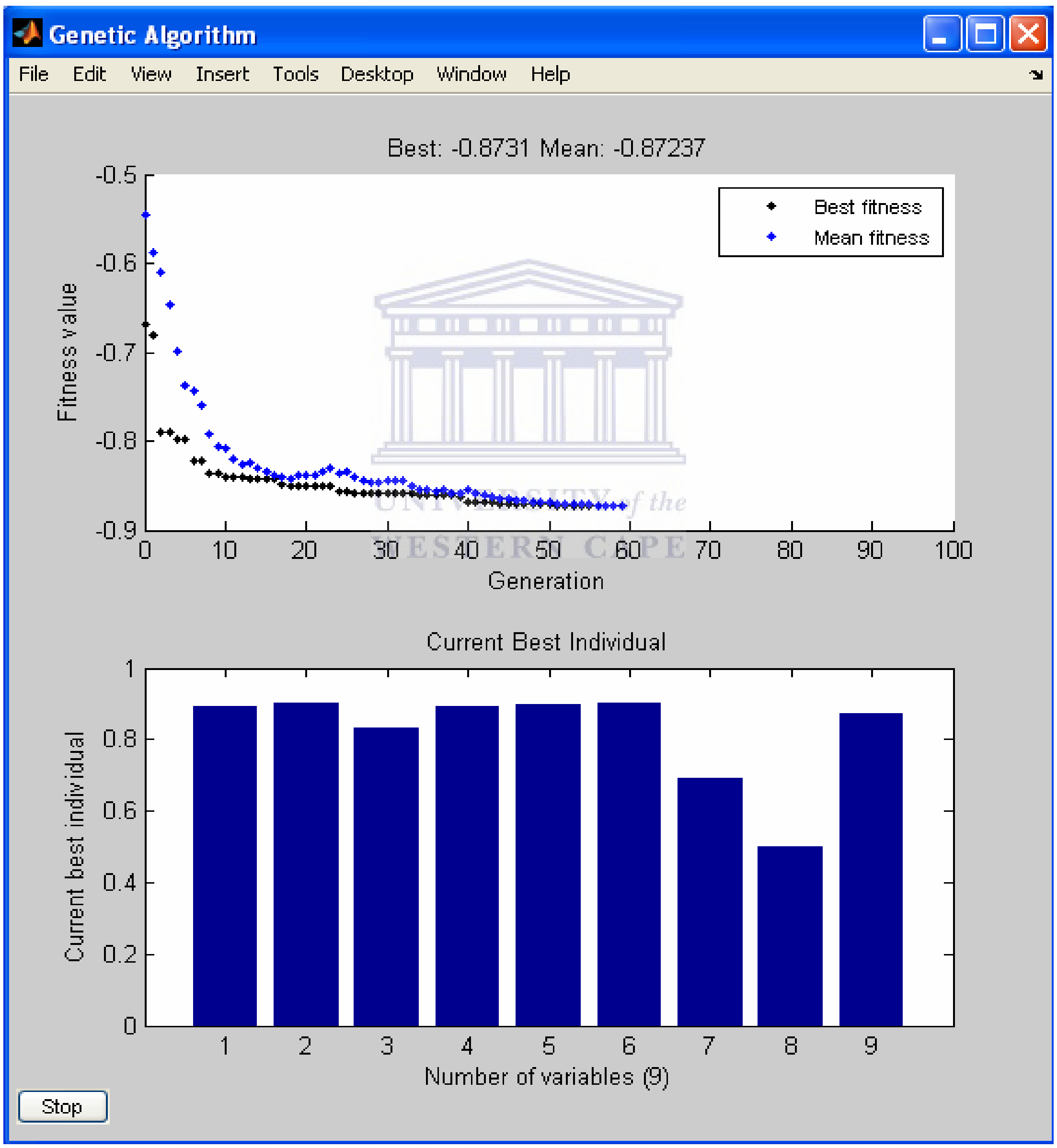
With seamless mobility, users can exploit all the access technologies to best meet their service costs and QoS requirements by automatically selecting the best access network through changing weight factors and constraints in a single objective optimization function. At the same time, operators may exploit seamless mobility in order to improve their network capacity as well as offer and improve the availability of compelling value-added services. It will be interesting to look at designing a vertical handoff decision algorithm from both the user and operator views as a dual-objective optimization problem containing two competing objectives and use fuzzy multiple objective decision making instead of FMADM. There is a premium on methods that can handle multiple objectives and discover optimal tradeoff (Pareto-optimal front) between these objectives. We shall use multiobjective genetic algorithms to solve the proposed dual-objective optimization problem since genetic algorithms and other population-based approaches are suited to handle multiple objectives as they can process a number of solutions in parallel and find all or majority of the solutions in the Pareto-optimal front. We shall also use multiobjective simulated annealing methods.

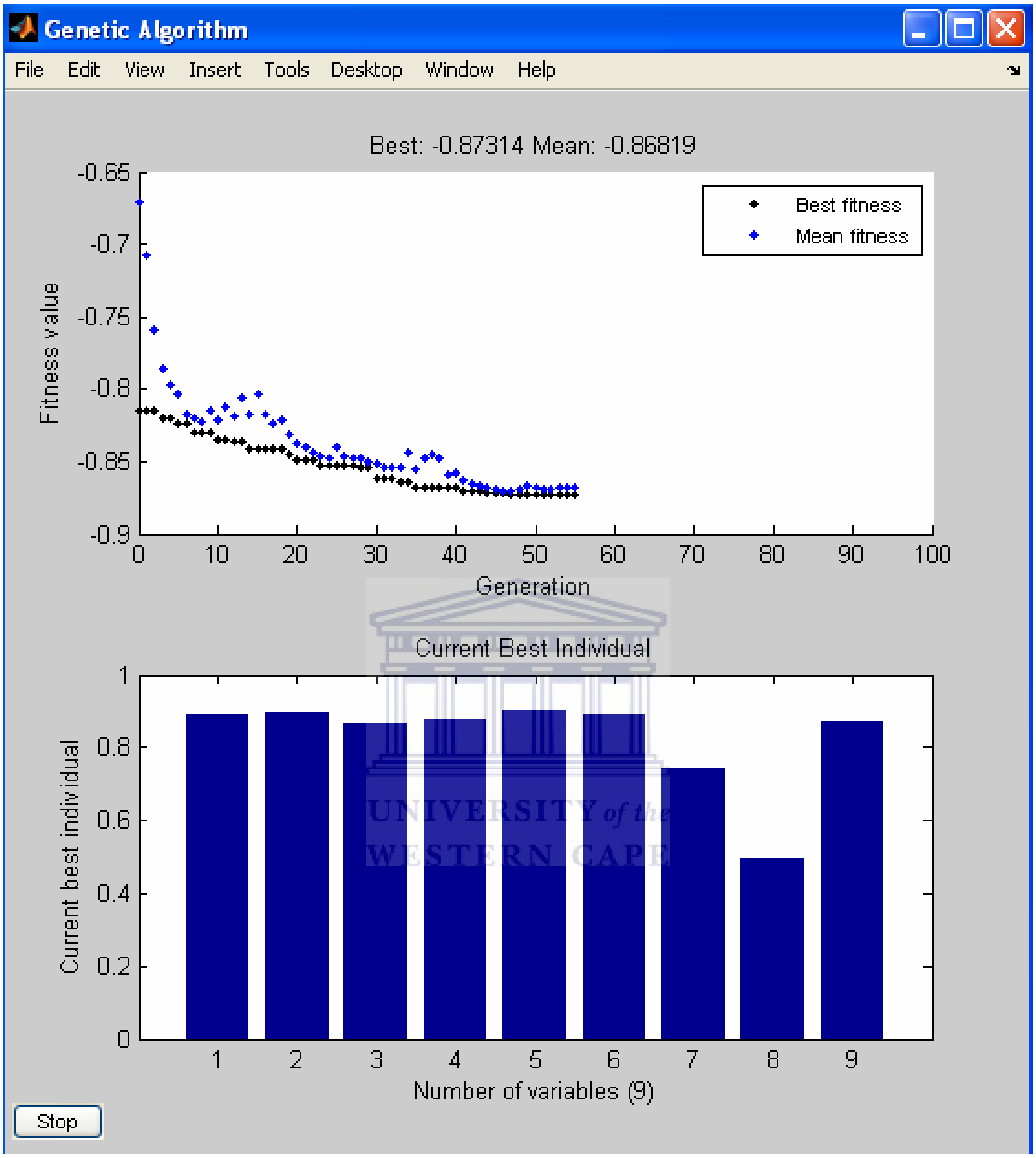
The thesis introduced the seamless mobility experience as the missing dimension needed to make the field workforce automation software become a full field workforce automation solution that delivers business benefits. Fuzzy logic and the fuzzy multiple attribute decision making method were used in designing an intelligent vertical handoff decision algorithm. The thesis demonstrated the potential of using genetic algorithms and simulated annealing for optimizing the access network selection problem in vertical handoffs in the next generation wireless networks. Whilst many issues have been dealt with, more research is required in the fields of seamless vertical handoff, seamless mobility, and mobile (field) workforce automation.

Appendix A

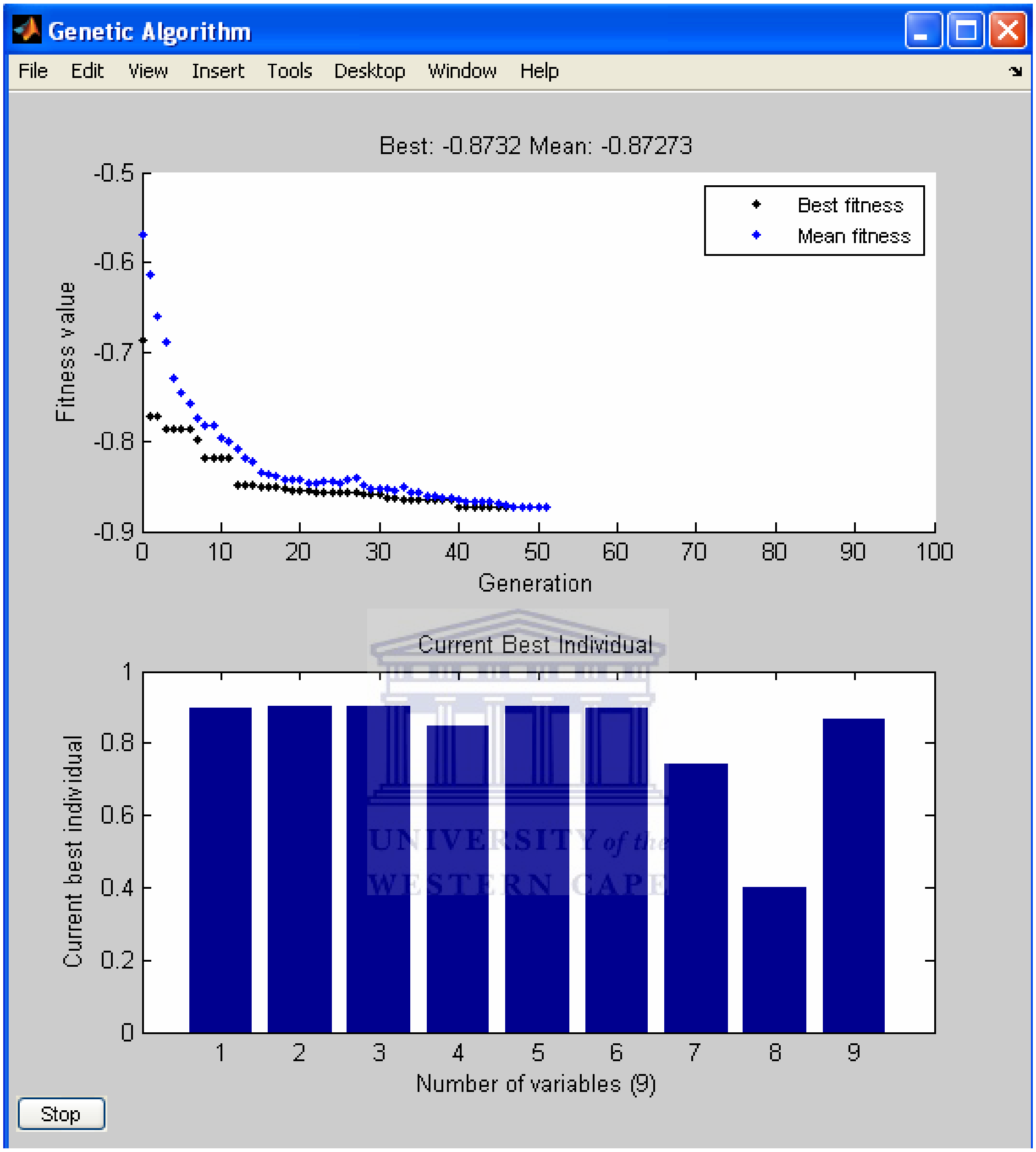
Additional Plots for Use Case1 and Use Case 2

Case 1: Additional plots of the best fitness value and the best individual given by the GA and the SA for the WiMAX_3 network are shown in Figure A.1 and Figure A.2 respectively.





Genetic Algorithm
File Edit View Insert Tools Desktop Window Help
Best: -0.87314 Mean: -0.86819
Best fitness
Mean fitness
Fitness value
-0.65
-0.7
-0.75
-0.8
-0.85
-0.9
0 10 20 30 40 50 60 70 80 90 100
Generation
Current Best Individual
Current best individual
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8 9
Number of variables (9)
Stop


Genetic Algorithm
File Edit View Insert Tools Desktop Window Help
Best: -0.8732 Mean: -0.87273
Best fitness
Mean fitness
Fitness value
Generation
Current Best Individual
Current best individual
Number of variables (9)
Stop

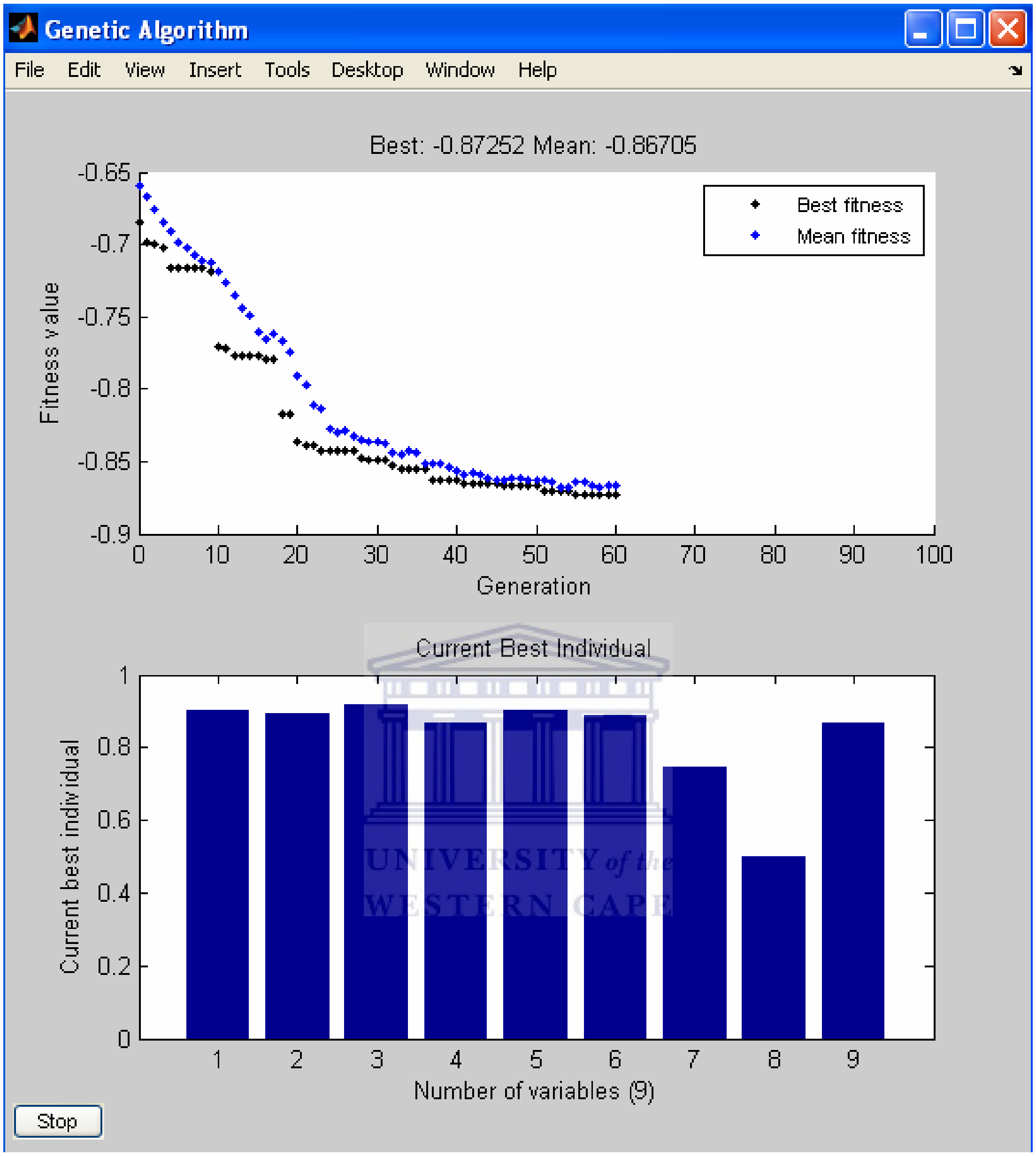
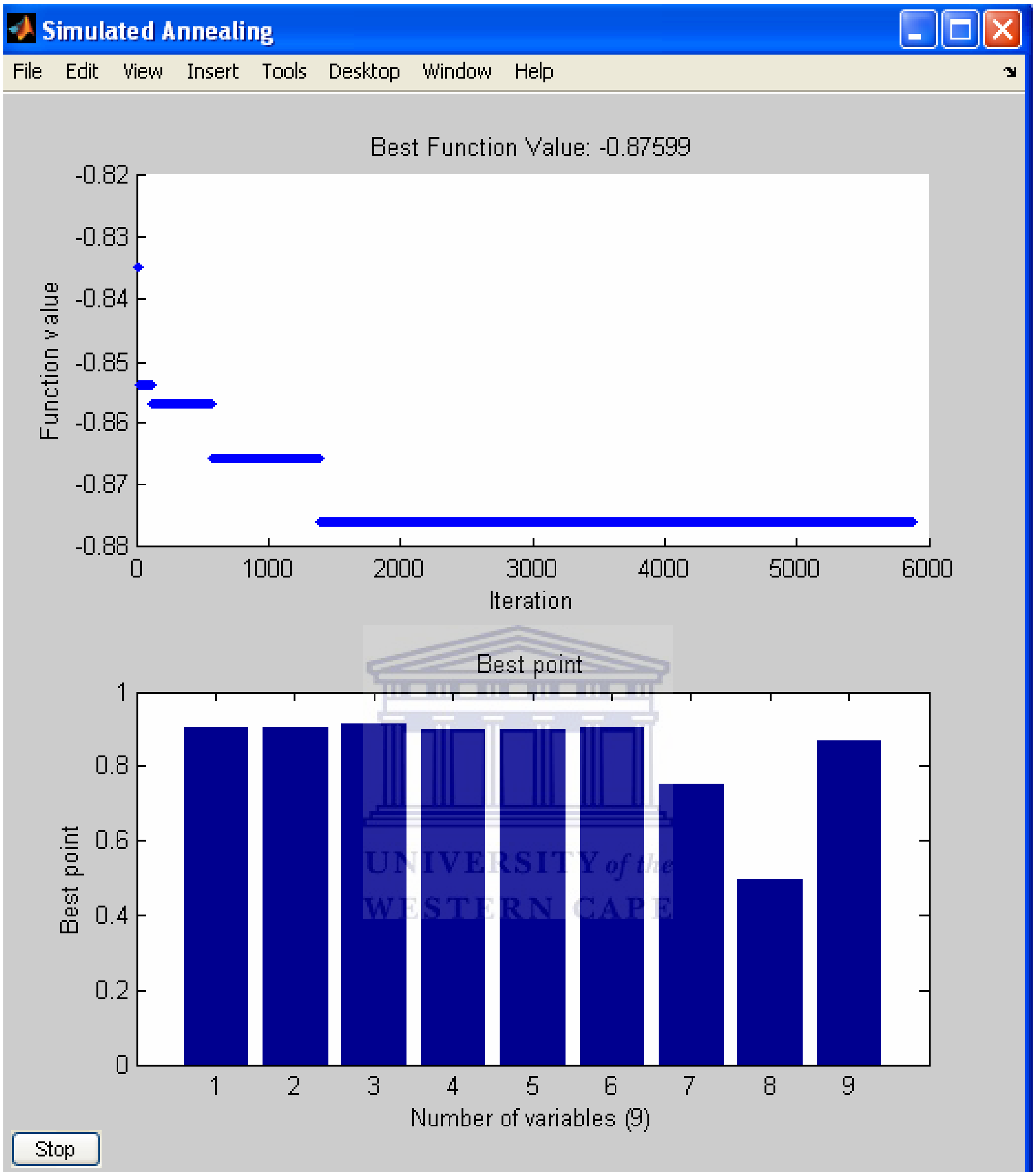
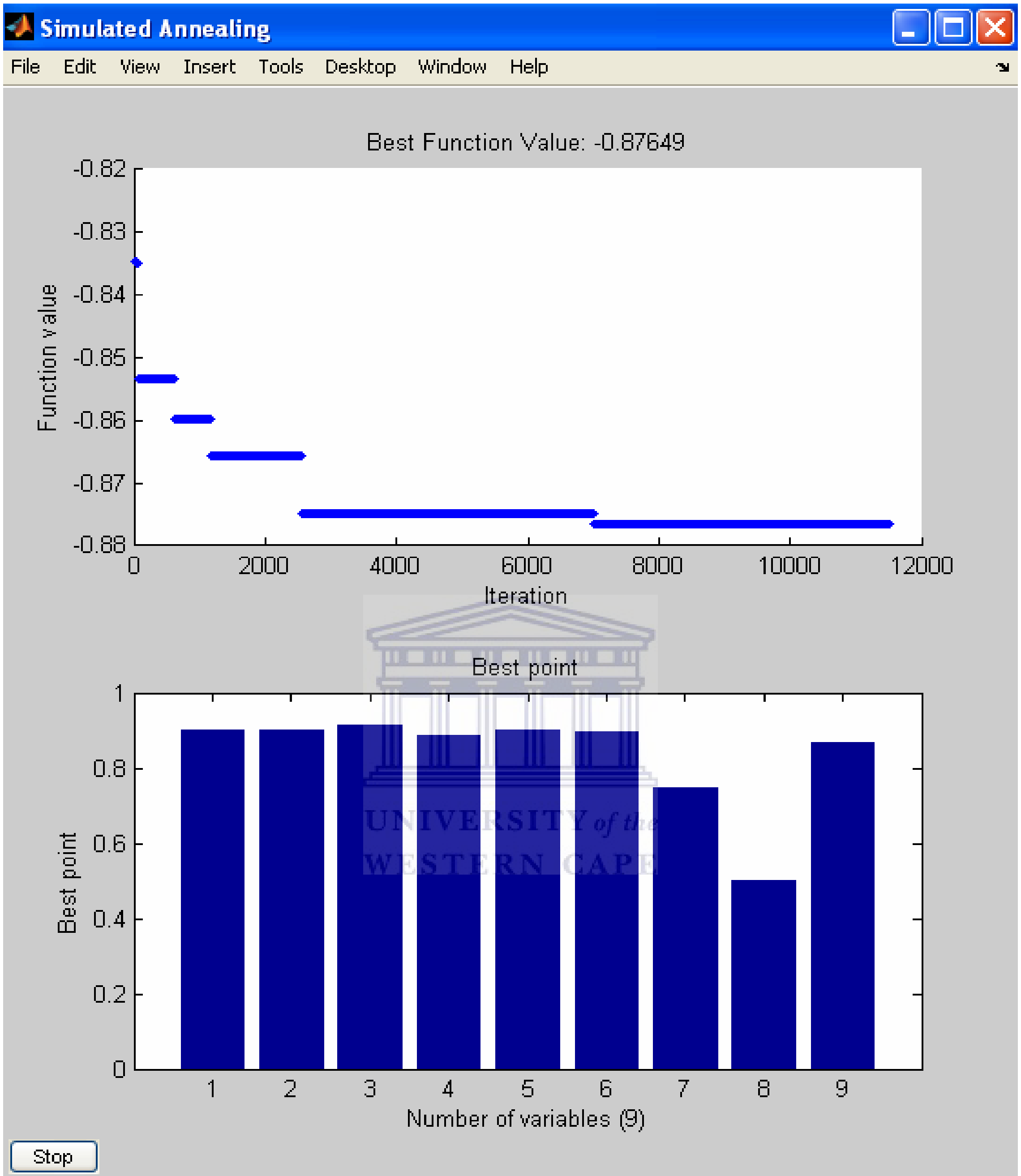
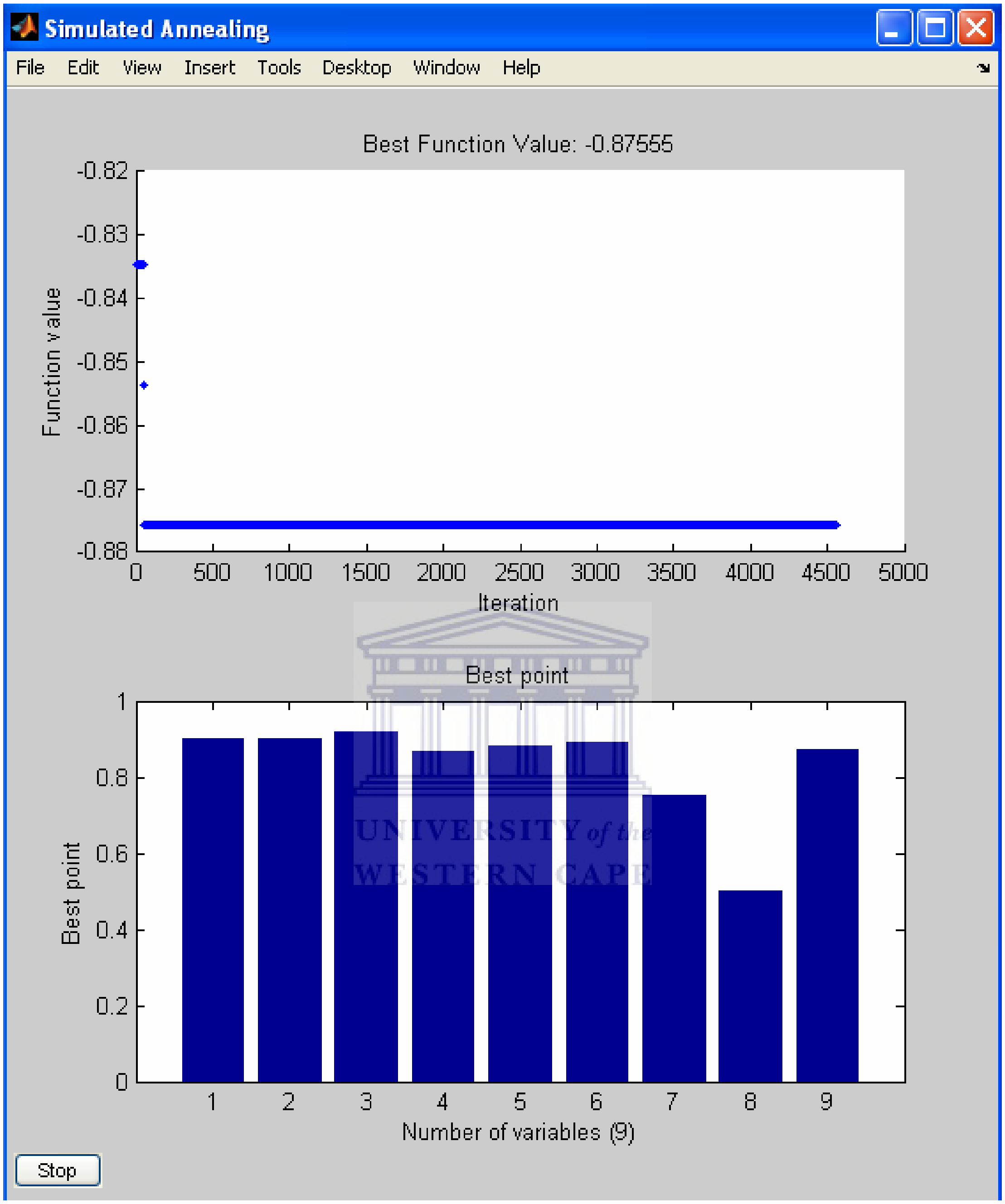


Figure A.1: *The Case 1 GA Best Fitness Value and Best Individual for WiMAX_3*


Simulated Annealing
File Edit View Insert Tools Desktop Window Help
Best Function Value: -0.87599
Function value
-0.82
-0.83
-0.84
-0.85
-0.86
-0.87
-0.88
0
1000
2000
3000
4000
5000
6000
Iteration
Best point
Best point
1
0.8
0.6
0.4
0.2
0
1
2
3
4
5
6
7
8
9
Number of variables (9)
Stop


Simulated Annealing
File Edit View Insert Tools Desktop Window Help
Best Function Value: -0.87649
Function value
-0.82
-0.83
-0.84
-0.85
-0.86
-0.87
-0.88
0
2000
4000
6000
8000
10000
12000
Iteration
Best point
Best point
1
0.8
0.6
0.4
0.2
0
1
2
3
4
5
6
7
8
9
Number of variables (9)
Stop


Simulated Annealing
File Edit View Insert Tools Desktop Window Help
Best Function Value: -0.87555
Function value
-0.82
-0.83
-0.84
-0.85
-0.86
-0.87
-0.88
0 500 1000 1500 2000 2500 3000 3500 4000 4500 5000
Iteration
Best point
Best point
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8 9
Number of variables (9)
Stop

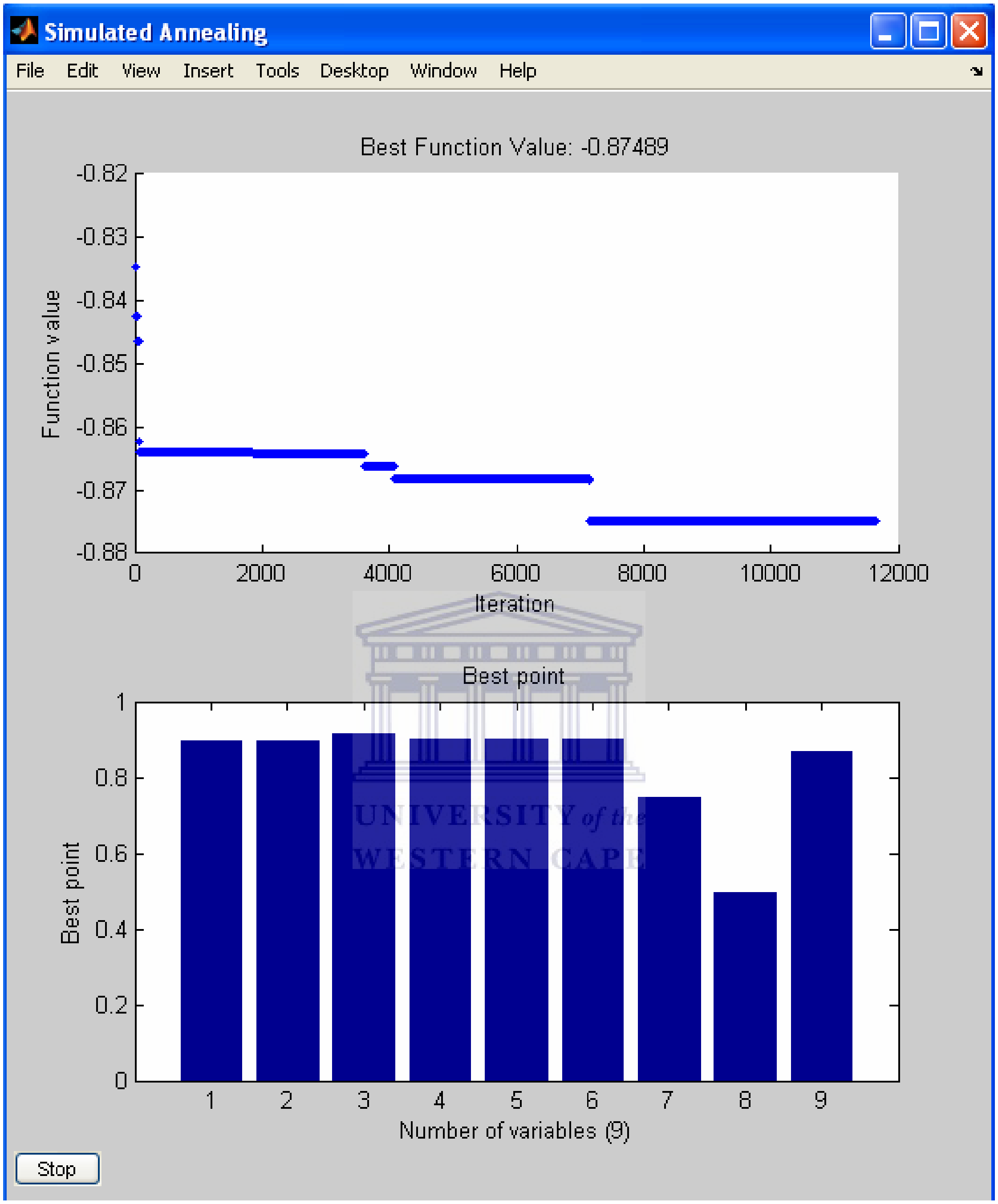
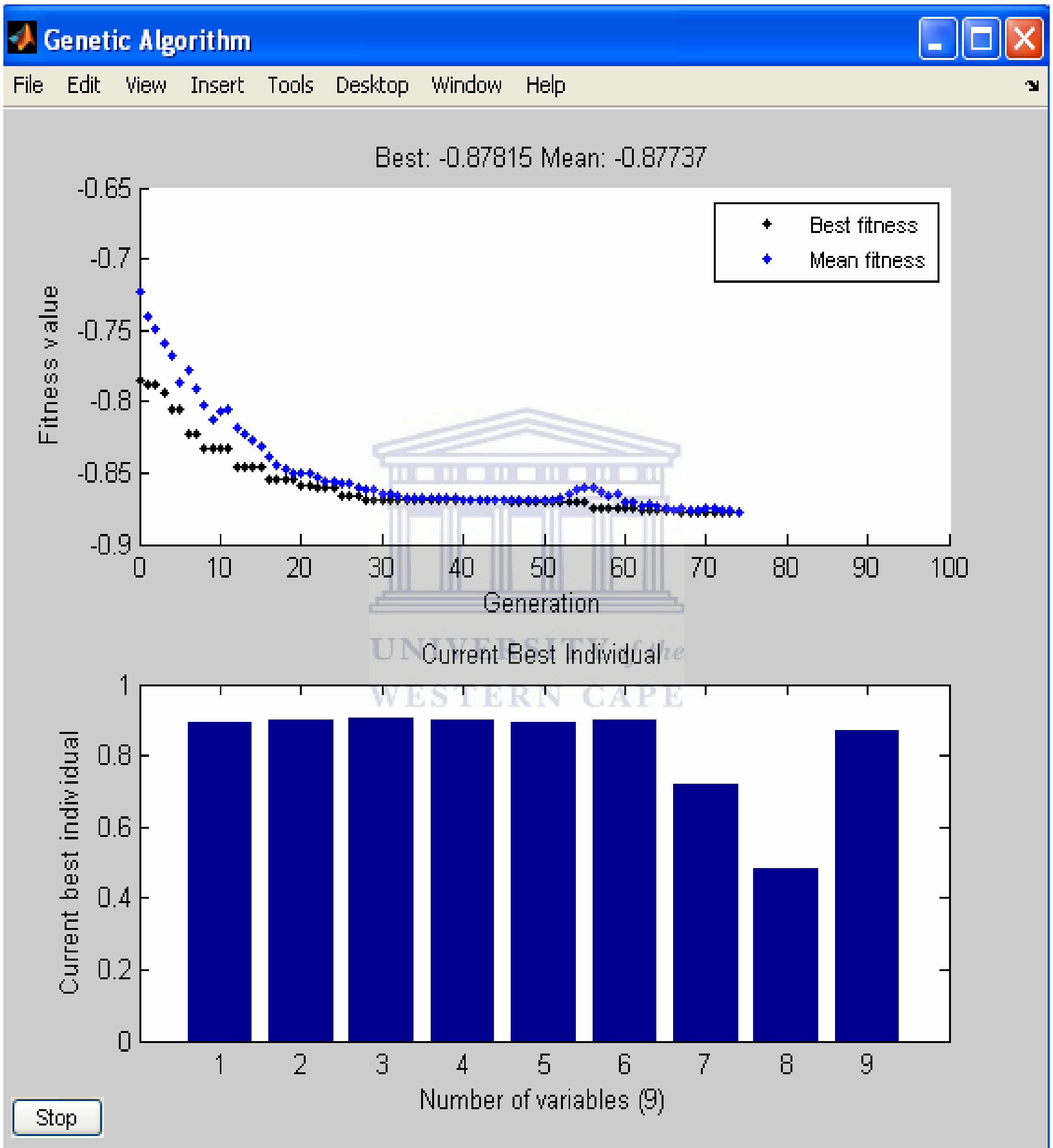
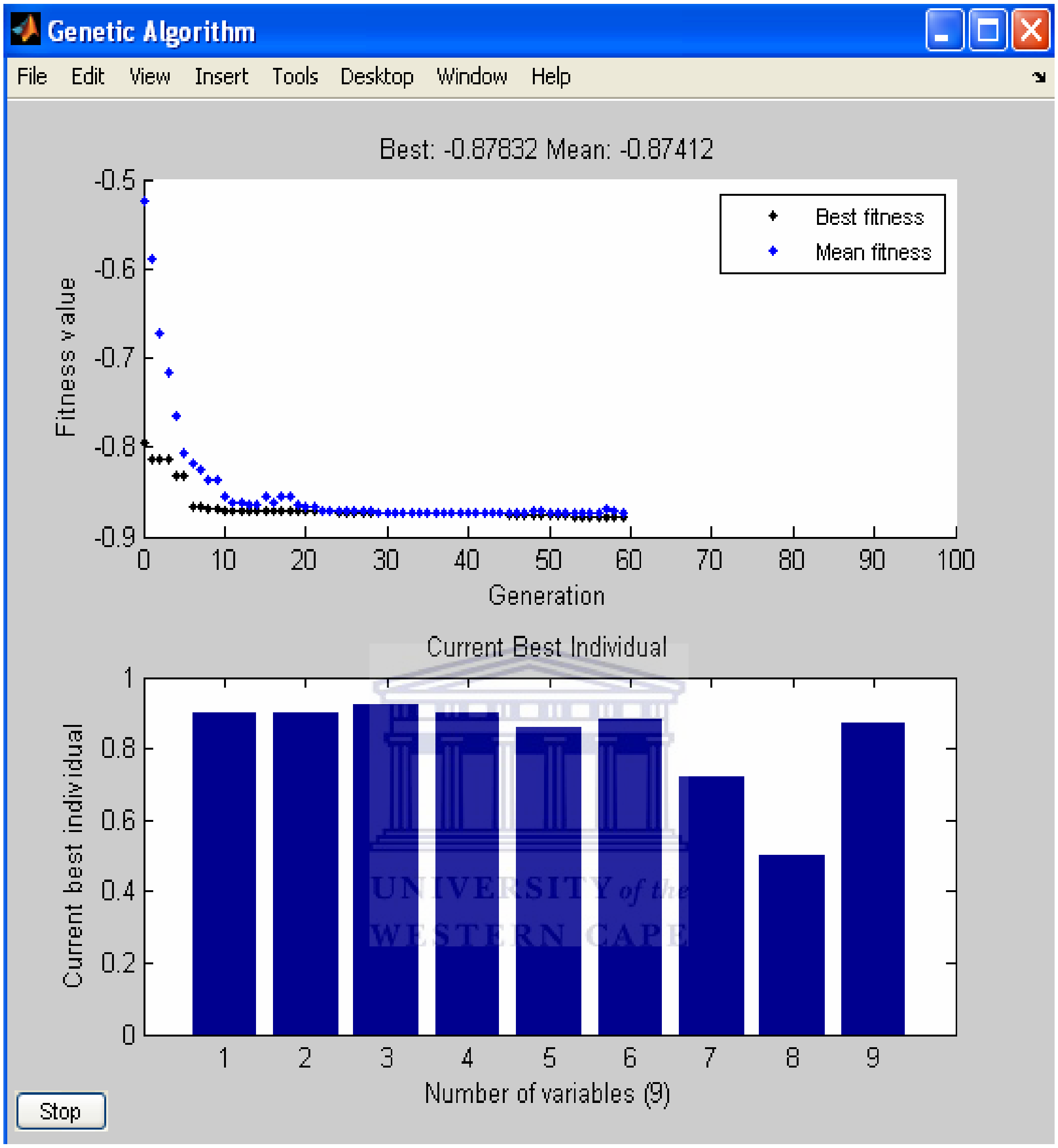
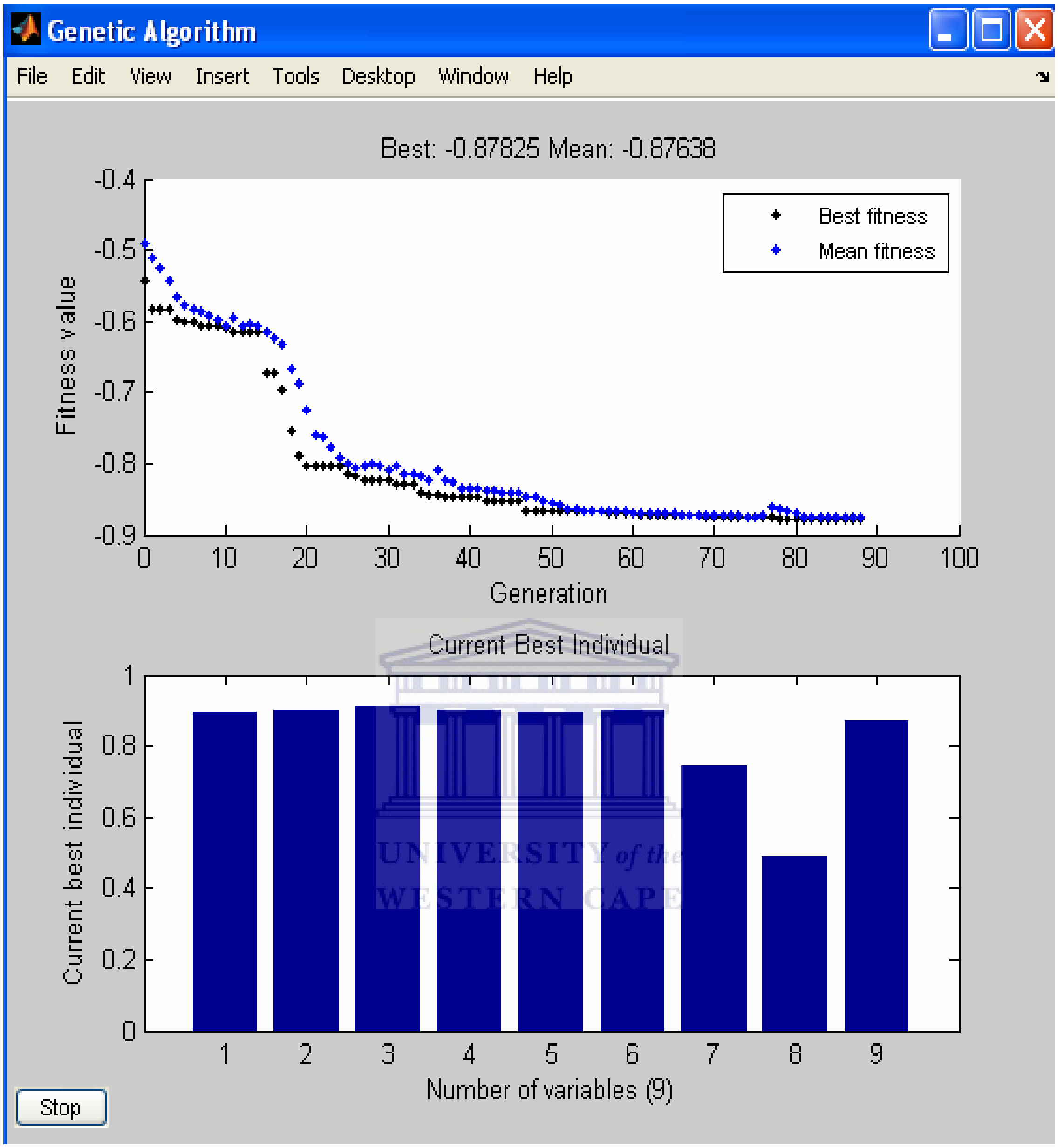


Figure A.2: *The Case 1 SA Best Fitness Value and Best Individual for WiMAX_3*

Case 2: Additional plots of the best fitness value and the best individual given by the GA and the SA for the WiMAX_3 network are shown in Figure A.3 and Figure A.4 respectively.




Genetic Algorithm
File Edit View Insert Tools Desktop Window Help
Best: -0.87832 Mean: -0.87412
Fitness value
Best fitness
Mean fitness
Generation
Current Best Individual
Current best individual
Number of variables (9)
Stop



Genetic Algorithm
File Edit View Insert Tools Desktop Window Help
Best: -0.87825 Mean: -0.87638
Best fitness
Mean fitness
Fitness value
Generation
Current Best Individual
Current best individual
Number of variables (9)
Stop

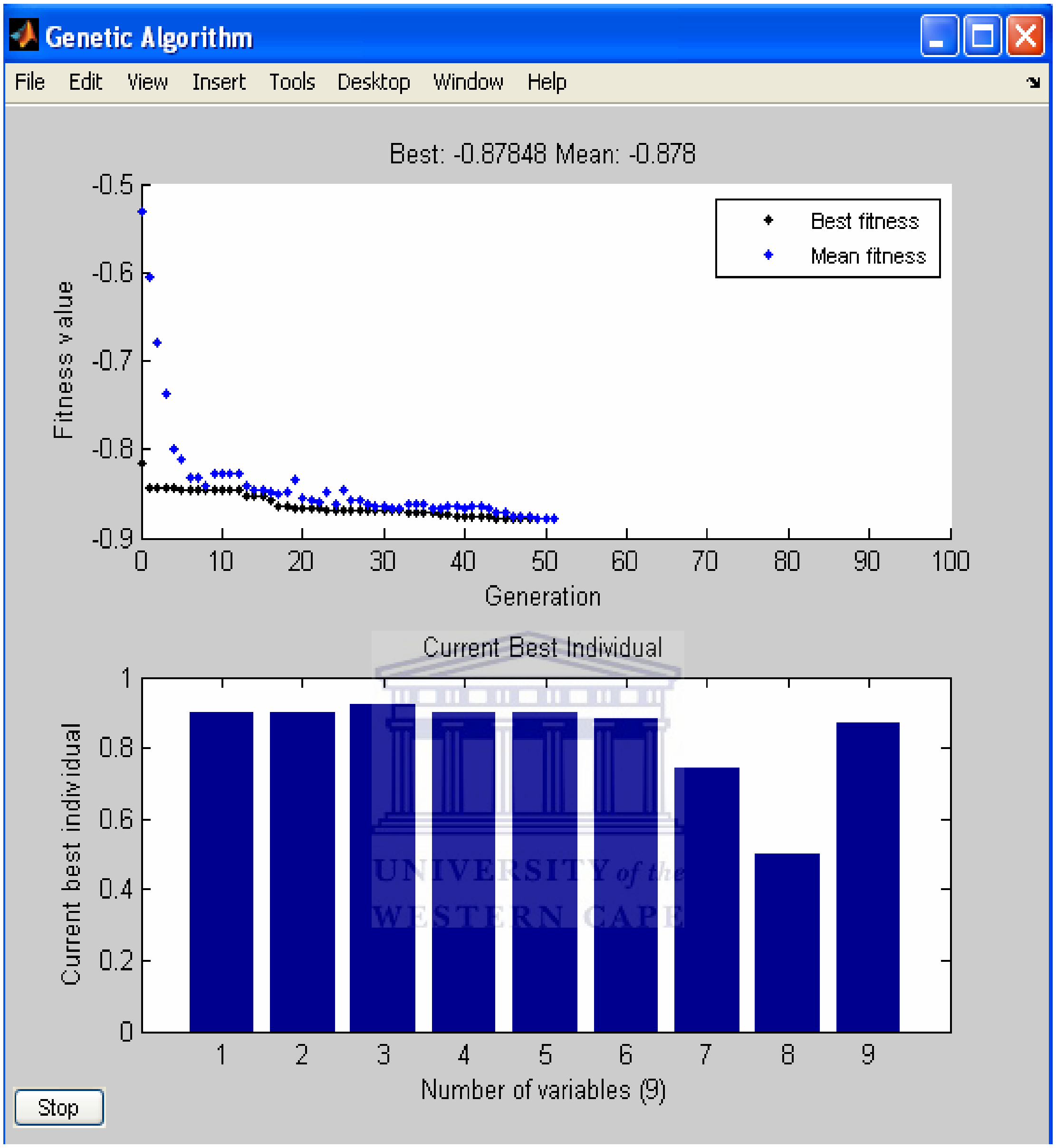
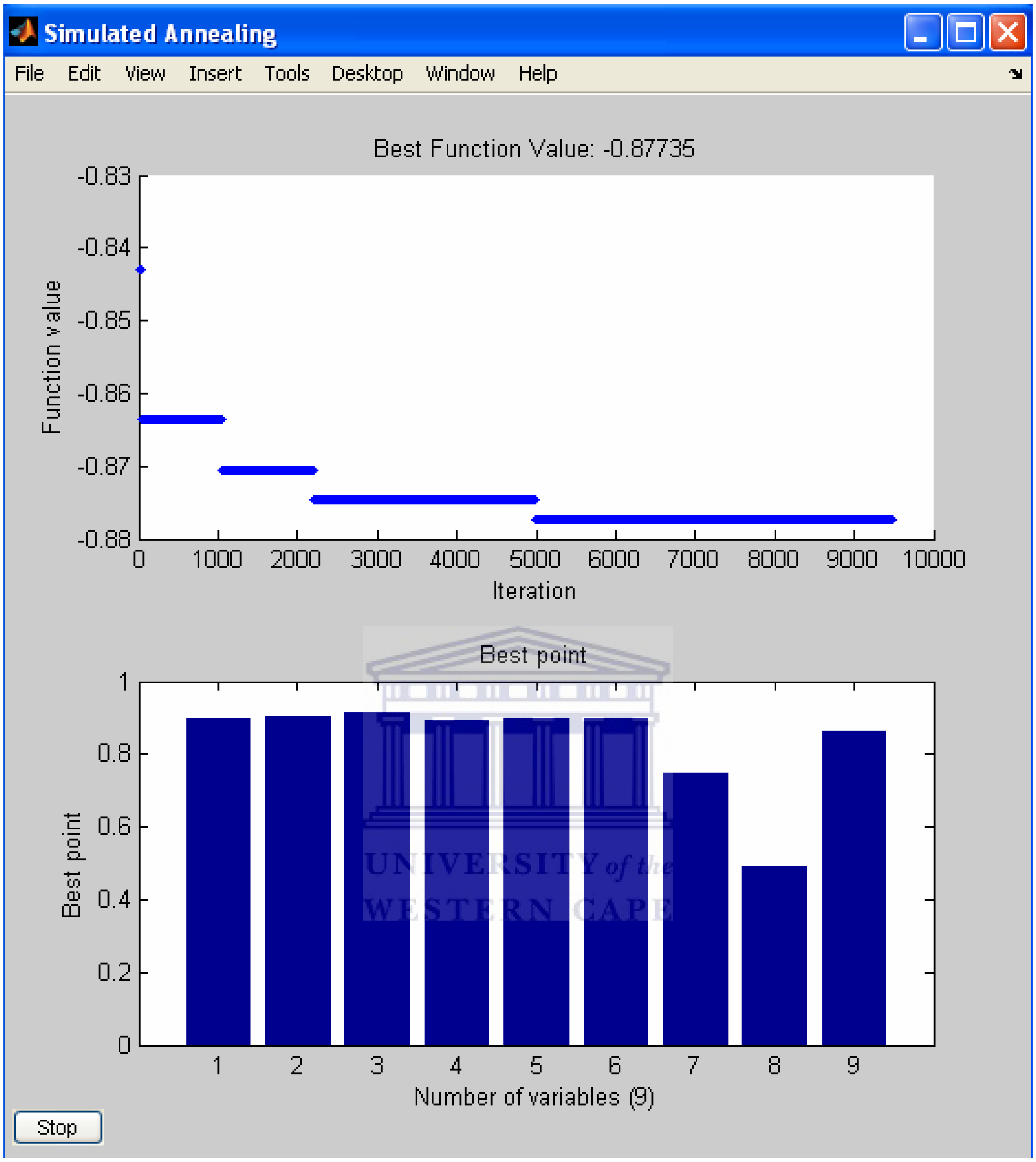
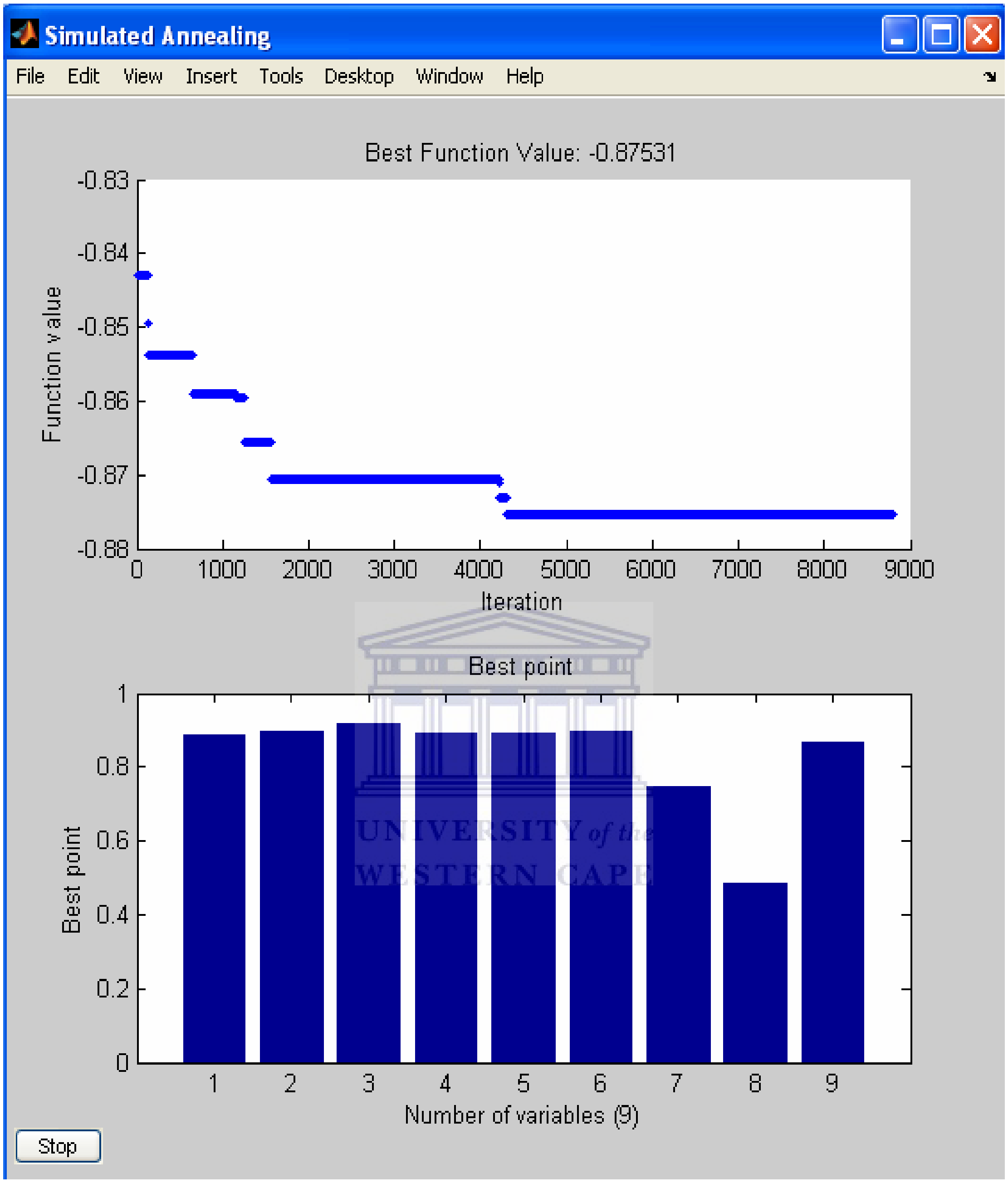
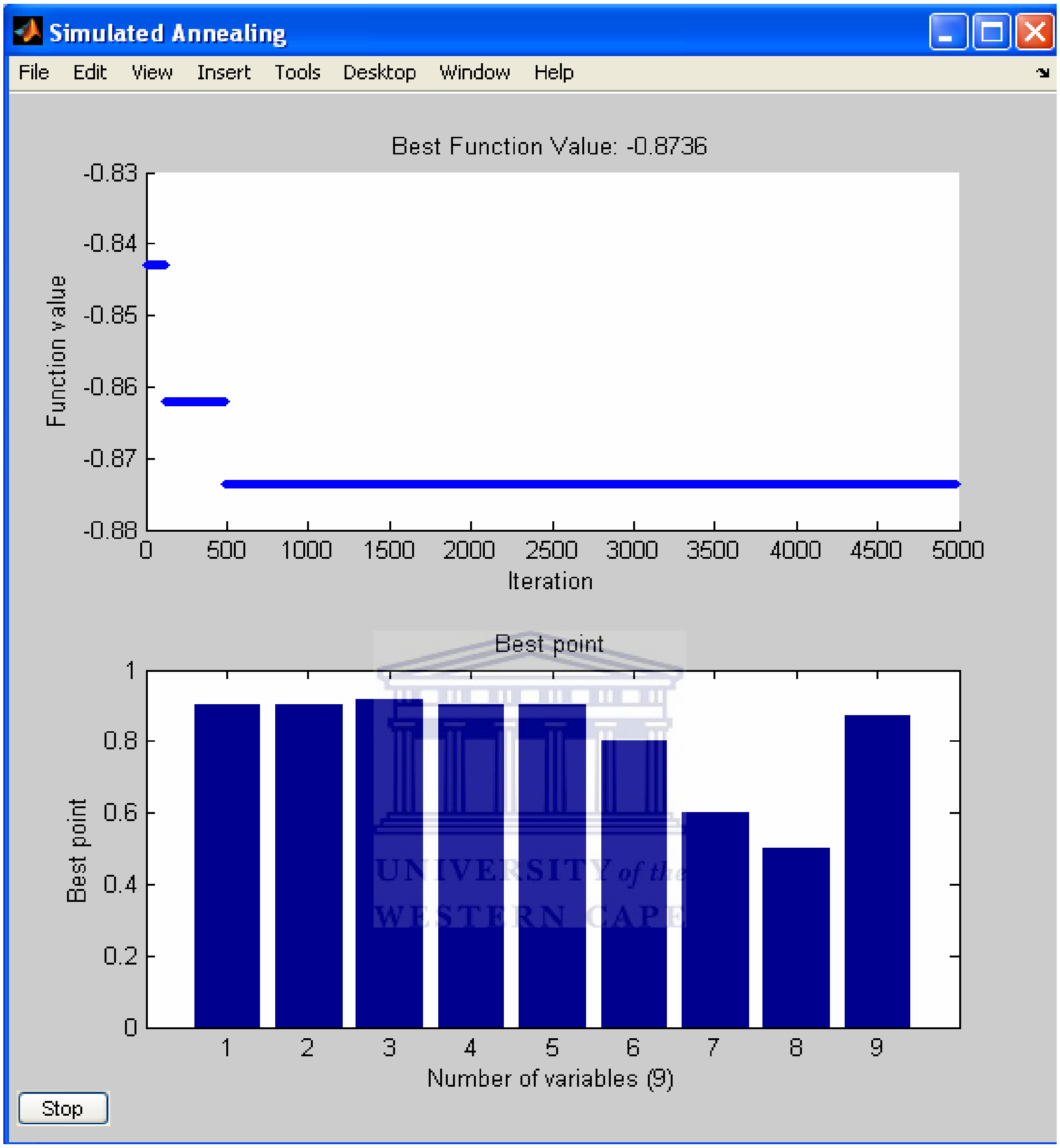


Figure A.3: *The Case 2 GA Best Fitness Value and Best Individual for WiMAX_3*


Simulated Annealing
File Edit View Insert Tools Desktop Window Help
Best Function Value: -0.87735
Function value
-0.83
-0.84
-0.85
-0.86
-0.87
-0.88
0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000
Iteration
Best point
Best point
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8 9
Number of variables (9)
Stop



Simulated Annealing
File Edit View Insert Tools Desktop Window Help
Best Function Value: -0.87531
Function value
-0.83
-0.84
-0.85
-0.86
-0.87
-0.88
0 1000 2000 3000 4000 5000 6000 7000 8000 9000
Iteration
Best point
Best point
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8 9
Number of variables (9)
Stop



Simulated Annealing
File Edit View Insert Tools Desktop Window Help
Best Function Value: -0.8736
Function value
-0.83
-0.84
-0.85
-0.86
-0.87
-0.88
0 500 1000 1500 2000 2500 3000 3500 4000 4500 5000
Iteration
Best point
Best point
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8 9
Number of variables (9)
Stop

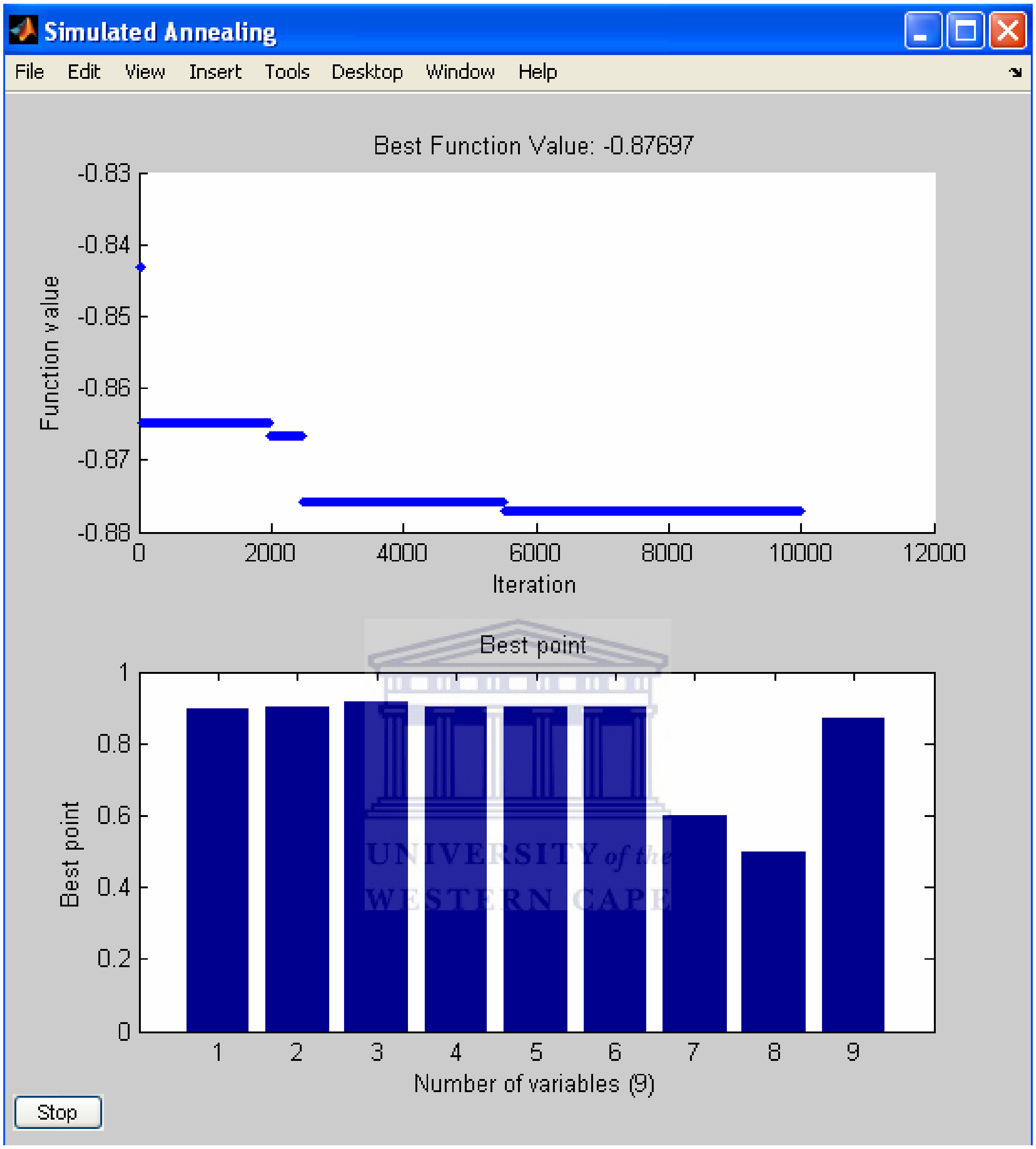


Figure A.4: *The Case 2 SA Best Fitness Value and Best Individual for WiMAX_3*

8 REFERENCES

[1] G. J. Garwood, “Work Manager”, *BT Technology Journal*, Vol. 15, No. 1, January 1997, pp. 58-68.

[2] Ericsson, “Field Force Automation Brochure”, October 2008, http://www.ericsson.com/au/solutions/utilities/images/ericsson_ffa_brochure_oct08.pdf

[3] Ericsson, “Lessons Learnt Implementing Field Force Automation at ENERGEX”, September 2008, http://www.ericsson.com/au/solutions/utilities/images/ericsson_energex_paper_nov08.pdf.

[4] D. Lesaint *et al.*, “Field Workforce Scheduling”, *BT Technology Journal*, Vol. 21, No. 4, October 2003, pp. 23-26.

[5] S. Y. Hui, and K. H. Yeung, “Challenges in the Migration to 4G Mobile Systems”, *IEEE Communications Magazine*, December 2003, pp. 54-59.

[6] B. Xie, A. Kumar, and D. P. Agrawal, “Enabling Multiservice on 3G and Beyond: Challenges and Future Diections”, *IEEE Wireless Communications,* June 2008, pp. 66-72.

[7] S. Frattasi, H. Fathi, F. H. P. Fitzek, and R. Prasad, “Defining 4G Technology from the User’s Perspective”, *IEEE Network*, January/February 2006, pp. 35-41.

[8] G. Lampropoulos, A. K. Sankintzis, and N. Passas, “Media-Independent Handover for Seamless Service Provision in Heterogeneous Networks”, *IEEE Communications Magazine*, January 2008, pp. 64-71.

[9] P. Taaghol, A. K. Salkintzis, and J. Iyer, “Seamless Integration of Mobile WiMAX in 3GPP Networks”, *IEEE Communications Magazine*, October 2008, pp. 74-85.

[10] C. Yiping and Y. Yuhang, “A New 4G Architecture Providing Multimode Terminals Always Best Connected Services”, *IEEE Wireless Communications,* April 2007, pp. 36-41.

[11] E. Gustafsson, and A. Jonsson, “Always Best Connected”, *IEEE Wireless Communications,* February 2003, pp. 49-55.

[12] G. Fodor, A. Eriksson, and A. Tuoriniemi, “Providing Quality of Service in Always Best Connected Networks”, *IEEE Communications Magazine*, July 2003, pp. 154-163.

[13] M. Stemm, and R. Katz, "Vertical Handoffs in Wireless Overlay Networks", *ACM Mobile Networking,* Special Issue on Mobile Networking in the Internet 3 (4), 1998, pp. 335-350.

[14] S. McCann, *et al.*, "Next Generation Multimode Terminals", http://www.roke.co.uk/download/papers/next_generation_multimode_terminals.pdf.

[15] A. A. Koutsorodi *et al.*, "Terminal Management and Intelligent Access Selection in Heterogeneous Environments", *Mobile Networks and Applications,* (2006) 11, pp. 861-871.

[16] Motorola, "Seamless Mobility", www.motorola.com/mot/doc/5/5764_MotDoc.pdf.

[17] N. Golmie, "Seamless Mobility: Are We There Yet?", *IEEE Wireless Communications,* August 2009, pp. 12-13.

[18] F. Siddiqui and S. Zeadally, "Mobility Management across Hybrid Wireless Networks: Trends and Challenges", *Computer Communications*, May 2006, pp. 1363-1385.

[19] I. F. Akyildiz, J. McNair, J. S. M. Ho, H. Uzunalioglu, and W. Wang, "Mobility Management In Next Generation Wireless Systems", *Proceedings of the IEEE*, Vol. 87, No. 8, August 1999, pp. 1347-1384.

[20] K. Pahlavan *et al.*, "Handoff in Hybrid Mobile Data Networks", *IEEE Personal Comm.*, April 2000, pp. 34-47.

[21] N. Nasser, A. Hasswa, and H. Hassanein, "Handoffs in Fourth Generation Heterogeneous Networks", *IEEE Communications Magazine*, October 2006, pp. 96-103.

[22] N. D. Tripathi *et al*., "Adaptive Handoff Algorithm for Cellular Overlay Systems Using Fuzzy Logic", *IEEE* 49^{th} *VTC*., May 1999, pp. 1413-1418.

[23] N. D. Tripathi, "Generic Adaptive Handoff Algorithms using Fuzzy Logic and Neural Networks", PhD Thesis, Virginia Polytechnic Institute and State University, 1997.

[24] F. Zhu and J. McNair, "Vertical Handoffs in Fourth-Generation Multinetwork Environments", *IEEE Wireless Communications,* June 2004, pp. 8-15.

[25] Draft IEEE Standard, "IEEE Standard for Local and Metropolitan Area Networks: Media Independent Handoff", Draft IEEE P802.21/D03.00, December 2006.

[26] M. Ylianttila *et al.*, "Optimization scheme for Mobile Users Performing Vertical Handoffs between IEEE 802.11 and GPRS/EDGE Networks", *Proc. of IEEE GLOBECOM'01*, San Antonio, Texas, USA, Nov 2001, pp. 3439-3443.

[27] A. H. Zahran, B. Liang, and A. Saleh, "Signal Threshold Adaptation for Vertical Handoff in Heterogeneous Wireless Networks", *Mobile Networks and Applications*, 11 (4) (2006), pp. 625-640.

[28] H. Wang *et al.*, "Policy-enabled Handoffs across Heterogeneous Wireless Networks", *Proc. of Mobile Comp. Sys. and Apps.,* New Orleans, LA, Feb 1999.

[29] R. Ribeiro, "Fuzzy Multiple Attribute Decision Making: A Review and New Preference Elicitation Techniques", *Fuzzy Sets and Systems*, 1996, pp. 155-181.

[30] C. Hwang and A. Masud, *Multiple Attribute Decision Making: Methods and Applications – A State-of-the-Art Survey*. Berlin: Springer-Verlag, 1981.

[31] J. L. Deng, "Introduction to Grey System Theory", *The Journal of Grey System*, Volume 1, Issue 1, 1989, pp. 1-24.

[32] E. Stevens-Navarro, and V.W.S. Wong, "Comparison between Vertical Handoff Decision Algorithms for Heterogeneous Wireless Networks", *Proceedings of IEEE Vehicular Technology Conference (VCT-06)*, Melbourne, Australia, 2006, pp. 947-951.

[33] Q. Song and A. Jamalipour, "Network Selection in an Integrated Wireless LAN and UMTS Environment using Mathematical Modeling and Computing Techniques", *IEEE Wireless Communications*, June 2005, pp. 42-48.

[34] D. Charilas *et al.*, "Packet-switched Network Selection with the Highest QoS in 4G Networks", *Computer Networks*, January 2008, pp. 248-258.

[35] P. M. L. Chan *et al.*, "Mobility Management Incorporating Fuzzy Logic for a Heterogeneous IP Environment", *IEEE Communications Magazine*, December 2001, pp. 42-51.

[36] J. Hou and D. C. O'Brien, "Vertical Handover Decision-Making Algorithm Using Fuzzy Logic for the Integrated Radio-and-OW System", *IEEE Transactions On Wireless Communications*, Vol. 5, No. 1, January 2006, pp. 176-185.

[37] Y. Nkansah-Gyekye and J. I. Agbinya, "Vertical Handoff Decision Algorithm for UMTS-WLAN", *Proc. of the 2nd International Conference on Wireless Broadband and Ultra Wideband Communications*, Sydney, Australia, August 2007.

[38] Q. Guo, J. Zhu, and X. Xu, An Adaptive Multi-criteria Vertical Handoff Decision Algorithm for Radio Heterogeneous Network", *Proceedings of IEEE International Conference on Communications*, 2005, pp. 2769-2773.

[39] Q. Wei, K. Farkas, C. Prehofer, P. Mendes, and B. Plattner, "Context-aware Handover Using Active Network Technology", *Computer Networks*, October 2006, pp. 2855-2872.

[40] M. Kassar, B. Kervalla, and G. Pujolle, "An Intelligent Handover Management System for Future Generation Wireless Networks", *EURASIP Journal on Wireless Communications and Networking*, Volume 2008, 12 pages.

[41] W. M. Eddy, "At What Layer Does Mobility Belong?", *IEEE Communications Magazine*, October 2004, pp. 155-159.

[42] J. W. Floroiu, R. Ruppelt, and D. Sisalem, "Seamless Handover in Terrestrial Radio Access Networks: A Case Study", *IEEE Communications Magazine*, November 2003, pp. 110-116.

[43] H. Schulzrinne, and E. Wedlund, "Application-Layer Mobility using SIP", *ACM Mobile Comp. and Commun. Rev.*, vol. 4, no. 3, July 2000, pp. 47-57.

[44] W. Wu *et al.*, "SIP-Based Vertical Handoff between WWANs and WLANs", *IEEE Wireless Communications*, June 2005, pp. 66-72.

[45] L. Budzisz *et al.*, "Towards Transport-layer Mobility: Evolution of SCTP Multihoming", *Computer Communications*, March 2008, pp. 980-998.

[46] R. Stewart, ed., "Stream Control Transmission Protocol", IETF RFC 4960, September 2007; http://tools.ietf.org/html/rfc4960.

[47] The OpenSS7 Project downloads at www.openss7.org/download.html.

[48] Y. Nkansah-Gyekye and J. I. Agbinya, "Vertical Handoff between WWAN and WLAN", *Proceedings of the 5th International Conference on Networking (ICN 2006)*, Mauritius, 23-28 April 2006.

[49] R. Stewart *et al.*, "SCTP Dynamic Address Reconfiguration", IETF RFC 5061, September 2007; http://tools.ietf.org/html/rfc5061.

[50] L. Ma, F. Yu, and V. C. M. Leung, "A New Method to Support UMTS/WLAN Vertical Handover using SCTP", *IEEE Wireless Communications,* Aug 2004, pp. 44-51.

[51] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice-Hall PTR, 1995.

[52] N. Baldo and M. Zorzi, "Fuzzy Logic for Cross-Layer Optimization in Cognitive Radio Networks", *IEEE Communications Magazine*, April 2008, pp. 64-71.

[53] L. A. Zadeh, "Fuzzy Sets", *Information and Control*, volume 8, 1965, pp. 338-353.

[54] J. M. Mendel, "Fuzzy Logic Systems for Engineering: A Tutorial", *IEEE Proceedings*, Volume 83, March 1995, pp. 345-377.

[55] J. M. Mendel, "Type-2 Fuzzy Sets and Systems", *IEEE Computational Intelligence Magazine*, February 2007, pp. 20-29.

[56] J. M. Mendel, R. I. John, and F. Liu, "Interval Type-2 Fuzzy Logic Systems Made Simple", *IEEE Transactions on Fuzzy Systems*, December 2006, pp. 808-821.

[57] L. A. Zadeh, "The Concept of a Linguistic Variable and Its Application to Approximate Reasoning", Parts 1 and 2, *Information Sciences*, 1975, **8**: pp. 199-249, pp. 301-357.

[58] J. Yen, "Fuzzy Logic – A Modern Perspective", *IEEE Transactions on Knowledge and Data Engineering*, January 1999, pp. 153-165.

[59] L. A. Zadeh, "A Fuzzy Set Theoretic Interpretation of Linguistic Hedges", *Journal of Cybernetics*, 1972, **2**(2): pp. 4-34.

[60] L. A. Zadeh, "Fuzzy Logic = Computing With Words", *IEEE Transactions on Fuzzy Systems*, May 1996, pp. 103-111.

[61] J-S. R. Jang and C-T. Sun, "Neuro-Fuzzy Modeling and Control", *Proceedings of the IEEE*, March 1995.

[62] T. J. Ross, *Fuzzy Logic With Engineering Applications*, McGraw-Hill, 1995.

[63] 3GPP TS 23.402, "3GPP System Architecture Evolution: Architecture Enhancements for Non-3GPP Accesses (Release 8)".

[64] S. J. Koh, M. J. Lee, M. Riegel, M. L. Ma, and M. Tuexen, "Mobile SCTP for Transport Layer Mobility", Internet Draft, draft-sjkoh-sctp-mobility-04.txt, June 2004.

[65] Y. Nkansah-Gyekye and J. I. Agbinya, "Vertical Handoff Decision Algorithm Based on Fuzzy Logic and Genetic Algorithm", *Proceedings of Southern African Telecommunication Networks & Applications Conference (SATNAC 2008)*, Wild Coast Sun, Eastern Cape, South Africa, 7-10 September 2008.

[66] Y. Nkansah-Gyekye and J. I. Agbinya, "An Optimum Vertical Handoff Decision Algorithm for UMTS-WiMAX", *Proceedings of the 2nd IFIP Symposium on Wireless Communications and Information Technology in Developing Countries (WCITD 2008)*, Pretoria, South Africa, 6-7 October 2008.

[67] Y. Nkansah-Gyekye and J. I. Agbinya, "A Vertical Handoff Decision Algorithm for Next Generation Wireless Networks", *Proceedings of the 3rd International Conference on Broadband Communications, Information Technology & Biomedical Applications (BroadCom 2008)*, Pretoria, South Africa, 23-26 November 2008.

[68] R. R. Yager and D. P. Filev, *Essentials of Fuzzy Modeling and Control*, John Wiley & Sons, 1994.

[69] J. M. C. Sousa and U. Kaymak, *Fuzzy Decision Making in Modeling and Control*, World Scientific, 2002.

[70] R. C. Eberhart and Y. Shi, *Computational Intelligence: Concepts to Implementations*, Morgan Kaufmann, 2007.

[71] MATLAB. http://www.mathworks.com.

[72] R. R. Yager, "Multiple Objective Decision Making using Fuzzy Sets", *International Journal of Man-Machine Studies*, Vol. 9, 1977, pp. 375-382.

[73] H.-J. Zimmermann, "Fuzzy Set Theory and Its Applications", 4th Edition, Boston/Dordrecht/London: Kluwer Academic Publishers, 2001.

[74] J. Lu, G. Zhang, D. Ruan, and F. Wu, *Multi-Objective Group Decision Making*. London, Imperial College Press, 2007.

[75] R. E. Bellman and L. A. Zadeh, "Decision-making in a Fuzzy Environment", *Management Sci.*, 1970, pp. 141-164.

[76] R. R. Yager, "Fuzzy Decision Making Including Unequal Objectives", *Fuzzy Sets and Systems*, 1 (1978), pp. 87-95.

[77] Y. Nkansah-Gyekye and J. I. Agbinya, "Vertical Handoff Decision Using Fuzzy Logic', *Proceedings of Southern African Telecommunication Networks & Applications Conference* (*SATNAC 2006)*, Western Cape, South Africa, 3-6 September 2006.

[78] J. J. More, R. A. Tapia, and M. H. Wright, "Optimization", *ACM SIGNUM Newsletter*, Volume 15, Issue 4, December 1980, pp. 16-26.

[79] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[80] C. Blum and A. Roli, "Metaheuristics in Combinatorial Optimization: Overview and Conceptual Comparison", *ACM Computing Surveys*, Vol. 35, No. 3, September 2003, pp. 268-308.

[81] Metaheuristics Network Website. http://www.metaheuristics.org/. Visited in June 2008.

[82] D. Fogel, *Evolutionary Computation*, IEEE Press, Piscataway, NJ, 1995.

[83] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press, 1992.

[84] F. Glover, "Tabu Search – Part I", *ORSA Journal on Computing* **1**(3), pp. 190-206.

[85] F. Glover, "Tabu Search – Part II", *ORSA Journal on Computing* **2**(1), pp. 4-32.

[86] J. Kennedy, R. C. Eberhart, and Y. Shi, *Swarm Intelligence*, Morgan Kaufmann Publishers, 2001.

[87] A. P. Engelbrecht, *Fundamentals of Computational Swarm Intelligence*, John Wiley & Sons, 2005.

[88] M. Dorigo and T. Stutzle, *Ant Colony Optimization*, MIT Press, Cambridge, MA, 2004.

[89] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning,* Addison-Wesley, Reading, Massachusetts, 1989.

[90] D. E. Goldberg, *Design of Innovation: Lessons from and for Competent Genetic Algorithms*, Kluwer Academic Publishers, Boston, Massachusetts, 2002.

[91] J. H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI, 1975.

[92] K. Sastry, D. E. Goldberg, and G. Kendall, "Genetic Algorithms: A Tutorial". In E. Burke, and G. Kendall (Eds.), *Introductory Tutorials in Optimization, Search, and Decision Support Methodologies* (Chapter 4, pp. 97-125), Berlin: Springer, 2005.

[93] D. E. Goldberg and K. Deb, "A Comparative Analysis of Selection Schemes Used in Genetic Algorithms". In *Foundations of Genetic Algorithms*, G. J. E. Rawlins (Ed.), Morgan Kaufmann, 1991, pp. 69-93.

[94] G. Syswerda, "Uniform Crossover in Genetic Algorithms". In J. D. Schaffer, Ed. *Proceedings of the Third International Conference on Genetic Algorithms*, San Mateo, CA: Morgan Kaufmann.

[95] R. L. Haupt and S. E. Haupt, *Practical Genetic Algorithms*, Second Edition, John Wiley & Sons, New Jersey, 2004.

[96] S. Kirkpatrick, C. D. Gelatt, Jr, and M. P. Vecchi, "Optimization by Simulated Annealing", *Science*, Volume 22, pp. 671-680, 1983.

[97] V. Cerny, "Thermodynamic Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm", *Journal of Optimization Theory and Applications*, Vol. 45, No. 1, January 1985, pp. 41-51.

[98] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines", *Journal of Chemical Physics*, Vol. 21, No. 6, 1953, pp. 1087-1092.

[99] S. Kirkpatrick, "Optimization by Simulated Annealing: Quantitative Studies", *Journal of Statistical Physics*, Volume 34, 1984, pp. 975-986.

[100] P. Persson, "Annealing Based Optimization Methods for Signal Processing Applications", PhD Thesis, Blekinge Institute of Technology, Sweden, 2003.

[101] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721-741, November 1984.

[102] L. Ingber, "Simulated Annealing: Practice versus Theory", *Mathematical and Computational Modeling*, vol.18, pp. 29-57, 1993.

[103] P. Salamon, P. Sibani, and R. Frost, "Facts, Conjectures, and Improvements for Simulated Annealing", Society for Industrial and Applied Mathematics, 2002, ser. Monographs on Mathematical Modeling and Computation, no. 7.