Mobility Prediction and Multicasting in Wireless Networks: Performance and Analysis

Suresh Venkatachalaiah

Doctor of Philosophy

2006

RMIT

Mobility Prediction and Multicasting in Wireless Networks: Performance and Analysis

by

Suresh Venkatachalaiah

B.Eng, M.Eng

Dissertation submitted in accordance with

the requirements for the Degree of

Doctor of Philosophy

School of Electrical and Computer Engineering

Science, Engineering and Technology Portfolio

RMIT University

Melbourne, Victoria, Australia

May 2006

Declaration

To the best of my knowledge and belief, this thesis contains no material previously published or written by any other person, except where due reference is made in the text of the thesis. This thesis has not been submitted previously, in whole or in part, to qualify for any other academic award. The content of the thesis is the result of the work, which has been carried out since the official commencement date of the approved research program.

Signed: .... Date: ....

Suresh Venkatachalaiah

Mobility Prediction and Multicasting in Wireless Networks: Performance and Analysis

To my Grandparents

Acknowledgements

It has been a long journey with many ups and downs. I thank God for giving me the opportunity to undertake this research program. Thankfully, on this journey, I have received so much help and support from many wonderful people to whom I would like to express my deepest gratitude.

First of all, I am deeply indebted to Prof. Richard J. Harris at Massey University, New Zealand (Formerly at the RMIT CATT Centre) for supervising me. Without his guidance and encouragement, I would not have been able to get through all the difficult times. Without his vision, deep insight, advice, and willingness to support, this work would not have been possible. His extensive knowledge, strong analytical skills, and commitment to excellence in research and teaching are truly treasures to his students. He gives students the freedom to explore uncharted territory while providing any necessary assistance at the right time. He is willing to share his knowledge and career experience and give emotional and moral encouragement. He is more than an adviser and a teacher but a role model and a friend. Working with him has proven to be an enjoyable and rewarding experience. He has been the key influence in my work and a father-figure responsible for the completion of my thesis and without his advice and expert knowledge my ideas would not have the necessary direction.

I should also acknowledge the financial support provided by the School of Computer and Electrical Engineering for my graduate studies. I am also thankful to have been awarded scholarships from the Australian Telecommunication Cooperative Research Centre (ATcrc), and an RPC Grant from Microsoft Corporation Inc., Redmond, USA. Mr. David Jones from the School deserves special thanks for his efforts in obtaining the Microsoft grant.

The CATT centre people ("CATTers") were always a source of inspiration and invaluable help to me, even though they themselves may not know it. They showed me how much one could achieve by hard work and dedication. Especially, I would like to thank Robert Suryasaputra, for his encouragement, helpful advice, his contribution and patience in my thesis and for being a wonderful friend and colleague. I also appreciated the encouragement from Dr. John Murphy during the "confusing time" early in my candidature. I would like to thank Andrew Cassin for keeping my computer in an excellent state, initially; and Robert who took care of my computer during the final stages of my research. I would also like to thank Suyong Eum and Rebecca Wang for making my time so much more enjoyable through their encouragement, friendship and support. I am grateful to Dr. Alexander Kist for his help in the final stages of my work. I am also thankful to Dr. Jidong Wang for his help in accepting the offer to be my second supervisor at very short notice.

Throughout my time in Melbourne, I was blessed with a great deal of emotional support from family and friends both in India and Australia. I must thank my parents, my brother-in-law Amarnath Sajapuram, my sister Savitha Amarnath, my brother Sridhar Venkatachalaiah and his wife Meera, my nieces Adithi and Nishitha, for the wonderful environment that they made for me and especially my mom, aunts (Dala aunty and Dachana aunty) and uncles for always believing in me in such a passionate way. Lastly, my cousin Amulya for being such a good friend to me and believing in my capability. The last year of my candidature would have been more difficult for me without the love and the unconditional support of my family and friends back home. Last, but not least, I would like to thank my friends

Shiva, Rashmi, Guru, Preeti, Birju, Sunil, Madhuri, Sunny, Sushmita, Sushant Goel, Suresh Matta, Jaggi, Pamma for their friendship, care, support and hospitality during my time in Melbourne.

Contents

List of Figures

List of Tables

Abstract

Handoff is a call handling mechanism that is invoked when a mobile node moves from one cell to another. Such movement may lead to a degradation in performance for wireless networks as a result of packet losses. A promising technique proposed in this thesis is to apply multicasting techniques aided by mobility prediction in order to improve handoff performance. Superior handover processing techniques are vital in meeting the QoS requirements of future mobile communication services. Some proposed methods in the literature include the use of signal strength thresholds and hysteresis to make handoff decisions. However, it is difficult with these methods to achieve efficient handoff due to a number of associated problems. In this thesis, we present a method that uses a Grey model for mobility prediction and a fuzzy logic controller that has been fine-tuned using evolutionary algorithms in order to improve prediction accuracy. We also compare the self-tuning algorithm with two evolutionary algorithms in terms of accuracy and their convergence times.

In this thesis, we also shall examine the issues that surround handover strategies in wireless networks. Our proposed method takes into account signal strengths from the base stations and predicts the signal strength of the next candidate base station in order to provide improved handover performance. The primary decision for mobility prediction is the accurate prediction of signal strengths obtained from the base stations and remove any unwanted errors in the prediction using suitable optimisation techniques. Furthermore, the model includes the procedures of fine-tuning the predicted data using fuzzy parameters. We also propose suitable multicasting algorithms to minimise the reservation of overall network resource requirements during handoff with the mobility prediction information. The considered problem is formulated as an optimisation problem depending on application requirements. To be able to efficiently solve the problem, the situation is modelled using a multicast tree that is defined to maintain connectivity with the mobile node, whilst ensuring bandwidth guarantees and a minimum hop-count. In this approach, we have tried to solve the problem by balancing two objectives through putting a weight on each of two costs. Simulation studies that have been performed show a significant performance improvement when the proposed scheme is employed.

We provide a detailed description of an algorithm to implement join and prune mechanisms, which will help to build an optimal multicast tree with QoS requirements during handoff as well as incorporating dynamic changes in the positions of mobile nodes. An analysis of how mobility prediction helps in the selection of potential Access Routers (AR) with QoS requirements – which affects the multicast group size and bandwidth cost of the multicast tree – is presented. The proposed technique tries to minimise the number of multicast tree join and prune operations. We have examined the performance of our algorithm using simulations under various environmental conditions and have obtained good performance results. Our results show that the expected size of the multicast group increases linearly with an increase in the number of selected destination AR's for multicast during handoff. We observe that the expected number of joins and prunes from the multicast tree increases with group size. Thus, for an increasing number of destinations, the estimated cost of the multicast tree in a cellular network also increases.

A special simulation model was developed to demonstrate both homogeneous and heterogeneous handoff which is an emerging requirement for fourth generation mobile networks. The model incorporates our mobility prediction model for heterogeneous handoff between the Wireless LAN and a cellular network. Solutions have been obtained for each of these problems with the development of efficient algorithms and techniques. The results presented in this thesis for mobility prediction, multicasting techniques and heterogeneous handoff include proposed algorithms and models which aid in the understanding, analysing and reducing of overheads during handoff.

Chapter 1

Introduction

1.1 Introduction

Based on the high growth of mobile and wireless networks, user expectations are driving today's world has high demands for QoS in wireless networks. All that is needed is "anytime, anywhere" access to information. Future Generation cellular networks are going to be the catalyst for a whole new set of mobile services enabling users to access advanced services anywhere and at anytime. We shall be free from the confines of cables, fixed access points and low speed connections. Next generation networks introduce wide-band radio communication with access speeds of up to 2 Mbits /sec. Compared with today's mobile networks, next generation networks will significantly boost network capability. So, operators will be able to support more users as well as offer more sophisticated services to their customers. It aims at high-speed data, superior quality voice and location based services. The objectives of next generation cellular networks will be to efficiently support real-time services for mobile users whilst ensuring that delays in the signalling and bearer traffic should be minimal. Global roaming is one of the design objectives for the next generation of cellular networks for which host mobility is very essential.

The aim of this research is to develop methods to minimise handover failures in future mobile communication networks. Before we embark on a detailed description of our approach to addressing the problems of minimising handover failure in cellular wireless networks, we shall present some background on future mobile communication networks and describe ways of using mobility prediction for improving handover - thus saving on network resources by using multicasting techniques. We move on to explain the methodologies used to improve handover performance in wireless networks. The performance of our proposed strategies are evaluated using analytical models supported by numerical results.

1.2 Handover in Cellular Networks

The use of cellular systems is a popular means of enhancing the capacity of mobile networks, as it is well known that spectrum is a limited resource. The benefit is that users can move while communicating, i.e., while they are mobile. In order to implement cellular technologies, we need to provide handoff capabilities to maintain link quality and reduce interference in the system during a call. Whenever the relative signal strength of a given base station rises above a threshold signal level, handoff is triggered between appropriate base stations. The radio link is transferred from the current channel to another channel in the other cell that can provide better quality. This kind of handoff is known as inter-cell handoff. As cells are made smaller and smaller in order to increase the capacity of the mobile network it is found that this increases the number of inter-cell handoffs. Handoff algorithms have been using bit-error rates (BER's) and Received Signal Strength Indicator (RSSI) as indicators for deciding whether to handover a call.

Several handoff criteria published in the literature are based on the strongest received signal from the base station and a hysteresis margin that allows the user to perform handoff only if the new base station has a greater signal strength than the current one [1]. While hysteresis reduces the frequency of unnecessary handoff, it also increases the decision delay. Also, many algorithms have been proposed that use fuzzy logic based approaches. In this thesis, we propose a technique, which uses the Grey prediction model to predict future received signal strength and then we shall fine-tune any errors using fuzzy logic and evolutionary algorithms. The parameters considered in this thesis utilise the RSSI values from the base station. The Grey system was devised in 1982 and was used for systems that have very little data from which to perform an analysis or predict future data [2].

1.3 Improving Handover Performance

The number of mobile users has increased over recent years. Traditionally, the first generation of wireless networks was targeted to include voice and some simple data communications operating at low data rates. Recently, with the advent of second and third generation networks we have seen the inclusion of broadband features. These improvements reinforce the use of the cellular concept as it provides good spectrum efficiency. This spectrum efficiency is obtained by dividing the coverage area into small cell areas. More cells meant that there would be more handovers that needed to be handled in an appropriate way. To effectively have good handoff performance, mobility prediction of signal strengths will play an important role here.

Previous analysis of communication networks has shown that handoff performance can be improved by accurate mobility prediction in both cellular networks as well as ad hoc networks using resource reservation. In other words, the amount of resources allocated to support traffic in a particular area has to be increased to improve handoff performance. Due to the finite nature of radio resources, this leads to a reduction in the customer base that can be supported by the network provider

- ultimately leading to an increase in tariffs for mobile communication networks. Therefore, other methods must be sought such as mobility prediction to predict the future base station to which a mobile user is likely to move so that we can be ready to provide the resources required by the user.

It is widely accepted that it is more important to support an ongoing call than accepting a new user (call) into the system. This suggests the implementation of a scheme which specifies priority for handover users versus new users. Various strategies have been proposed in the past to achieve this. The allocation of a certain proportion of the channels exclusively to handover users has been the most common and straightforward approach [3]. The main drawback of this strategy is that it increases the chances of a new call being rejected and the blocking probability may increase. Therefore, the approach has only minimal system utilisation. Furthermore, it is difficult to determine the exact requirement for reserved handoff channels.

In this research, we try to optimise handover performance by continuously predicting signal strengths and determine the conditions for determining when handover is to take place and where the correct base station is located. In the current urban cellular environment, handover traffic is typically generated by mobile users travelling in vehicles. The higher speeds of these users can cause them to traverse through more than one cell during the average length of a mobile call leading to a requirement for one or more handovers. Furthermore, if we know the direction of travel then it is easy to predict the next cell based on signal strengths – even with slow fading. This leads to the creation of a prediction model, which is one of the main aims of this research, that can provide insight into the future movement of a mobile user. Parameters used to characterise the users' mobility in the cells provide the input for the mobility prediction model.

1.4 Focus of this Thesis

The main research focus of this work was to study the advantages offered by accurate mobility prediction in wireless mobile networks and their application to improve handoff performance using a combination of mobility prediction and multicasting techniques. Our investigations were focused on ascertaining the suitability of the proposed mobility prediction scheme for the development of next generation mobile networks. In this thesis, we propose a novel framework for mobility prediction that takes into account new technologies that can provide QoS in future next generation wireless networks. The proposed method incorporates procedures that allow for accurate prediction of signal strengths that will help in improving handoff performance. Assuming that the mobile node is moving at a constant speed, the prediction methodology is able to predict the cell that will be used next by the mobile user and to have resources reserved and ready for a handoff.

To provide a foundation for the solution of the accurate prediction of the signal strengths from the base station, we first survey and categorise the prediction schemes currently proposed and deployed and identify their implications to handoff performance in wireless networks. In this way, we define the set of prediction parameters employed in the development of the prediction schemes. In order to predict the signal strengths we need a prediction model that has good prediction accuracy. We have selected the Grey model, that has previously been used for weather predictions and predicting parameters in control systems, in order to solve our prediction problem. To compare the performance of the Grey model, we considered a number of well-known approaches including weighted moving averages and exponential moving averages. The Grey model proved to be much more accurate than some of these traditional estimation techniques.

The Grey model, although it has good prediction accuracy, still has some errors which can be improved upon. Our approach to solving this problem is to use a hybrid technique that involves fuzzy IF-THEN inference rules and evolutionary algorithms. By using fuzzy IF-THEN rules, we formulate the decision parameters required to reduce the error in our prediction of signal strength. More precisely, we have used evolutionary algorithms such as the well-known Genetic Algorithm (GA) and Particle Swarm Optimisation (PSO) methods as our main optimisation techniques to fine-tune the decision parameters of the fuzzy rules. After treating the errors, a comparison is made of the compensated models from the two evolutionary algorithms and a self-tuning algorithm. These comparisons are made based on their convergence to a specified fitness function. All the procedures are implemented in order to reduce the errors in signal strength and to increase the prediction accuracy of the model.

To further improve the QoS factors during handoff, we propose to use multicasting techniques to construct a minimum cost spanning tree from a source to certain prescribed destinations using the mobility prediction algorithm. Firstly, we formulate a problem to simultaneously solve two possible objectives, viz: optimising the residual bandwidth or determining the minimum number of hops required to establish the connection between the source and the mobile node. Thus, there could be two possible alternative solutions, one for minimum cost and the other for minimum hops. Secondly, solving the minimum cost and minimum hop count problems gives two extreme solutions which may conflict. As a result of this potential conflict, we propose an approach to find a "near optimal tree" with respect to these two objectives by exploring all the other trees "in-between" these two extreme solutions. Ultimately, we propose a weighted average of these two objectives as the solution in order to reflect the best solution according to the type of application. To solve the problem, we propose two algorithms that will be referred to as the MMP algorithm and the K-Minhop algorithms respectively.

A second model goes a step further with the parameters like delay and residual bandwidth to reduce the number of joins and minimise the number of prunes in the multicast tree. An analysis of how mobility prediction helps in the selection of potential Access Routers (AR) with QoS requirements that affect the multicast group size and bandwidth cost of the multicast tree, is presented. The proposed technique tries to minimise the number of multicast tree join and prune operations. Here, the optimal tree is chosen to meet a bandwidth objective and to satisfy a delay constraint. It has been used to measure parameters like the number of unnecessary joins and the number of prunes to the multicast tree. An algorithm called the MBWDC (Multicast BandWidth Delay Constraint) Algorithm is proposed. The proposed algorithm also takes care of the QoS constraints involved during the join operation. However, we propose a solution that reduces overheads by performing accurate mobility prediction which can select a potential AR. Advance knowledge of the new AR can be used to initiate a mechanism for avoiding packet losses and accomplishing a new handover (with low packet losses). To improve accuracy, a set of new potential AR's can be determined using our approach.

Minimising the cost of a multicast tree is an important issue. When a mobile node wishes to join an existing multicast tree, a route from the existing multicast tree to the node must be computed. Here, we deal with the problem of optimally connecting a mobile node to an existing multicast tree such that the selected node by the prediction algorithm still satisfies the QoS requirements. In this thesis, the formation of a near optimal multicast tree problem is considered which requires the prediction of a new Access Router(AR) and setting up a path pro-actively to it. It is shown that by using the proposed mobility prediction algorithm, we not only select the right access router to minimise the packet losses but also reduce the amount of resources that are used.

Another major challenge was to produce a mobility management tool to demonstrate and analyse the handoff procedure. Furthermore, our research concentrated on developing a mobility simulator as a part of a research grant that examined embedded IPv6 issues on heterogeneous technologies. The main focus of this research was to develop a simulator to demonstrate mobility in homogeneous and heterogeneous technologies. Roaming across heterogeneous wireless networks such as a cellular network and a wireless LAN network poses considerable challenges, as it is usually difficult to maintain an existing connection and guarantee the necessary QoS. A **vertical handover** is defined the process of switching over between two different networks. The simulation tool that was developed enabled us to demonstrate practical aspects of handover in homogeneous and heterogeneous technologies, which can be used for further research activities.

1.5 Contributions made in this Thesis

In this thesis, a novel hybrid model for mobility prediction is presented. An initial method is proposed based on a Grey model but this was found to have errors. Using a variety of different optimisation techniques, it will be demonstrated that these errors can be reduced to give accurate prediction of signal strength. This hybrid model used two types of evolutionary optimisation techniques which are well known in the literature, viz: Particle Swarm Optimisation (PSO) and Genetic Algorithms (GA). The results of these studies and enhancements to the basic prediction model have led to publications [4][5]. The prediction model was used to select the minimum number of AR's for handover. Based on these, a set of algorithms was proposed which led to publications [6], [7], [8], [9], [10], [11], [12]. A simulation model is built to study the basic handoff mechanism considering existing architectures and the performance results are evaluated. The research also aims to achieve a theoretical understanding of relative signal strength, radio behaviour, movement detection and prediction accuracy and their applicability to handoff algorithms. New algorithms are verified both theoretically and practically with our simulation. The improved algorithms proposed will support seamless handoff with improved data transfer and without any loss of packets. In the following, the key contributions of the thesis is summarised below.

The work in this Chapter 3 led to publications [4, 5]. The contributions of this chapter can be summarised as follows:

1. Development of a basic mobility prediction model based on the Grey model.
2. The models take into account the received signal strength from the base station. The prediction was tested for a window size of 4 which included slow fading.
3. The prediction model was tested and compared to some standard moving average filters and found to be superior.

The work in Chapter 4 led to publications [4, 6]. The contributions of this chapter can be summarised as follows:

1. Development of a hybrid model that includes the Grey model, fuzzy inference rules and particle swarm optimisation techniques.
2. Problems with the Grey model errors are fine-tuned with the hybrid model for accurate mobility prediction in order to improve handoff performance in wireless networks.
3. Simulation was carried out using a simple mobility scenario where the mobile node moves from one base station to another and the performance results were plotted and analysed. A fitness function was also chosen to optimise the values from the fuzzy inference rules.

The work in Chapter 5 led to publications [10]. The contributions of this chapter can be summarised as follows:

1. Further development of the hybrid model was done to select the best evolutionary optimisation technique for fine tuning the fuzzy parameters. Two algorithms are considered, viz: the particle swarm optimisation and the genetic algorithm. Each of these algorithms was compared in terms of the fastest convergence time to achieve the optimal value.
2. Hybrid models, viz: the genetic algorithm based mobility prediction and the particle swarm optimisation were tested and simulated. Along with this, a self tuning algorithm that was proposed in [13] was also tested for comparison purposes.

The work in Chapter 6 led to publications [9, 11]. The contributions of this chapter can be summarised as follows:

1. Development of an optimisation mathematical model for building a minimum cost multicast tree. A situation is modelled where a multicast tree is defined covering multiple access routers (AR) to maintain connectivity with the mobile node using mobility prediction (by selecting the least number of access routers) whilst ensuring guarantees of bandwidth and minimum hop count. To simultaneously solve the above two problem formulations gives rise to a multi-objective optimisation problem.
2. To solve the problem formulation outlined above, two algorithms are proposed based on the residual bandwidth and minimum hop count viz: The MMP algorithm and K-MinHop algorithm.
3. Numerical results are presented for various network sizes and topologies to determine the accuracy of the proposed algorithms.

The work in Chapter 7 led to publication [12]. The contributions of this chapter can be summarised as follows:

1. A situation was modelled that involved mobility prediction to select the can-

didate access router for building the multicast tree based on the residual bandwidth as an objective and the delay bound as a constraint.

2. Development of the MBWDC algorithm for building an optimal multicast tree based on the selection of candidate access routers that incorporates efficient join and leave procedures.
3. A numerical study is conducted to evaluate the accuracy of the proposed algorithm for different topologies that it employs for reducing the size of the multicast tree i.e., to reduce the total cost of the tree. It was also compared to the CAR-set (Coverage Access Router) algorithm proposed by Helmy et. al.

The work in Chapter 8 led to the publication [7] and a research grant [8]. The contribution of this chapter can be summarised as follows.

1. Based on the mobility prediction model, a simulator was developed as a part of the research grant (RPC) obtained from the Microsoft Corporation. A model was developed to demonstrate homogeneous and heterogeneous handoff using simulator which was built using Microsoft Visual Studio .NET and using C# as the programming language.
2. The model consisted of a discrete event simulation for handover from wireless LAN to a cellular network. The protocol specification used Mobile IPv6 as the standard protocol.
3. The tool consisted of a standard GUI and used XML to store the simulation results.

1.6 Organisation

Chapter 2: Background. In order to provide the foundation for development of the mobility prediction methodology, this chapter reviews existing schemes that are deployed in wireless networks. The chapter discusses the requirements of mobility prediction schemes and multicasting schemes to improve handoff in wireless networks. Mobility based on received signal strengths, their advantages and disadvantages, are discussed. This chapter concludes with a literature survey on the advantages and methods for using mobility prediction techniques and multicasting and their importance for future generation networks.

Chapter 3: Prediction Methodology. This chapter investigates issues related to mobility prediction and prediction accuracy. Grey theory is chosen to build the necessary prediction model. The parameters considered in our problem utilise the Received Signal Strength Indicator (RSSI) values from the base station. The Grey system model has a unique way of ordering the sequence of data that is known as the Accumulated Generation Operation (AGO). The basic prediction model is formulated and used to predict signal strengths from the base station. Handoff is performed based on this prediction. This chapter also compares our prediction algorithms with traditional estimation techniques such as the moving average filter and the exponential moving average. However, the prediction model based on the Grey method is shown to be better than any of the other “considered" estimation techniques.

Chapter 4: Hybrid Prediction Model. In this chapter, an improved technique is proposed that is a combination of Grey prediction, fuzzy logic and Particle Swarm Optimisation. The parameters considered in this chapter utilise the RSSI values from the base station. Since prediction error is inevitable, the output from the Grey model can be compensated for by the use of a fuzzy controller and then fine-tuned using PSO algorithms. In the past, many research papers have discussed reducing these errors by employing learning constants, which is a very tedious process. Our proposed technique minimises the number of handoffs and it is also shown to have a very short calculation time and better prediction accuracy compared with hysteresis based decisions. We have evaluated the Grey model and further work has been done to perform error compensation. To improve prediction accuracy, we take the output and calculate the error between the predicted output and the actual output. This would be treated using the fuzzy rules by actually compensating for the error and fine-tuning it with a PSO algorithm. Grey prediction uses very little data (as little as four measurements can be used to make a prediction) to predict the next signal strength. Even though the prediction accuracy of the Grey model is accurate it still deviates from accurately predicting these future values, which have a large variation – but our simulation model shows an improvement by using fuzzy inference rules and the PSO algorithm. The simulation results show that the model can improve prediction performance.

Chapter 5: Optimisation Algorithms for Hybrid Model. In this chapter, we discuss mobility prediction model using the Grey model which tracks the signal strength curve quite closely, but (inevitably) there is still some error. It should be noted that the Grey model does not predict large variations in the input data. The simulation model developed helps to serve two purposes: first, to decide which evolutionary algorithm best suits our problem and second, to see the performance of our prediction methodology with the two evolutionary algorithms and the self tuning algorithm. An optimisation criterion has been adopted and this is discussed in this chapter together with a comparative study between the self-tuning algorithm and the two evolutionary algorithms in terms of accuracy and speed of convergence. The improved accuracy of the approaches is shown by comparing results of simulations and experiments. The chapter discusses the genetic algorithm based prediction and Particle Swarm Optimisation based prediction and compares them in detail. The objective of this chapter is to describe how to utilise the benefits of these new evolutionary optimisation techniques and test their potential and competitiveness with respect to function optimisation.

Chapter 6: Handoff and Multicasting in Wireless Networks. This chapter illustrates the use of multicasting techniques aided by mobility prediction to improve handoff performance in wireless networks. Handoff holds the key to defining the performance of wireless networks since there are potentially some packet losses during handoff as the mobile node moves from one point of attachment to another. A new method of determining a multicast tree routing scheme with specific performance objectives is presented in this chapter. As previously mentioned, the Grey model has been used as the prediction methodology as it has been shown to provide good prediction accuracy [2]. A situation is modelled where a multicast tree is defined covering multiple access routers (AR) to maintain connectivity with the mobile node using mobility prediction (by selecting the least number of access routers to use in a multicast tree) whilst ensuring guarantees of bandwidth and minimum hop count such that packet loss can be avoided. To simultaneously solve the above two problem formulations gives rise to a multi-objective optimisation problem. It can be shown that the optimal routing problem is an NP hard problem where network state information is not accurate, and this is a common feature in wireless networks. To solve the problem formulations, we propose two algorithms viz: the MMP algorithm and the K-Minhop algorithm.

Chapter 7: Join/leave Algorithms. In this chapter, we provide a detailed description of the algorithm for join and prune mechanisms that will help to build an optimal multicast tree with QoS requirements during handoff. An analysis of how mobility prediction helps in the selection of potential Access Routers (AR) with QoS requirements which affects the multicast group size and bandwidth cost of the multicast tree is presented. The proposed technique tries to minimise the number of multicast tree join and prune operations. We have examined the performance of our algorithm using simulations under various conditions and we have obtained good performance results. Our results show that the expected multicast group increases linearly with the increase in the number of selected destination AR's for multicast during handoff. We observe that the expected number of joins and prunes from the multicast tree increases with group size. Thus, for an increased number of destinations, the estimated cost of the multicast tree in a cellular network also increases. We also hope that the discussion presented in here will help researchers and providers and will pave the way for future research.

Chapter 8: A Handoff Simulation Tool. In this chapter, we present the simulation modeler "NeTSim-version.3.0" which demonstrates the handoff procedures in homogeneous and heterogeneous networks. By developing such a tool we can plan the mobile environment in a efficient manner so that we can simulate the real-world problems. The tool itself is built on Microsoft Visual studio .NET and coded using C# as the programming language.

Chapter 9: Conclusions. Although each chapter in this thesis has its own concluding remarks, we dedicate this chapter to summarising and evaluating some final conclusions and discuss topics for future research.

1.7 Publications

Journals:

1. Suresh Venkatachalaiah, Richard. J. Harris and Robert Suryasaputra, "Improvement of handoff in wireless networks using mobility prediction and multicasting techniques", http://www.wseas.org, WSEAS TRANSACTIONS on COMMUNICATIONS, Issue 2, Volume 4, February 2005. (ISSN: 1109-2742), pg 104-111.

Conferences:

1. Suresh Venkatachalaiah, Richard. J. Harris, John E. Murphy, "Improving Handoff in Wireless Networks using Grey and Particle Swarm Optimisation", "2nd International Conference on Computing, Communication and Control Technologies": CCCT 2004 August 14-17, Austin, Texas, USA.
2. Suresh Venkatachalaiah, Richard. J. Harris, David Jones, Poster titled "CE. NET Embedded IPv6 Performance Issues - Heterogeneous Technologies " was presented at "DevCom-04", Microsoft Windows Embedded Developers' Conference 2004, June 28-July 1, 2004, San Diego, California, USA.
3. Suresh Venkatachalaiah, Richard. J. Harris and Robert Suryasaputra, "Improvement of handoff in wireless networks using mobility prediction and multicasting techniques ", in Proc. of 4th WSEAS Int. Conf. on Electronics, Hardware, Wireless & Optical Communications (includes the Symposium: Microwaves, Antennas and Radar Systems) (EHAC 2005), Salzburg, Austria, February 13-15, 2005.
4. Suresh Venkatachalaiah and R.J Harris, "Improving Handoff in wireless Networks using mobility prediction and evolutionary algorithms", 2nd International Conference on E-Business and Telecommunication Networks (ICETE '05), Vol 2., October 3-7, pp. 112-118, Reading, UK, 2005.

5. Suresh Venkatachalaiah, Robert Suryasaputra, Richard J. Harris, “An Algorithm for Join/Prune mechanisms for Improving Handoff using Mobility Prediction in Wireless Networks", IEEE Malaysia International Conference on Communications and IEEE International Conference on Networks (MICC & ICON 2005), pp. 1101-1107, 16-18 Nov 2005, Berjaya Times Square Hotel & Convention Centre, Kuala Lumpur, MALAYSIA.

Research Grant:

1. Suresh Venkatachalaiah, Richard. J. Harris, David Jones, awarded Microsoft research Grant for “CE.NET Embedded IPv6 Performance Issues - Heterogeneous Technologies", June 2004, Microsoft Research, Redmond, USA.

Local Conferences:

1. Suresh Venkatachalaiah, Richard. J. Harris, “An approach to improve handoff using mobility prediction and in wireless networks", “The 3rd Australian Telecommunications CRC (ATcrc) Conference and workshop" held on 12 - 13 December 2003 at University of Melbourne, Melbourne, Victoria, Australia.
2. Suresh Venkatachalaiah, Richard. J. Harris, “Mobility prediction and multicasting in wireless networks", “The 4th Australian Telecommunications CRC (ATcrc) Conference and workshop" held on 13 - 14 December 2004 at Nedlands, Perth, Western Australia.

Chapter 2

Background

2.1 Introduction to Wireless Networks

Years ago, the notion of transmitting information without wires or media seemed to be impossible. In 1896, the first patent for wireless communication was granted to Marchese Guglielmo Marconi [14]. Ever since that time, the advances have been revolutionary. With the proliferation of compact laptops and other internet enabled devices such as mobile phones, PDA's and other devices; more people than ever before are accessing the web while on the move. Streaming video and audio are also becoming increasingly popular. With the increased use of multimedia, it has become important for these wireless networks to provide seamless connectivity to these devices. The evolutionary path for mobile and wireless technology started with Analog Voice - Kbps in 1G - first generation networks; it then evolved to Second Generation (2G) networks which offered Voice + low-rate data 10 Kbps - 10 Mbps. A Third Generation of mobile networks is currently being rolled out around the globe with a Fourth Generation technology already being developed. 3G networks (Voice + data + multimedia) offer 100 Kbps - 100 Mbps) whilst Fourth Generation

(4G) networks (Voice + data + multimedia + QoS + IP) will offer 10 Mbps - Gbps [14].

Wireless networks make use of limited spectrum resources, which need to be conserved. Significant advances in digital modulation techniques have also influenced these advances in available resources. In wireless networks, bandwidth is limited, wireless links are error prone and there are frequent changes in the position of the mobile. Performance of mobile networks in a cellular environment is also influenced by the requirement to perform handoff as the users cross cellular boundaries. New applications are being developed for use with mobile devices and each of these presents new challenges in mobile system performance. Example: There has been a significant demand for online gaming operations (multiple), where players are located at different locations and users play games using their PDA's or handheld devices. A broad range of research challenges are associated with future mobile wireless systems networks that will be required to support anywhere/anytime multimedia communication and seamless ubiquitous access to information. Although resources are limited when compared with fixed line technologies, this technology releases users from the confines of cables, fixed access points and low speed connections.

2.2 Evolution of Wireless Communication Networks

The early development of mobile radio was driven by public safety needs. In 1921, Detroit became the first city to experiment with radio-dispatched police cars. The earliest wireless technology dates back to military and defence purposes only. Prior to this, Major Edwin Howard Armstrong in 1912 developed the regenerative circuit, which "revolutionised wireless radio communication because it could amplify weak radio signals without distortion far more effectively than other radio receivers of that time", (Microsoft Encarta 1998 edition). Armstrong's invention of Frequency Modulation technology brought success with the technology for FM broadcasting [14]. It wasn't until 1933, when two-way radio communications were established, that major advances took place in this field and this time it was another police department leading the way. The N.J. police department began using two-way radios in March, 1933. The system operated in a "push-to-talk" mode (i.e., half-duplex): simultaneous transmission and reception, or full-duplex mode, was not possible at the time. While one-way mobile radio communications was a dramatic improvement, it still had deficiencies. In 1940 Motorola developed the first handheld AM two-way radio, the Handie-Talkie, a 2.3-kg AM unit with a range of 1.6 to 4.8 km.

Prior to 1946, no civilian had access to a mobile wireless telephone, although portable radio units were used by the World War II military forces. On June 17, 1946 a major milestone was reached when AT&T and South Western Bell introduced in St. Louis the Mobile Telephone Service (MTS). A radio control terminal, which was monitored by an MTS operator, served as a gateway between the land mobile radio systems and the PSTN. The central transmitting antenna on a tall structure transmitted at 250 watts to provide coverage of about 20 to 25 miles in radius.

It was not until 1964 when AT&T inaugurated the improved mobile telephone service (IMTS), to replace the badly aging MTS [14]. Enabled by a number of recent technological advances on low cost, low power, small size and more complex circuits made possible by new semiconductors, IMTS provided automatic channel searching, direct dialing, and full-duplex operation. The widespread commercial introduction of IMTS began in 1964. IMTS was first introduced to the 150 MHz band, and was later extended to the 450 MHz band in 1969.

The big breakthrough came when AT&T Labs researchers Frenkiel and Engel divided wireless communications into a series of cells, and then automatically switched callers as they moved so that each channel could be reused. This led to the development of cellular phones and made today's mobile communications possible. The cellular concept was demonstrated by AT&T in 1947. In a cellular system, there would be many towers, each low in height and operating at low power, each covering a cell a few miles in radius, and all of them collectively covering an entire metropolitan area. Each tower would use only a few of the total frequencies allocated to the system, and as cars moved across the area, their calls would be handed off from tower to tower and among the available frequencies. The same frequency could be re-used in several simultaneous conversations at far smaller distances than was possible on conventional systems. Cellular towers would be linked to each other and to the wire line telephone network by a central switch. When a cell's capacity was exhausted in the future, they could be split into several smaller cells, each with almost the same capacity as the original, larger cells. The advantages of cells would be low power operation, handoff, reuse, and cell-splitting. The beauty of the cellular concept was that a finite number of frequencies could accommodate a large, and theoretically infinite, number of customers.

The first wireless phones utilised analog transmission technologies. The dominant analog standard was known as “AMPS", which stands for “Advanced Mobile Phone Service”. Analog standards operated on bands of the spectrum with a lower frequency and greater wavelength than subsequent standards, providing a significant signal range per cell along with a high propensity for interference. The first generation systems (1G) came in the 80's mostly for cellular analog voice using AMPS. This standard evolved into the second generation standard (2G) in the 90's to support digital voice and low bit rate data services. An example of such a cellular system is IS-54. The layout of a typical cellular network is as shown in Fig. 2.1. At the same time, wireless local area networks started entering into service starting at 1 Mbps for 802.11 standards and extending to 11 Mbps for the 802.11b standards close to the year 2000. In second-generation systems, TDMA systems transmit bits

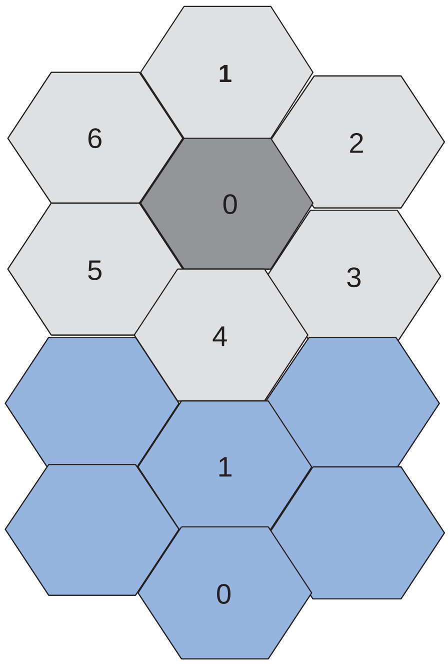

Figure 2.1: A cellular network model.

of digitised voice through individual data channels, one at a time. As the bits arrive at the other end, they are reconstructed and converted back into sound. GSM systems are based on technologies similar to TDMA. However, they operate at a higher frequency. CDMA systems represent the most advanced second-generation technologies. CDMA transmits digitised bits across a few adjacent voice channels to be reconstructed on the other side. CDMA provides much more reliable error recovery than TDMA or GSM. There are also 2.5 generation networks which provide the regular GSM services along with some packet based services. In third generation systems, generation of the standards (3G), cellular services have progressed to higher data rates in terms of 100's of Kbps to support voice, data and multimedia and wireless LANs have evolved via 802.11a and 802.11g to support data rates around 100 Mbps. In fourth generation systems (4G), the data rates are expected to continue to increase and will be able to provide IP-based services along with QoS

(Quality of Service).

2.3 Mobility in Wireless Communication Networks

Mobile communications technology has developed very rapidly over the past decades. For achieving high speed data rates, mobility is considered to be a important feature of wireless networks. A number of solutions attract special attention for seamless mobility support based on geographical areas where the network is deployed. This causes the mobility range to increase and depending on this, the mobility can range from campuses to cites, countries or even continents. Mobility presents a challenging issue for protocol designers since the protocol design must adapt to frequent changing in topology in a manner that must be transparent to the end user. To cater for the needs of different scales of mobility some of the wireless technologies use wireless LAN to serve small areas, cellular networks to serve mobility across towns and cities, as well as cellular 3G and satellite networks for mobility across countries. These technologies differ in the types of service they offer such as voice, data and real multimedia applications.

In all of the technologies that have been proposed, the mobile node has a point of attachment as the access point or a base station which can serve its mobility needs. In mobile communications information is exchanged by radio signals between mobile nodes (MN) and base stations (BS). Since each mobile node can communicate with other mobile nodes within a finite area, there is a need for several base stations in order to serve the whole service area. The area covered by the base station is called a “cell". These base stations are, in-turn, connected to an MSC (Mobile Switching Centre) which can carry information to the various base stations. The architecture discussed is common among most of the popular cellular networks. The important factor that governs this mobility is “handover" or “handoff" which is discussed in the section below. These terms have been interchangeably used in the subsequent chapters and sections.

2.3.1 The Handoff Process

When a mobile user is engaged in communication, the MN is connected to the BS via a communication radio link. If the user moves from the current coverage area (or cell) of another cell then the communication link to the old BS should be disconnected and a radio link to the new BS should be established without any losses. The process should take care that all the resources such as the frequency and time slot associated with the current connection should be restored in the similar manner to that which has existed before. This is often determined by the signal strength availability of the potential BS. The handoff process is often initiated by the deterioration of the received signal strength from the current base station. The handover process is initiated if and only if the resources are available in the potential BS. These initiations are based on the type of control which is available either on the BS or the MN.

2.3.2 Handoff Initiation

Handoff is referred to as “the sequence of events following mobility detection till the instance at which the MN starts receiving data packets from the new base station". Handoff essentially consists of handoff decision algorithm and the associated path set up scheme (mechanisms that enable packet delivery to the new base station). The handoff initiation criteria are based on the analysis of signal strength which is the most commonly used approach in GSM systems. The handoff initiation phase involves the following steps: monitoring the radio link, deciding to commence the handoff process and selection of a new base station. The handoff process should be initiated whenever the received signal quality deteriorates inside a cell or between two adjacent cells or when the mobile unit is moving along the common boundary of two cells. The criterion for handoff initiation is based on the received signal strength at appropriate sampling instants. Some of the handoff initiation algorithms that are mentioned in the literature are based on Relative Signal Strength, Relative Signal Strength with Threshold, Relative Signal Strength with Hysteresis, Relative Signal Strength with Hysteresis and Threshold, and Prediction Techniques. Handoff algorithms have been extensively studied in the context of mobile and cellular systems [15, 16, 17, 18] and can be summarised as follows:

“Relative Signal Strength method" selects the strongest received by the BS at all times. The decision is based on a mean measurement for the received signal. Due to channel fading this can cause oscillations and can lead to inappropriate handoff decisions. To overcome this problem an algorithm based on the “Relative Signal Strength with Threshold" approach was proposed. This method allowed a MN to hand off only if the current signal was sufficiently weak (less than a specified threshold) and the other signal is the stronger of the two. The effect of the threshold depends on its relative value when compared to the signal strengths of the two BSs at the point at which they are equal. If the threshold is higher than this value, this scheme performs exactly like the relative signal strength scheme, so handoff will occur. If the threshold is lower than this value, the MN would delay handoff until the current signal level crosses the threshold. Another scheme called “Relative Signal Strength with Hysteresis" allows a user to hand off only if the new BS is sufficiently stronger (by a hysteresis margin, 'h') than the current one. This technique prevents the so-called ping-pong effect, i.e., the repeated handoff between two BS's caused by rapid fluctuations in the received signal strengths from both BS's. In the Relative Signal Strength with Hysteresis and Threshold approach, the scheme hands a MN over to an new BS only if the current signal level drops below a threshold and the target BS is stronger than the current one by a given hysteresis margin. In prediction techniques the handoff decisions are based on the expected future value of the received signal strength. In our research, a technique will be proposed and verified by simulation to demonstrate better results, in terms of reduction in the number of unnecessary handoffs, than the relative signal strength approach, both with and without hysteresis, and threshold methods.

2.3.3 Handoff Decision

Most of the handoff decision algorithms proposed in the literature require the MN to sample multiple beacons or advertisements before making a decision. The beacon or advertisement interval typically being $50 - 200$ms, the mobility detection and handoff decision forms the lengthiest procedure when an MN switches from one sub-domain to another. Handoff decision algorithms are one of the very critical parts of any micro-mobility protocol. A bad mobility detection algorithm will adversely affect the performance of the micro-mobility protocol. A fast mobility detection and handoff decision can result in the MN not handing off to the best base station. An extremely slow mobility detection algorithm can lead to packet loss.

Two of the most commonly used handoff schemes are hard handoff and soft handoff. In hard handoff schemes, the mobile node is assumed to be able to talk to only one base station at a time. No schemes exist to forward packets from the old base station to the new base station after handoff. This scheme is very simple to implement, but incurs packet loss during handoff. In soft handoff schemes, the mobile node either has the ability to talk to more than one base station during handoff (typically the old base station and the new base station) or elaborate schemes are used to forward the packets arriving at the old base station on to the new base station, after handoff. Soft handoff schemes are more complex than the hard handoff schemes, but incur significantly smaller packet losses.

In second generation cellular systems such as GSM (Global Systems for Mobile communications) or PACS (Personal Access Communication Systems), different types of handoff decisions are made; which mainly depend on the handoff algorithms used. Depending on the decisions algorithms they can be categorised into three possible systems that are referred to as: MAHO, MCHO and NCHO [19].

Mobile Assisted Handoff (MAHO - e.g., supported in GSM systems) scheme moves most functional support and complexity to the mobile device. In this scheme, the access point or base station continuously measures the signal strength from the mobile node. If the signal strength falls below a certain threshold, a detection algorithm running on the current access point or base station collects the measurements to invoke the handoff. In the case of mobile assisted handoff, the handoff execution is distributed making shorter handoff completion times.

In Mobile Controlled Handoff (MCHO), the signal strength measurements are taken from the mobile device. If the candidate access point having a better signal strength is detected, then the handoff process is initiated. The mobile controlled handoff scheme moves most of the complexity for detecting handoff to the mobile device alleviating the network from centralised control of the handoff process. In this respect, mobile controlled handoff is more scalable and more distributed than other schemes.

In a Network Controlled Handoff (NCHO - e.g., supported in the AMPS cellular systems), a signal strength monitoring of the mobile device takes place in order to assist the handoff. The signal strength monitor initiates the handoff when the signal strength is below a certain threshold. Handoff is executed when a candidate access point is detected with a better service quality. Network controlled handoff moves most of the complexity for controlling handoff from the mobile device to the network.

2.3.4 Types of Handoff Procedures

Mobile communication has become a vital part of the current communication infrastructure. The future will see a higher quality of communication than currently available today. Demand for communication has increased rapidly but the allocation of spectrum is a finite resource. Cellular networks work on the principle of frequency reuse to optimise the available frequency bandwidth. Each base station is assigned a frequency band which is assigned in such a way that there is no interference with other base station frequency allocations. The base stations using the same frequency bands are separated by sufficient distances so as to eliminate the co-channel interference. A suitable multiple access scheme such as TDMA (Time division multiplexing) or FDMA (Frequency division multiplexing) is used to distribute the bandwidth assigned to a cell among the users, when the user switches to the neighbouring cell and has another channel.

Inter cellular handoff and Intra cellular handoff

During the call, the mobile can communicate with the same base station as long as the signal quality is good. When the signal quality decreases (as the user moves around) the call has to be transferred to a neighbouring cell and a handoff has to be initiated. In second generation networks such as GSM networks, there are two basic types of handoffs, namely Internal and External Handoffs. The internal handoffs are of 2 types, which are referred to as Intra-Cell handoff and Inter-Cell handoff respectively [20]. In Intra-Cell handoffs, a call is transferred from one channel to another within the same cell; whereas in Inter-Cell handoff, a call is transferred from one cell to another, both of which are under the control of the same Base Station Controller(BSC). The external handoffs are of 2 types, known as Intra-MSC handoff and Inter-MSC handoff. In Intra-MSC handoff, the calls are transferred between different BSC's, but belonging to the same Mobile Service Switching Centre (MSC). With Inter-MSC handoff, the calls are transferred between different MSC's. The old MSC is usually referred to as the anchor MSC and the new MSC is referred to as the relay MSC. A typical handover scenario in GSM networks is shown in Fig. 2.2.

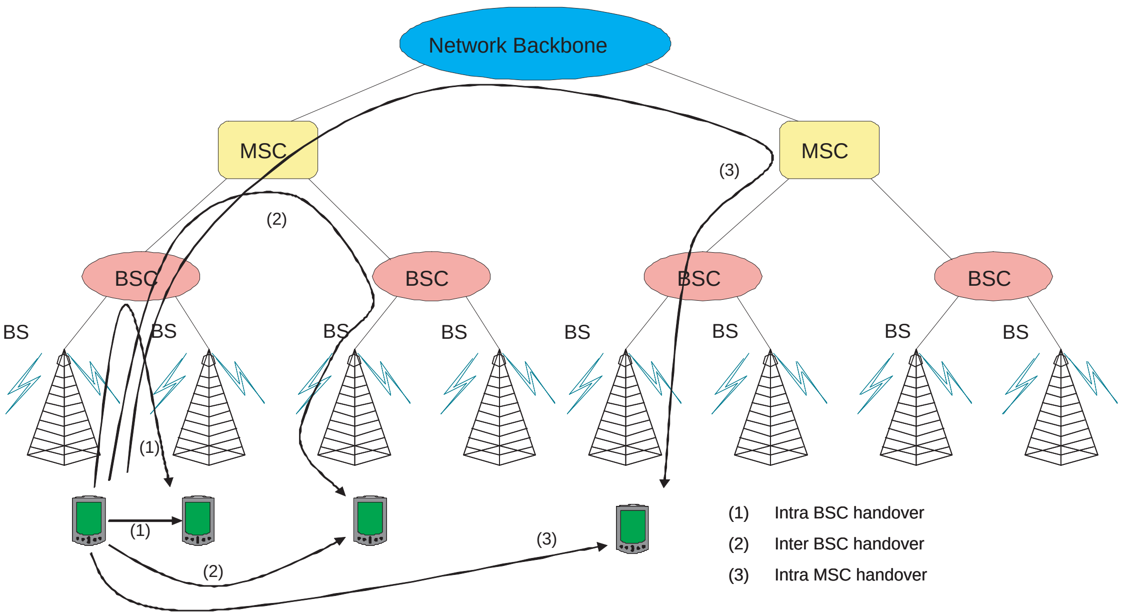

Figure 2.2: Typical handover scenario in a GSM network.

2.4 The Need for Mobility Prediction in Wireless Networks

The answer to this is not straight-forward, and with each attempt at an explanation, one can include new points for or against the need for this mobility prediction requirement! Both wired and wireless networks provide QoS support with appropriate control and signalling information. They are well defined and robust. With the growth of the Internet due to developments in WWW (World Wide Web) and the widespread popularity of email, there is a similar need for good QoS requirements in wireless networks as for traditional wired networks. The integration of wireless cellular networks and the Internet is an obvious scenario, one that is being realised in the standardisation process for the current generation mobile networks. In the telecommunications sector, the basic philosophy is always towards a balance between cost and quality. With the rapid developments taking place in these two technologies, Internet users will seek connection to the Internet while they are on the move.

Wireless network resource use depends significantly on the mobility of network users. The ability to predict this mobility, at least in part, enables the network to anticipate resource use in the future and take precautionary measures, if necessary. To ensure minimum disruption to mobile services, there should be minimal packet losses, minimum signalling requirements and enough system resources reserved. The next anticipated step in wireless communication is the delivery of data services, specifically internet services to mobile users. It is anticipated that mobile users in the near future will not only be concerned with the availability of these wireless services, but also with the quality of these services. The second item worth mentioning is the notion that a prediction is usually based on some previous knowledge. The exact specification of what knowledge is used to make a prediction is very crucial in determining the appropriateness of that prediction scheme. If a prediction is based on data that is simply not available in a given situation, that prediction scheme is useless in that scenario regardless of how well it performs in other scenarios.

Mobility prediction is focused on supporting the next expected handoff. It is a method to anticipate where the mobile node is going to move. As we know, a mobile node will typically experience a number of handoffs during a call depending on the speed of the mobile node and the coverage area of the base station or a given cell area. Further, it is not only necessary to predict the next cell or base station but also to reserve the resources in such a way that there is minimum interference and disruption to communications that are in progress. For next generation networks which promise much larger data rates, bandwidth and minimal delays, it is important for mobility prediction to consider all the available QoS metrics. But, due to the nature of the wireless networks such as delays, poor bandwidth and high error rates it is a very challenging task to meet many QoS requirements – especially when the mobile nodes experience frequent handoffs. This is where mobility prediction plays an important role which is a proactive method that helps in improving the performance during handoff. Thus, the schemes proposed to date are based on the reservation of resources and for this reservation we should have appropriate methods to predict where exactly these resources are needed. Some of the methods that have been proposed to predict these resources include user pattern based models, global positioning system based models, sectorised cell models etc. The Fig. 2.3 shows the mobility prediction schemes that have been proposed so far in the literature.

2.4.1 Application of Mobility Prediction to Different Networks

As future generation networks will offer higher data rates with QoS factors, it will be necessary to improve their coverage – irrespective of the type of network. In future generation networks such as 3G networks and beyond, a cell size varies from less than 100 metres for a pico cell, 1000 metres for a micro cell and to less than 20 Km for a macro cell. The actual choice of cell depends on the environment and the potential number of users, with the pico cells expected to be installed in city centres. These smaller cell sizes allow the available bandwidth to be shared among a smaller group of users or, in the case of a communication hotspot, to allow a larger number of users to communicate. These smaller cell sizes will increase the likelihood that the user, during any communication, will leave one cell and enter another (handoff).

The increased frequency of changing of cells means that there are more handovers and the call has to be handed over to the appropriate base station.

In order for a prediction scheme to be deployed in existing 2G/3G networks and to smoothly transition to future network technologies it needs to be efficient in both infrastructure supported (cellular) and infrastructure-less (ad hoc) environments. The prediction algorithm should not be sensitive to randomness in user movement patterns, as future networks are required to support global mobility. Future next generation networks are also envisioned as networks offering seamless and heterogeneous mobility which requires a potential mobility prediction scheme to be non-technology specific. A viable prediction scheme for a hybrid/ad hoc network should offer a high level of prediction accuracy while incurring minimal amounts of control overhead. These schemes include Location Criterion (based on current location), the Direction Criterion (based on current direction), the Segment Criterion (based on pattern matching), Bayes' Rule (based on conditional probability of future direction), and the Time Criterion (based on Direction Criterion with additional time constraint).

These mobility prediction schemes have been proposed for both wireless cellular networks as well as mobile Ad hoc networks. Recently, a number of schemes that apply user movement prediction to various aspects of mobility management have been reported in the literature. It has been shown through modelling and simulation, that the use of movement prediction is effective in enhancing the performance of resource reservation [21], handover management [22], and location management [23] schemes. Some schemes uses pattern matching techniques and a self-adaptive extended Kalman filter for next cell prediction based on cell sequence observations, signal strength measurements, and cell geometry assumptions.

2.5 Multicasting

Recently, streaming real-time multimedia content over the Internet has been gaining momentum in the communications, entertainment, music, automotive, and interactive game industries. Multicasting is the simultaneous transmission of data to multiple destinations. Multicasting will play a prominent role in the Internet in coming years, as group-based real-time applications (e.g., video conferencing) are continuing to become more popular. All of these applications also find use in a military context, including coordination, education, situation awareness, distributed simulation, and battlefield communication.

The multicast applications are generally categorised into one-to-many, many-to-many and many-to-one situations. Examples of one-to-many applications include scheduled audio/video distribution, push media, file distribution, caching, and monitoring of stock prices. Multimedia conferencing, distance learning, chat groups, distributed interactive simulations, and multiplayer games fall into the many-to-many category. Some of the many-to-one applications include resource discovery, data collection, and auctions. The first applications running on the multicast backbone (MBone) were real time audio and video for news, entertainment, and distance learning [24].

Multicasting is a method for group communication and it involves sending a message to a specific group of nodes in a network which is aimed at reducing the network resources that are used. As the number of mobile devices has increased and users are capable of moving at higher speeds and they require good QoS, it becomes necessary to use multicasting during handoff to support this requirement. Due to the higher speeds of the mobile nodes, there is increase in number of handoffs during a call and it becomes a challenging task to maintain continuity. Thus it becomes necessary to maintain a routing tree from the source to the set of destinations which can be done using multicasting algorithms. The multicasting algorithms activate the pre-established paths and the old paths are disabled. The algorithms should have the capability of determining the path with the necessary QoS as the mobile node moves through the network. The main advantage of using such a technique helps the handoff process to be executed quickly. However, the disadvantage is that multicasting to a group could increase the utilisation of network resources. But this disadvantage can be reduced by minimising the number of destination nodes which is done by the mobility prediction algorithm as in the case of our research. There are some protocols which address the problems of multicasting in wireless environment in the literature and the study of these protocols are discussed in the following sections.

2.6 Multicasting in Wireless Networks

To improve handoff performance, several multicast routing protocols for wireless mobile networks have been proposed in the literature [25, 26, 27, 28]. Multicasting is an efficient means of group communication. It has been used for video-conferencing and many other real-time applications, the advantage being bandwidth savings. Providing multicast support for mobile hosts in a wireless network is a challenging problem for several reasons. Here the multicast routing algorithm must not only deal with dynamic group membership but also dynamic membership location. Reconstructing the delivery tree every time the MN moves is not always a viable option because of the overheads involved. For efficient multicasts, data must be transmitted in parallel to various destinations along the branches of the tree and a minimum number of copies must be transmitted to only the necessary forks in the tree.

With the emergence of mobile Internet enabled devices and multimedia real-time applications, delay and packet loss during handoff have become important concerns in the existing mobility architecture. Some of the performance metrics like handoff delay, packet reordering, packet duplication, and routing efficiency play an important role. In our problem, mobility prediction is used to highlight the minimum number of access routers required to build a multicast tree with members who are the routers for the mobile node.

2.6.1 Multicast Assisted Handover in Wireless Networks

Deployments of the General Packet Radio Service (GPRS) and rapid development of the Universal Mobile Telecommunications System (UMTS) and third-generation cellular networks in general have been driven by increasing demand for high-speed wireless Internet access. Multicast assistance has been used to improve the handoff performance. One of the advantages of using multicasting is the ability to use advance registration with a new call in order to minimise the packet loss during handoff. In a mobile computing system, there are two types of nodes which are usually called wired nodes which could be a Access Router(AR) and a wireless node which is Mobile Node(MN). When the old AR sees that the signal from a mobile node is fading (and this is an indication of the onset of a handover condition), it triggers the AR's in the vicinity to join the multicast group. To avoid packet losses, handover must be detected early enough to provide an adequate time margin before actual handover takes place [29]. Once the mobile node is connected to the new AR, the remaining set of AR's will be removed from the group.

There are dynamic algorithms designed to identify probable new access routers [AR] based on observed signal strengths. After the identification of the probable AR's by mobility prediction, the packets for the destination are multicast to all potential AR's. Multicasting in mobile computing environments can be provided through the multicast routing protocols considering the properties of wireless networks. Hence, an efficient scheme is needed to support multicast for mobile nodes roaming in the smaller cells. The scheme should minimise the re-computation times for the multicast tree and reduce packet losses when the mobile crosses from one cell to another. Some of the schemes that are proposed to improve handoff performance using multicast based protocols are MobiCast, MoM (Multicast based Mobility), HAWAII (Handoff Aware Wireless Access Internet Infrastructure), CIP (Cellular IP), DVMRP (Distance Vector Multicast Routing Protocol), CBT (Core Based Trees), MOSPF (Multicast Open Shortest Path First), AMTree protocol etc. Some of the goals of these algorithms are to achieve

- Smooth handoff: which is expressed in terms of handoff delay, jitter, packet losses and communication overhead.
- Efficient routing: to deliver the resources in optimal way.
- Low wastage of bandwidth: to minimise overhead during detection and handoff process.

2.7 Related work on Prediction and Multicasting

There are many ways of solving the mobility prediction problem which has resulted in many prediction mechanisms being proposed. These mechanisms have been proposed to reserve radio resources and to configure cellular wireless networks due to increased demand for QoS in wireless networks. Each of the prediction mechanisms are unique and have been developed to solve a specific problem or particular type of problem. Most of them are related and can be used in scenarios that are different from their original intent. The details presented in this chapter provide a general overview of attempts to solve the problem of mobility prediction, based on some fundamental features found in most prediction mechanisms. Most of these mechanisms regard prediction as being based on location, movement history, move-

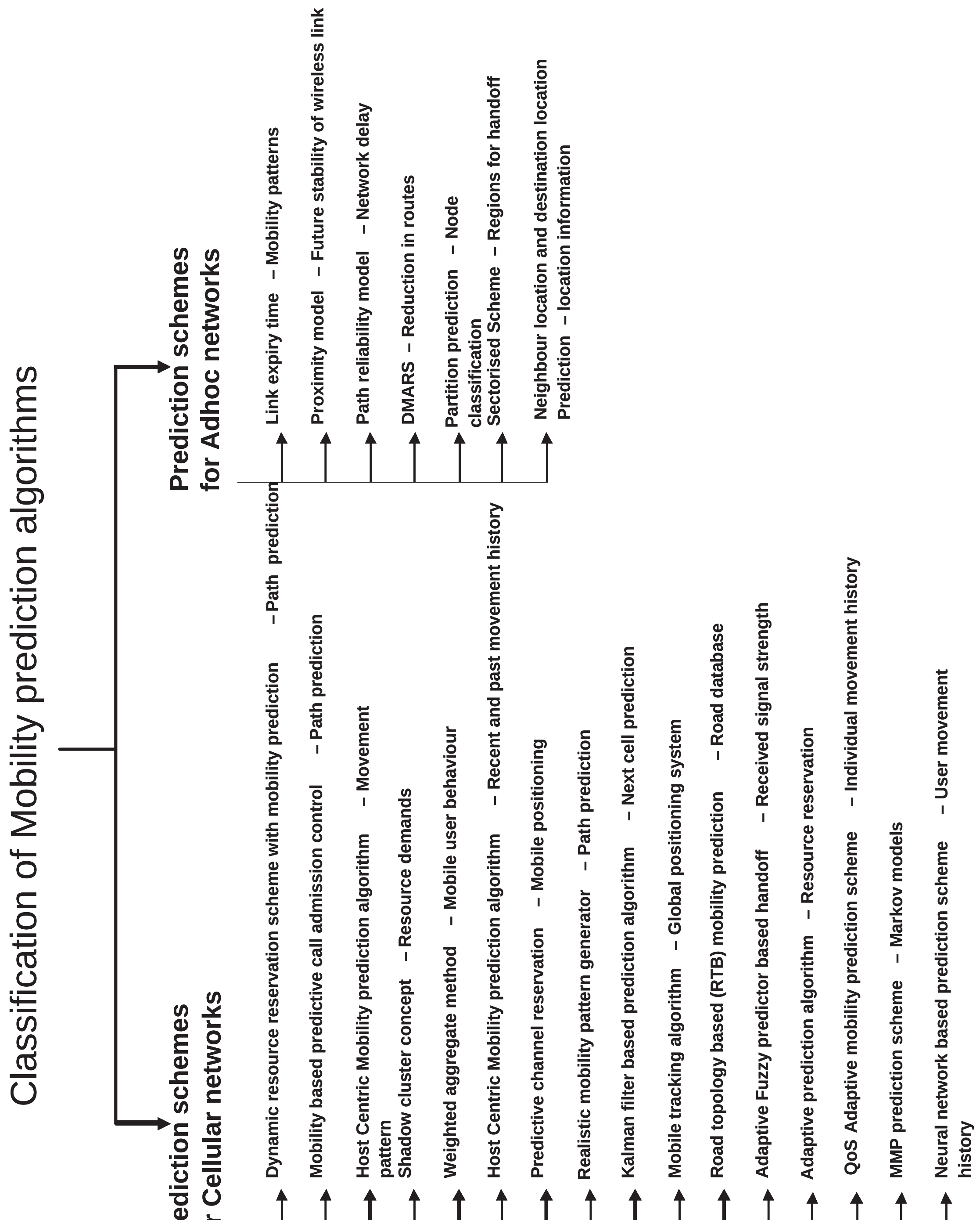

Figure 2.3: A comparison chart showing the prediction schemes proposed for wireless networks.

ment patterns, location and velocity, neural network based prediction etc. There are two main types of wireless networks where mobility is important namely, system supported by infrastructure, such as a cellular systems and systems that have no supportive infrastructure, such as ad hoc networks. The main difference is that an infrastructure supported system can refer to fixed base station for location while an infrastructure-less system needs an abstract location reference.

Mobility prediction research has mainly focused on determining the next expected handoff [30]. In reality, MNs will be able to move throughout the network and experience multiple handoffs during the lifetime of a call. It may therefore be necessary to predict more than just the next location visited by the MN or the next event the MN will experience [31, 32]. A major portion of this research still concentrates on user mobility and the connection trace for an MN. Many prediction systems proposed [33, 34] attempt to measure or capture some regularity of the user's mobility in order to extrapolate from this knowledge to determine the future behaviour of the user's MN. These are based on the movement history of the mobile node. Real life mobility traces have shown that this assumption of user mobility and connection trace of the MN is not as valid as most researchers believe. This raises the issue that it may not be possible to accurately predict the behaviour of an MN by studying the movement behaviour of the user. It will most likely be necessary to study the behaviour of the MN and its interaction with the network directly. The following sections will discuss prediction schemes proposed for both cellular and ad hoc networks and include details about the parameters used, their advantages and disadvantages. The sections also discuss the literature available on multicasting techniques proposed to improve the handoff performance.

2.8 Mobility Prediction Schemes for Cellular Networks

2.8.1 Prediction Schemes on Movement History and Patterns:

Hierarchical Position-Prediction Algorithm (movement history and RSSI measurements): A novel hierarchical position-prediction (HPP) [35, 36] algorithm improves connection reliability and overall system performance by accurately predicting the future movement pattern of the mobile user. HPP aims to improve connection reliability and overall system performance by accurate user movement pattern prediction. This algorithm adopts a two-level approach - at the top (global) level, approximate pattern matching is applied to determine the mobile's overall inter-cell movement direction, and at the bottom (local) level, an optimum self-learning Kalman estimator is applied that uses real-time signal strength measurements to estimate the mobile's intra-cell movement direction and velocity.

To provide prediction for all users with different mobility characteristic, an accurate model is modelled to explore the regularity and rationality in the seemingly random movement. In this scheme they propose algorithms which are composed of a approximate pattern matching algorithm that extracts any existing regular movement pattern that might exist to predict the global inter-cell direction, and a Extended Self-learning Kalman filter that deals with “unclassifiable" random movement by tracking intra-cell trajectory and predicting the next cell crossing. The study of the performance of these methods consists of the taking into account the presence of path loss, shadow fading and random user movement.

The global mobility model is motivated by the fact that most mobile users exhibit some regularity in their daily movement, and this regularity can best be characterised by a number of user mobility patterns (UMPŠs), recorded in a profile for each user and indexed by their time of occurrence. Global level mobility is in terms of inter-cell movements, the cell sequence the user might achieve during its connection lifetime. The UMPŠs proposed here are similar to the movement patterns, but are more robust in the sense that there is a decreased sensitivity to small deviations from the userŠs actual path (UAP). This is done while maintaining their effectiveness for estimating the mobileŠs inter-cell directional movement intention by modelling UAP as the edited version of a UMP, and employing an approximate pattern-matching technique to find the UMP that most resembles UAP.

At the local level, resolution is in terms of intra-cell movements modelled as a stochastic process with state variables that vary dynamically with time. The motivation for the local level modelling is to reduce the uncertainty of inter-cell mobility by tracking the user's intra-cell movement. The results reported from the HPP algorithm show that they remain reasonably accurate (83%) despite the influence of random movements. The time taken for the Kalman filter to stabilise is attributed to the initial inaccuracy of prediction. The main improvements of this scheme are setting up and reserving resources along a mobile path and planning quick handoffs between the base stations.

Autonomous Host Centric Mobility Prediction Algorithms (movement history and RSSI measurements): This scheme develops and evaluates mobility prediction algorithms for predicting the departure time of the mobile user at a certain location to another location - based on the mobile users' recent and past movement histories as well as their movement patterns. In autonomous host-centric mobility prediction algorithms [37], the authors predict future movements of mobile hosts by analysing recent and past movement histories. These mobility prediction algorithms are "autonomous" and "host -centric" since they require individual mobile hosts, rather than fixed base stations, to collect and maintain mobility prediction data, thus making them applicable to both cellular and ad hoc wireless networks. Each mobile host under these autonomous host-centric mobility prediction algorithms is capable of predicting the probability that it will leave the current location to the next location within a time period, thus facilitating resource reservation and route maintenance decisions, as well as for estimating the mobile host's residence time.

An on-going connection may be dropped during a location handoff if there is insufficient bandwidth in the destination cell to support it. If a mobile user can predict its arrival in a cell before it enters the cell, then the necessary bandwidth can be reserved to maintain the connection. This approach relies on the use of an accurate mobility prediction algorithm to supply the required mobility information. This work also considers using the current path to predict the location of the next mobile node except that they use a path definition based on the notion of resident locations. In addition, they also consider the residence time distribution to predict the probability that a mobile node will move out of the current location within a prescribed period. The work by [37] considers autonomous host-centric mobile nodes for collecting mobility prediction information to offload the base station. The mobility prediction information is summarised in the form of a probability distribution that predicts the mobility pattern of the mobile user when given a cell that it has just entered. The summarised prediction information is calculated by each mobile user, based on its movement history and is updated to its HLR (Home Location Register) periodically. Whenever a mobile node crosses a cell boundary and enters a new cell, the summarised prediction information is passed from the HLR to the cell's base station, which subsequently uses the mobility prediction information to perform necessary admission control and bandwidth reservation in its cell and neighbouring cells, if necessary.

Elliptical Shadow Algorithm (movement history): This scheme [33] involves a distributed algorithm for enhancing the connection quality and bandwidth utilisation in a cellular mobile network, which can improve the quality of service

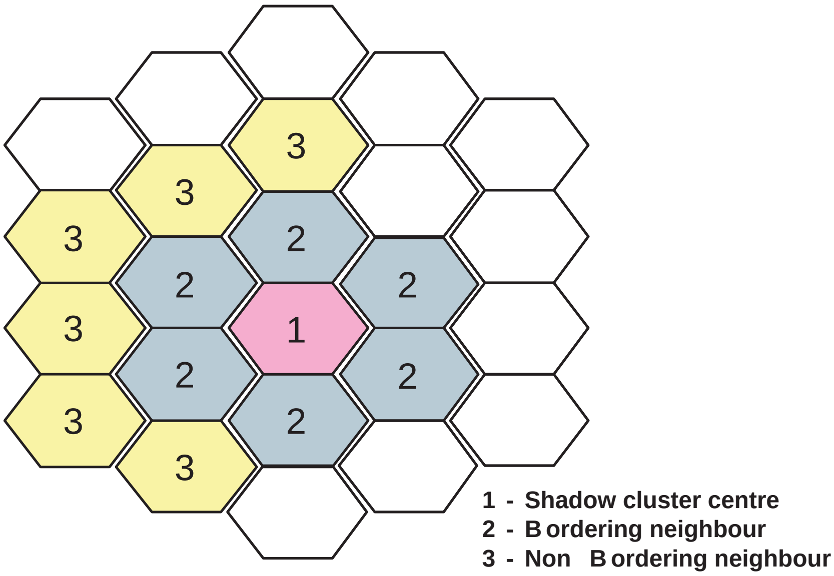

Figure 2.4: Shadow cluster concept.

(QoS) in wireless cellular networks. This algorithm allows the history information of the mobile entity to be used to predict an elliptical shadow around it, which covers all spots that the mobile would move to with high probability in a given time. The cells, and only the cells that overlap the shadow should reserve the resource for the mobile connection so as to achieve a low hand-off dropping rate and significantly reduce amount of resources reserved. The algorithm is self adaptive when the motion is very unpredictable, by observing the mobile's actual movement and automatically adjusting the elliptical shadow.

Shadow Cluster Concept (movement history): Levine et al. [38] have proposed the concept of a *shadow cluster*. A shadow cluster defines the area of influence of a mobile terminal (i.e., a set of base stations to which the mobile terminal is likely to attach in the near future). Like a shadow, this set moves along with the mobile, incorporating new base station's while leaving the old ones as they come under and out of the mobile's influence. Each base station in the shadow cluster anticipates the mobile's arrival and reserves resources for it. A close association exists between the prediction of the mobile's arrival and reservation of resources for it. The accuracy of the mobile's path prediction determines the number of base stations that reserve resources, and consequently determines the overall system efficiency. However, the proposed schemes based on the movement history requires base stations to collect and maintain mobility prediction information for all users and this generates a lot of traffic to the base station and leaves the approach infeasible. Also, according to [39], the proposal lacks the capability to determine the shadow cluster in real networks as it assumes either a precise knowledge of user mobility or totally random movements. An example of the shadow cluster concept is as shown in Fig. 2.4.

2.8.2 Prediction Schemes Based on Observed Traffic Patterns

Mobility Prediction Based on Observed Traffic Patterns: Mobility prediction algorithm can be used to predict the behaviour of mobile users and this prediction information can be used to define the optimal number of handover-only channels at a given time. The proposed scheme [20] tries to achieve lower blocking probabilities for handover and new users at higher system utilisation. This scheme uses user mobility information to optimise the resource reservation in cellular networks. The proposed handover strategy consists of two major components:

1. Mobility Prediction Model - A scheme that predicts the most probable target cells and the probabilities of handing over to these cells, dependent on the user's current state.
2. Advance Allocation Algorithm - An allocation strategy that determines the number of handover channels required in any cell, given the current state of the system and current position of the mobile users.

The method used to find the probability of a user making a handover to each of the neighbouring cells, given the users current position, speed and direction has been discussed in [40, 41]. The time taken to make predictions is important to the model, as the system will be able to determine possible handover targets and take necessary action before the user leaves the coverage area of the current cell. To make the mobility model more accurate, a model which can capture all the local variations of the probability has been used.

The idea is to create a weighted aggregate model by combining the effect of each point to form the probability density function of the whole distribution. This general concept has been used in this work to estimate the probability density function for each class. A class here represents one of the possible outcomes for a user, making a handover to a neighbouring cell or terminating the call before handover. The relative probability of certain events given the current position, direction and speed of users has been found. In order to estimate the relative probabilities of making handovers to each cell the probability density functions for all the classes have been normalised. The mobility model predicts the user's handover probabilities and a dynamic resource allocation algorithm estimates the number of channels required to support the predicted handover traffic. Results of the study show that the proposed solution increases the system utilisation by up to 5%.

2.8.3 Prediction Schemes Based on Mobile Positioning

Predictive Schemes Based on Handoff Prioritisation in Cellular Networks (Based on Mobile Positioning): The proposed method [1] analyses a new channel reservation approach, called predictive channel reservation (PCR) based on real-time position measurement and movement extrapolation. The authors propose and analyse a new channel reservation approach, called predictive channel reservation

(PCR) based on real-time position measurement and movement extrapolation. The underlying assumption of the scheme is that the position and orientation of the MN can be measured/estimated by the MN itself or by the base station (BS) or cooperatively by both the MN and BS. The measurements are used by the BS to predict the next cell for each MN. They assume that the cellular network employs fixed channel allocation (FCA), but the PCR concepts presented are potentially applicable to dynamic channel allocation (DCA).

One problem with all history-based schemes is the overhead to develop, store and update traffic histories for the different cells [1]. These histories can never be fully reliable as they continually go through short-term changes (e.g., diversion of traffic due to accidents), medium-term changes (e.g., traffic re-routing during road constructions and repair) and long-term changes. The PCR approach uses real-time position measurement to predict the future path of an MN. The utilisation of real-time measurements introduces a set of considerations not addressed in history-based schemes (e.g., changes in motion, cancelling false reservations, etc.). Most of the schemes assume that the cell shape is either hexagonal or circular but none of them have dealt with irregular sizes or fuzzy boundaries. Early work on handoff prioritisation proposed static reservation at each base station as a solution in which a fixed portion of the radio capacity is permanently reserved for handoff. However, this approach is unable to handle traffic load and mobility and it under-utilises precious radio resources when handoffs are less frequent and experiences high forced termination when the mobility is high.

2.8.4 Prediction Schemes Based on Path Prediction

Realistic Mobility pattern Generator, Design and Application in Path Prediction Algorithm Evaluation (Path Prediction): In [42], RMPG's design is based on the fact that every mobile user has some degree of regular patterns in his/her movement. This scheme discusses the use of RMPG for a mobility prediction algorithm used for proactive management of network resources. In this scheme, the design and implementation of a simulation platform, called RMPG takes into account the time-space regularity in the movement of real nomadic users. The simulation process is structured in two distinct stages: the generation of traces and their re-play. The new path prediction algorithm is based on the Responsive Learning Automaton (RLA) technique. RLA algorithms prevent the reduction of state transition probabilities below a certain threshold (i.e., probability values never drop down to zero). This approach is ideal for environments of dynamic nature where conditions are constantly changing.

In the first stage, the characteristics of the area where mobile users roam are defined. After completing the process that they call "Area Map Creation" the allowed moves in the space are recorded using the Cell IDs (CIDs) of visited Base Stations (BS). The result is a set of Legs (sets of CIDs) describing a "mobility map" of the area. Due to the volume of recorded Legs, the use of a database management system has been considered advantageous for storage/retrieval.

In the second stage, the movement of users along the area is simulated. Each user may have a different role profile (businessman, student, etc.) and, therefore, exhibit a quite different behaviour from others. The movement of users is constantly fed to the Mobility Function module allowing various tests on different kinds of mobility algorithms to be performed.

2.8.5 Prediction Schemes Based on location prediction

Predictive Distance-Based Mobility (location and velocity): In this scheme [43], a mobile's future location is predicted by the network, based on the information gathered from the mobile's recent report of location and velocity. When a call is made, the network pages the destination mobile around the predicted location. A mobile makes the same location prediction as the network does; it inspects its own location periodically and reports the new location when the distance between the predicted and the actual locations exceeds a threshold. To more realistically represent the various degrees of velocity correlation in time, a Gauss-Markov mobility model is used. For practical systems where the mobility pattern varies over time, they propose a dynamic Gauss-Markov parameter estimator that provides the mobility parameters to the prediction algorithm.

Based on the Gauss-Markov model, the authors describe an analytical framework to evaluate the cost of mobility management for the proposed scheme. They also present an approximation method that reduces the computational complexity of the cost evaluation for multidimensional systems. They compare the cost of predictive mobility management against that of the regular, non-productive distance-based scheme, for both the case with ideal Gauss-Markov mobility pattern and the case with time-varying mobility pattern. The performance advantage of the proposed scheme is demonstrated under various mobility patterns, call patterns, location inspection cost, location updating cost, mobile paging cost, and frequencies of mobile location inspections. As a point of reference, prediction can reduce the mobility management cost by more than 50% for all systems, where mobile users have moderate mean velocity and where performing a single location update is as least as expensive as paging a mobile in one cell. However since the scheme does not take direction of travel into account it is not considered too accurate.

Neural Network Techniques for Prediction in a Mobile Networking Location (Location Prediction): Neural Networks (NN) are very sophisticated modelling techniques capable of modelling extremely complex functions [44]. In particular, NN's are non-linear. These networks learn by example. NN user gathers representative data, and then invokes training algorithms to automatically learn the structure of the data. NN methodology for location prediction can be applied in two steps. In the first step, suitable NN is trained with observed motion pattern. It is important to note that prediction is not possible unless some regular subscriber movement patterns are detected and stored in the database. In the second step the trained NN is used for prediction, the actual movement and time is used to feed to the trained neural network to get the next location. In this study, the possibility of using NN technology to predict the mobile users next location prediction is based on the information of his current location as well as time. Conventional statistical techniques that include one of the advanced statistical techniques, like Box-Jenkins have also been explored to predict the mobile subscriber location. Results of the techniques are compared and used “in-house" developed software packages for both Neural Network and Box-Jenkins.

Mobile Motion Prediction Algorithms (Location Based): A set of Mobile Motion Prediction (MMP) algorithms [34] is used to predict the “future” location of a mobile user according to the user’s movement history patterns. The data or services are pre-connected and pre-assigned at the new location before the user moves into the new location. Thus, the user can immediately receive service or data with virtually the same efficiency as at the previous location. The MMP algorithms are based on the fact that everyone has some degree of regularity in his/her movement, that is, the movement of people consists of random movement and regular movement, and the majority of mobile users have some regular daily movement-patterns and follow these patterns more or less every day of the week.

The algorithm discussed in [34], is based on Mobile Motion Prediction (MMP) scheme for the prediction of the future location of a roaming user according to his movement history patterns. The scheme consists of Regularity-Pattern Detection (RPD) algorithms and a Motion Prediction Algorithm (MPA). Regularity Detection algorithm is used to detect specific patterns of user movement from a properly structured database (IPB: Itinerary Pattern Base). Three classes of matching schemes are used for the detection of patterns namely the state matching, the velocity or time-matching and the frequency matching. The Prediction Algorithm (MPA) is invoked for combining regularity information with stochastic information (and constitutional constraints) and thus, reach a decision - prediction for the future location (or locations) of the terminal. Simulations of the proposed scheme have shown a maximum prediction efficiency of 95%. Further, some of the papers argue that the performance of the MMP is accurate for regular patterns but decreases linearly with increase in random user movement.

2.8.6 Prediction schemes on Path Prediction

Dynamic Resource Reservation Scheme with Mobility Prediction for Wireless Multimedia Networks: To reduce the call blocking probability and call dropping probability, the schemes employed mainly consist of resource reservation [21]. However, pure resource reservation leads to inefficient resource utilisation. In order to prevent unnecessary reservation, the prediction of the direction of movement has been employed in this method. This scheme consists of dynamic resource adjustment, resource reservation, resource configuration and path prediction. To reduce unnecessary resource reservation, the prediction of the direction of movement has been used to enhance pure resource reservation. Therefore, a dynamic resource reservation scheme with mobility prediction is proposed in this method. In this scheme, a resource request with available range rather than a fixed resource request is specified by each of the traffic to make the reservation scheme flexible.

Utilising the strategies of basic resource reservation and dynamic resource adjustment, the scheme optimises resource utilisation and avoids dropping of service.

In order to use the resource in a more efficient way, a mechanism of path prediction is added to the proposed scheme. The method of path prediction used here is mainly determined by the relative distances between the mobile device and its neighbouring base stations. When the mobile device continues moving, its relative distances to the neighbouring base stations are changed, too. A square cell is considered as a cell shape and a threshold is set for distance. If the distance between the mobile device and a specific base station is less than this threshold, it is considered that the mobile device is moving in this direction. Thus, the resources in this particular cell will be passively reserved. The idea effectively enhances the utilisation of the system resources and reduces the burden of the system since not all neighbouring cells will reserve resources for the mobile device. This scheme tries to improve the QoS of a mobile device and achieve better resource utilisation.

2.8.7 Prediction Schemes on Road Topology Prediction

Dynamic Bandwidth Reservation in Cellular Networks Using Road Topology Based Mobility Predictions (Trajectory Prediction): In [22] the authors propose a solution to predict the trajectory of mobile terminals so as to perform bandwidth reservation in advance with the vision that future mobile devices will most-likely be equipped with a reasonably accurate positioning capability. They propose a novel mobility prediction technique that incorporates both mobile positioning information and road topology knowledge. They develop an adaptive bandwidth reservation scheme that dynamically adjusts the reservation at each base station according to both incoming and outgoing handoff predictions generated using a mobility prediction technique.

The method involves a novel predictive reservation scheme that utilises knowledge of road topology, in addition to positioning information. It could potentially achieve more accurate predictions at the cost of increased complexity, but the resulting gain in reservation efficiency may justify this cost. The proposed technique requires the serving BS to receive regular updates about each active MN's position every Δ, say, 1 sec. This will consume a small amount of uplink wireless bandwidth (several bytes per update for each MN), which might be negligible for future broadband services. The output of each prediction has the form of a 4-tuple: [target cell, prediction weight, lower prediction limit, upper prediction limit].

The target cell is the MN's predicted handoff cell. The prediction weight is a real number between 0 and 1 that indicates how likely accuracy of the prediction. The lower prediction limit (LPL) gives a lower statistical bound for the actual remaining time from handoff, t_{remain}, with probability ξ_l, i.e., $P[t_{remain} \geq LPL] = \xi_l$. The upper prediction limit (UPL) gives an upper statistical bound for t_{remain} with probability ξ_u, i.e., $P[t_{remain} \leq UPL = \xi_u$. It is noted that each MN may have more than a single 4-tuple; a 4-tuple is specified for each possible path from its current position that may lead to a handoff within a given time $T_{threshold}$. The essential information required for making predictions is held by the base stations. Unlike previous attempts which have assumed either hexagonal or circular cell geometries, this technique caters for irregular handoff regions. They also incorporate road topology information into the prediction technique, which could potentially yield better prediction accuracy for MNs that are carried in vehicles.

QoS Adaptive Mobility Prediction Algorithms (Real life movement traces): In a populated network environment, it seems that advance reservations and configurations are an effective way to decrease the call-dropping probability and to shorten handover latency. The QoS Adaptive Mobility Prediction (QoSAMP) algorithm [23] is a prediction scheme driven by QoS demand or tariff preference. It mainly concentrates on resource reservation and network pre-configuration at locations in the direction of travel. A parameter known as the Prediction Confidence

Ratio (PCR) is proposed to minimise the effect of statistical randomness in a user's movements. Basically, they avoid predicting random movements, and argue that there is a trade-off between the optimisation of resource usage and a guarantee of QoS measures. To achieve a stricter QoS compliance, they transform the QoS requirement of applications or the tariff preference of mobile users into a particular value of PCR, which is applied to the adaptive prediction mechanism. A prediction is derived from the probability distribution of all possible next moves. If the first predicted cell does not satisfy the required PCR, additional cells are involved until the probability sum exceeds the PCR. There is a strong dependence on the stored probability values of past user movements with prediction from new locations with a lack of history, being approximated to a group mobility model. Prediction from a new location is based on the movement pattern of the majority in the cell. The algorithm does not deal with the prediction of non-regular individual user movements allowing room for prediction inaccuracies. To ensure prediction accuracy even in the less complex non-random scenario the scheme requires increased PCR. A higher PCR value requires more number of cells to be involved. The advantage of using this approach is that it neither requires the implementation of complex cell structure nor any extra cost of GPS receivers. The results of these schemes provide improvement over arbitrarily selecting an adjacent cell, but the randomness of a user movement may prevent good prediction accuracy.

2.9 Mobility Prediction Schemes for Ad hoc Networks

A mobile ad hoc network (MANET) is a self-configuring network of mobile routers (and associated hosts) connected by wireless links, the union of which form an arbitrary topology [45, 46]. The routers are free to move randomly and organise themselves arbitrarily; thus, the network's wireless topology may change rapidly and unpredictably. Mobile ad hoc networks represent the purest form of decentralised systems and, therefore, they impose many challenges to cooperative communication. As a consequence, much ad hoc network research has focused on the investigation of fundamental algorithms for routing on which almost everything else relies. Routing a packet from a source to a destination in a mobile ad hoc network is a challenging problem, since nodes in the network may move and cause frequent, unpredictable topological changes. In accommodating the communication needs of the user applications, the limited bandwidth of wireless channels and their generally hostile transmission characteristics impose additional constraints on how much administrative and control information may be exchanged, and how often.

Ensuring effective routing is one of the greatest challenges for ad hoc networking. Several unicast routing protocols have been proposed for MANETs in an attempt to solve the routing problem. Example protocols include Dynamic Source Routing (DSR) [47, 48], Ad hoc On demand Distance Vector (AODV) [49], and the Zone Routing Protocol (ZRP) [50, 51]. To improve the performance of unicast communication, some of the proposed MANET unicast routing protocols use location information in the routing protocol.

Some of the proposed algorithms include

- the Location-Aided Routing (LAR) algorithm
- the Distance Routing Effect Algorithm for Mobility (DREAM)
- the Greedy Perimeter Stateless Routing (GPSR) algorithm
- the Geographical Routing Algorithm (GRA)
- the Terminode Remote Routing (TRR) protocol
- the Scalable Location Update-based Routing Protocol (SLURP)
- the Depth First Search (DFS) algorithm

- the Greedy-Face-Greedy (GFG) algorithm
- the GPS Zone Routing Protocol (GZRP)

One of the main challenges of a position-based routing protocol is to learn an accurate position or location for a packet's destination. While some protocols (e.g., DREAM and LAR) include the exchange of location information as a part of its protocol, most position-based routing protocols assume a separate mechanism that provides location information.

Mobility prediction has also been applied to design efficient routing algorithms [52] for MANETs before. Several approaches have been proposed to decrease overhead of route discovery by utilising location information for mobile hosts such as GPS. Such location information obtained using the global positioning system have been used in location aided routing (LAR) [53, 54] protocols for route discovery. Global Positioning System (GPS) [55] in QoS routing decisions was the first of its kind in mobility prediction to consider and predict the connection time (estimated lifetime) of wireless links. Some of the others proposed a simple mechanism to predict durations of wireless links in a MANET by assuming directions and speeds of end nodes of wireless links will not change in the future. The authors of [56] introduce prediction-based link availability estimation. They also propose to use their estimation algorithm to develop a metric for path selection in terms of path reliability, which is shown by simulations to improve network performance. In order to better understand and quantify mathematically the mobility of ad hoc network nodes, several mobility models and prediction have been proposed in literature.

The motivation behind the mobility models has been the potential of mobility prediction for application in various fields of ad hoc networking. In ad hoc wireless mobile networks, the mobility models focus on the individual motion behaviour between mobility periods, in which a mobile host moves in a constant direction at a constant speed. Ad hoc networks are more sensitive to mobility than cellular wireless networks since in the latter the topology changes only when a node leaves the cell, irrespective of relative connectivity with other mobiles. The application and critical study of various mobility prediction schemes has shown mobility prediction to offer potential improvements.

2.9.1 Mobility Prediction Scheme using Wavelet Neural Network

This scheme proposes to predict the mobility of mobile nodes with a wavelet neural network and the resulting clustering algorithm which provides a generic and stable cluster structure for the upper-layer protocols. For the clustering scheme proposed, they give a analytical model and performance evaluation. Many researchers have focused their attention on partitioning the ad hoc network into multiple clusters [57]. The clustering algorithms are crucial to ad hoc networks. A cluster head is elected within each cluster to form the upper-layer backbone. With the help of the cluster head, a hierarchical routing protocol can be implemented. Since the movement is the main cause, they propose a scalable mobility prediction scheme based on the accumulated past behaviour. In this work, a novel clustering protocol based on the Wavelet Neural Network mobility Prediction(WNNP)[58] is proposed. The major difference from the previous research is that they design the clustering operation from mobility's point of view, i.e., predicting the nodes mobility by a wavelet neural network. In contrast, most prior work focuses on the algorithm out of regard for management support, lacking an overall cluster stability evaluation. The main goal is to provide a generic and stable cluster structure for the upper-layer protocols. In this proposed clustering algorithm, the member node predicts its mobility by the wavelet neural network and computes its visit probability to the cluster.

From the wavelet neural network, they not only get the next possible cluster, but also a series of clusters that reflect the node trajectory. The simulations are compared with respect to the performance of the clustering algorithm with the Lowest-ID, maximum connectivity clustering algorithms, in terms of stability of clusters being formed. The cluster stability is measured by determining the number of each mobile node that either attempts to become or gives up the cluster head role.

2.9.2 Mobility Prediction Scheme using DMARS

For any active connection in ad hoc network, the end hosts as well as the intermediate nodes, which relay the information, can be mobile. Therefore, under moderate mobility, end users experience frequent disconnections on their established path. Current research in this area has focused on developing mobility-based clusters with less mobile node cluster heads for ad hoc networks, rather than on the challenges involved in using mobility predictions for routing. In this scheme they propose Distributed Mobility Aware Route Selection (DMARS)[59] for improving performance of existing routing protocols using mobility predictions. The mobility metric, which exploits non-random behaviours of mobility patterns, is used to select more stable routes and reduce the routing overhead caused by userŠs mobility.

Approaching the above goal in an ad hoc network is substantially more complex due to heterogeneity of available resources, multi-hop communication and movement of intermediate nodes serving as routers, thereby affecting the route. Combination of such factors complicates the task of using reliable transmission schemes and time critical applications. In ad hoc networks, route determination is done by flooding, which puts excessive overhead and reduces the efficiency in using the channel. Also mobility of node causes frequent route breaks and therefore re-route discoveries.

The purpose of this scheme is to present a distributed mobility-aware scheme for route selection, where each node evaluates its own mobility independently and provides feedback in determining and selecting the stable path. Since no central control is available in the ad hoc network, the scheme performs its operations in a distributed manner. They develop and use, two simple metrics, node mobility metric and path stability metric, to enhance the performance of existing routing protocols under wide range of mobility. The node does not depend on GPS or any special device for calculation of these metrics. It has been observed that in mobile networks, each mobile host exhibits some mobility pattern. For example, planes moving in the battlefield to destroy some target are likely to maintain their heading and speed for some period of time before they change or a person walking on the road or driving a car will follow the path and maintain the speed within some range. By making use of userŠs non-random mobility pattern, they predict the future state of a network topology and thus provide the most stable route.

In ad hoc networks, it is crucial to minimise the number of disconnections in the ongoing transmissions. In this scheme they have examined the mobility aware route selection scheme and the performance results show that DMARS exhibits much stable route when applied to existing protocol at negligible additional overhead of maintaining the various metrics. Routes that are more stable and stay connected for longer duration can be chosen by utilising mobility prediction.

2.9.3 Mobility Prediction Schemes using a Sectorised Algorithm

In this scheme a method of mobility prediction that can aid in achieving seamless mobility is presented [60, 61]. In order to optimise the efficiency of a resource reservation algorithm, accurate prediction of the future movements of the user is required. If the user movements can be predicted accurately in a hybrid network environment then handoff/cluster change, resource reservation and context transfer procedures can be efficiently completed as required by node mobility.

Clustering has been defined and proposed in ad hoc network for grouping in the literature. Every node in the network belongs to a cluster and will change its cluster of membership as effected by mobility. The proposed prediction scheme is built on the rationale that in order to achieve maximum prediction accuracy the prediction process is restricted to areas of high cluster change probability. To ensure prediction accuracy the process guards against under-prediction (i.e., commencing the prediction process too late so as to miss a cluster change) and over-prediction (i.e., predict too early along a user path). Prediction restricted to the last movement legs of a mobile user ensures higher accuracy of prediction. In this scheme a sectorised cluster structure based on cluster change probability to aid in mobility prediction is proposed.

The sectorised ad hoc mobility prediction scheme introduces the sectorised cluster structure and makes use of the cluster-sector numbering scheme to predict user movements in an ad hoc network. Due to the dynamic topology of an ad hoc network, prediction of user movements from mobility history bases (MHB) is proved to be not possible and/or efficient. Hence MHBŠs are not employed for prediction purposes in a purely ad hoc network. This work is extended to the sectorised mobility prediction algorithm to the ad hoc networking domain to assist in context and routing information [60, 61]. By exploiting a mobile users' non-random travelling pattern they predict the future states of the network topology changes and provide transparent access during the period of change.

2.9.4 Group Mobility and Partition Prediction Schemes

In wireless networks, network partitioning occurs when the mobile nodes move with diverse patterns and cause the network to separate into completely disconnected portions. Network partitioning is a wide scale topology change that can cause sudden and severe disruptions to ongoing network routing and upper layer applications. Its occurrence can be attributed to the aggregate group motion exhibited in the movements of the mobile nodes. By exploiting the group mobility pattern, the proposed scheme is used predict the future network partitioning, and thus minimise the amount of disruptions [57, 62]. A new characterisation of group mobility based on existing group mobility models is used, which provides parameters that are sufficient for network partition prediction. They demonstrate how partition prediction can be made using the mobility model parameters, and illustrate the applicability of the prediction information. Furthermore, they propose to use a simple but effective data clustering algorithm that, given the velocities of the mobile nodes in an ad hoc network, it can accurately determine the mobility groups and estimate the characteristic parameters of each group.

Researchers have proposed mobility prediction schemes that attempt to predict the future availability of wireless links based on individual node mobility model, in order to improve routing algorithm efficiency and build more stable routes. The changes in link availability are caused by local topology changes. However, global scale topology changes such as network partitioning cannot be predicted by these schemes. The main cause of network partitioning is the group mobility behaviour of the mobile nodes in wireless ad hoc networks, where the mobile nodes belonging to the same movement group exhibit similar movement characteristics, while the nodes of different groups show diverse mobility patterns. When a network partitions, the partitioned parts are completely disconnected from other parts of the original network. Upper layer routing and other applications involving nodes in separate partitions are severely disrupted, and may terminate if the partitions do not merge in time.

In order to predict network partitioning, they identify and characterise the group-based movements of the mobile nodes, and use the characterisation to quantitatively model the topology changes. Based on the topology changing pattern, they derive important information about future network partitioning. The main contributions of the scheme are: First, they propose a new and enhanced characterisation of the mobility groups based on existing models, which provides parameters sufficient for network partition prediction. Second, they show a method of predicting partition timing with the parameters provided by our enhanced group mobility model. Third, they use a simple data clustering algorithm such that, given the velocities of mobile nodes, it can accurately identify the mobility groups and estimate the characteristic parameters of each group necessary for the partition prediction. Finally, effectiveness of the clustering algorithm with respect to mobility group identification is illustrated.

Based on the mobility parameters provided by the model, they have shown how future network partitioning can be predicted. They also proposed to use a low-complexity data clustering algorithm that can accurately determine the mobility groups and their mobility parameters, and identify the group membership of each mobile node in the network. The main purpose of this scheme were to show the cause-and effect relationship between group mobility and network partitioning in the wireless ad hoc networks, and investigate how mobility groups can be determined from node velocities.

In [62] one of the important issues associated with group mobility in ad hoc networks is predicting the partition time. Existing algorithms predict the partition time assuming that the partitioned groups move to opposite direction with the same speed and the same coverage. So, these algorithms cannot accurately predict partition time in practical situation. In this scheme, a partition prediction algorithm considering network partition in any direction, at any speed, and with different coverage of group is proposed. The algorithm can predict the partition time more accurately in real situations. The main cause of network partition is the group mobility behaviour of the mobile nodes, in which some nodes belonging to a group exhibit similar movement characteristics while the others in another group show a different mobility pattern. If partition time is predicted behind time, nodes in partitioned group cannot receive some important messages. Some researchers have proposed partition prediction algorithms by assuming that a group is partitioned into two clusters and they move to opposite angle with the same speed and the same coverage. Therefore, these algorithms will not accurately predict partition time in practical situation. In this scheme, a partition prediction algorithm considering any angle, any speed, and any coverage of cluster is proposed.

RVGM(Reference Vector Group Mobility) model [63] considers group partition. The node movement can be characterised by the velocity $W = (W_x, W_y)^T$, where W_x and W_y are the velocity components in the x and y directions. Each mobility group has a characteristic group velocity. The member nodes in the group have velocities close to the characteristic group velocity but deviate slightly from it. RVGM model is represented by modelling the group velocity, $W_j(t)$, and the local velocity deviation of the member nodes, $U_{j,i}(t)$, as random variables each drawn from the distribution $P_{j,t}(w)$ and $Q_{j,t}(u)$, respectively. Also, this model uses three factors, so that it is able to predict each group and node mobility as well as group partition. The existing partition prediction algorithm [57] predicts the partition time assuming that each group moves to the opposite direction, at the same speed, and with the same coverage. The existing algorithm cannot predict the partition time accurately in a practical situation. In this scheme, a partition prediction algorithm considering group movement in any direction, at any speed, and with different coverage has been proposed which is an improvement to the existing technique.

2.9.5 Mobility Prediction Scheme Based on Proximity Model

This scheme [64] examines the problem of finding stable links for routing in ad hoc networks. It considers the question of link availability: how long will two nodes remain in close enough proximity for a link between them to remain active? More precisely, with what probability will two nodes remain within a given distance threshold of one another over time? The scheme builds on the probabilistic model for link availability presented in [65] by proposing an adaptive learning strategy to discover, with high probability, when two nodes are effectively moving together. The objective is to limit the cost of incorrectly predicting that a link is stable using minimal information. The purpose for defining this metric is to enhance the performance of routing algorithms and better facilitate mobility-adaptive dynamic clustering in ad hoc networks. In this scheme, an analytical framework is presented for extending the random-independent link availability model from [66] to capture the effects of correlated movement. Link availability is a probabilistic model which predicts the future status of wireless link. It is assumed that link status (up/down) is determined by a distance threshold.

Link availability is defined as the probability that there is an active link between two mobile nodes at time $t_0 + t$, given that there is an active link between them at time t_0. Let ${}_{m,n}(\tau)$ be the status of the link between nodes n and m at time τ : ${}_{m,n}(\tau) = 1$ if the link is up, and ${}_{m,n}(\tau) = 0$ if the link is down. Then the link availability is defined as:

$$ {}_{m,n}(t) \equiv P_r({}_{m,n}(t_0 + t) = 1 \mid {}_{m,n}(t_0) = 1) \qquad (2.1) $$

The salient feature of this model is that it uses knowledge of the mobility profiles of adjacent nodes, which can be learned periodically through acquisitions by Global Positioning System (GPS) [55] information, and computes the probability that the nodes remain within a threshold distance, or proximity of each other at any time t in the future – assuming a random-independent mobility pattern. They observe that the mobility of two nodes may be either independent or correlated; however, in most cases it will be uncertain which condition holds. Hence, an expression for the total link availability, ${}^{T}_{m,n}(t)$, can be given by equation 2.1, where, ${}^{i}_{m,n}(t)$ is the link availability when nodes n and m move independently, ${}^{c}_{m,n}(t)$ is the link availability when nodes n and m move in a correlated manner, and P_i is the probability that the two nodes are moving independently.

$${}^{T}_{m,n}(t) = {}^{i}_{m,n}(t)P_i + {}^{c}_{m,n}(t)(1 - P_i) \tag{2.2}$$

${}^{i}_{m,n}(t)$ is equivalent to the link availability as derived in equation 2.2. This metric assumes independent movement, consequently, with respect to the total availability it reflects the case when $P_i = 1.0$.

The idea is to initially assume that the endpoints of a link are moving independently, and to evaluate the link availability based on the independent model, ${}^{i}_{m,n}(t)$. Any link which survives longer than its expected time-to-failure begins to gradually transition into Associated Mode based upon an exponential smoothing function. If the transition is excessively abrupt, there is a greater chance of making an error. However, if the availability is sufficiently small, such that very few independent pairs survive this long, the overall risk of making an error is lower. Consequently, the rate of this function depends on the elapsed time since the expected time-to-failure and the magnitude of the Independent Mode (IM) availability. In this scheme, a node proximity model is developed that is designed to efficiently compute a metric that reflects future link stability based on minimum information.

2.9.6 Mobility Prediction Scheme Based on Link Expiration Time

A prediction mechanism [56] for the link expiration time (LET) between any two ad hoc nodes has been observed to enhance various unicast and multicast ad hoc routing protocols [67]. It is proposed that by exploiting the non-random movement patterns of a user they can predict the future state of the network topology providing transparent access during the time of topology change. By piggy-backing GPS based position [55] information on data packets, the link expiration time between any two nodes is estimated. If two nodes node i and node j at positions (x_i, y_i) and (x_j, y_j) are travelling at speeds v_i and v_j with moving directions θ_i and θ_j respectively with a transmission range r then the time period D_t during which they would stay connected is predicted as:

$$D_t = \frac{-(ab+cd) + \sqrt{(a^2+c^2)r^2 - (ad-bc)^2)}}{a^2+c^2} \tag{2.3}$$

where , $a = v_i \cos\theta_i - v_j \cos\theta_j$, $b = x_i - x_j$, $a = v_i \sin\theta_i - v_j \sin\theta_j$ and $d = y_i - y_j$. By predicting the LET's of all links of a route, the Route Expiration Time is given as the minimum of the LET values. This enables route reconstruction to take place prior to route failure. Reported results show that unicast protocols using mobility prediction such as the Flow Oriented Routing Protocol (FORP) [56] and Distance Vector with mobility prediction (DV-MP) [49] were the least affected by mobility maintaining delivery ratios of 0.9 for speeds of up to 70 km/hr. The On-Demand Multicast Routing Protocol with mobility prediction (ODMRP-MP) [27], performs better than its counterpart without mobility prediction offering a delivery ratio of 0.9 (i.e., 10% of packet loss) up to speeds of 70km/hr. The proposed method offers accurate prediction support to networks with simple mobility patterns with no sudden change in direction and constant speed. Since ad hoc networks find application in environments that prompt sudden changes in speed and direction of user movement, the assumption of constant user speed and direction is problematic for most scenarios in an ad hoc network [68].

2.9.7 Mobility Prediction Scheme for Link Availability Estimation

One critical issue for routing in mobile ad hoc networks (MANET) [65] is how to select a reliable path that can last longer since mobility may cause radio links to break frequently. To answer this question, a criterion that can judge path reliability is needed. The reliability of a path depends on the availability of the links constituting the path. However, how to measure link availability is the question raised in this scheme. In this scheme, a prediction-based link availability estimation is introduced and verified through computer simulations. This estimation algorithm can be used to develop a metric for path selection in terms of path reliability, which can improve the network performance as to be shown by the simulation results.

A probabilistic link availability model which can predict the future status of a wireless link is proposed in [66]. Prediction based link availability [65] proposes a metric for path selection based on path reliability. The method first allows a node to predict a continuous time period T_p during which a currently available link would last from t_0 assuming that both nodes of the link would keep their current movements unchanged in terms of both speed and direction. Then, the probability that the link will last to $t_0 + T_p$, (t_p) is estimated by calculating possible changes in the nodesŠ movements that might occur between t_0 and $t_0 + T_p$. The link availability estimation consists of Śunaffected T_p Š with node movements being unchanged and Śaffected T_pŠ with node movements being changed. Assuming that the mobility epoch (a random length interval during which node movement is unchanged) is exponentially distributed with mean λ^{-1} and node mobility is uncorrelated, link availability is given as,

$$(t_p) = {}_1(t_p) + {}_2(t_p) \tag{2.4}$$

where ${}_1(t_p)$ is the link availability estimation for the unaffected case and ${}_2(t_p)$ is the estimate for the affected case. Since the node movements are independent of each other and the exponential distribution is memoryless, ${}_1(t_p)$ is given as:

$${}_1(t_p) = [1 - E(T_p)]^2 = e^{-2\lambda T_p} \tag{2.5}$$

An accurate calculation of ${}_2(t_p)$ is challenging due to the difficulties in learning the changes in link status caused due to node mobility. A conservative prediction of link availability ${}_{min}(t_p)$ is proposed.

$${}_{min}(t_p) = \frac{1 - e^{-2\lambda T_p}}{2\lambda T_p} + \frac{\lambda T_p e^{-2\lambda T_p}}{2} \tag{2.6}$$

Based on the above estimation a routing metric based on (t_p), T_p is proposed. Reported results show that it offers improved network performance in terms of network loss, delay and goodput. In highly volatile environments it is possible that the mobility epochs (during which the mobility of the user is unchanged) can be very small. This will necessitate a large number of estimations increasing the control overhead. Also the accuracy of the link availability prediction requires that the original estimation of T_p by the nodes is accurate. This algorithm tries to predict the probability that an active link between two nodes will be continuously available for a predicted period, T_p, which is obtained based on the current nodeŠs movement.

2.10 Multicasting Based Routing Schemes to Improve Handoff Performance in Wireless Networks

Multicasting in wireless/mobile networks is defined as the ability to send data to a set of mobile units in one operation regardless of the mobility of the mobile units. This requires a routing tree to connect the source node to the multicast set. Mobile units change their access point over time. Hence, the multicast routes must be updated. This poses several challenges if efficient multicast routing is to be provided. Several attempts have been proposed [25] to build multicast trees that are suitable for wired networks. However, multicast routing in wireless mobile networks is not equally investigated.

The handoff problem in IP networks can be considered as a special instance of the broader mobility management problem, which arises because of the change in a mobile node's point of attachment to the network as it moves around. There are several mobility management solutions that have an emphasis on enabling network routing to and from alternate points of attachment for a mobile node. Design issues involved in mobility management are - changes or enhancements to the existing protocols, the resource requirement to support mobility, signalling involved during handoff, data loss incurred during handoff, and the mechanisms to overcome this data loss, etc. Also, every mobility management solution makes specific assumptions regarding whether the handoff is forward or backward and soft or hard, and regarding the capabilities from the underlying hardware. Different proposed mobility management solutions differ in terms of approaches to address these issues and of their assumptions.

Performance during handover is a significant factor in evaluating performance of wireless networks. Several protocols have been proposed to reduce handoff delay and improve the system performance. Some of these protocols include Mobile Multicast protocol (MOM) [69, 70, 71]. Multicast based Mobility (M&M) [72, 73, 74], Mobicast [75], RBMOM [76] etc. Most of the proposed protocols try providing an efficient scheme for multicasting or try to improve the handoff performance with the help from multicasting. The details of the proposed algorithms are discussed in detail in the sections below. Although most protocols proposed have discussed the improvements on performance by using multicasting there are not many of them that discuss the use of mobility prediction and multicasting in improving handoff performance. Although we discuss each of the algorithms in this chapter, the details of improvements of each these protocols by comparison with the algorithms proposed in the thesis are discussed in individual chapters.

2.10.1 Multicast Mobility Protocol - (MOM)

The IETF has proposed two approaches to provide multicast over Mobile IP. They are called remote subscription and bidirectional tunnelling. The remote subscription approach is simple, and works well if the mobile host spends a relatively long time at each foreign network, compared to the join and graft latencies. It has the further advantage of offering good (i.e., shortest path) routes for delivery of multicast datagrams to mobile hosts.

With bi-directional tunnelling, mobile hosts send and receive all multicast datagrams by way of their home network, using unicast Mobile IP tunnels from their Home Agents. This approach handles source mobility as well as recipient mobility, and in fact hides host mobility from all other members of the group [72, 77]. The drawbacks, however can be that the multicast delivery is far from optimal as it has to traverse the entire network and offers little scalability. To solve the tunnel convergence problem of bidirectional tunnelling, the MoM [69, 70] avoids the duplicate data being tunnelled to the common foreign agent (FA) using the Designated Multicast Service Provider (DMSP). The basic idea in MoM is to use the home agent functionality of IETF Mobile IP for delivery of multicast datagrams to mobile hosts, achieving scalability through the use of a designated multicast service provider (DMSP) optimisation per multicast group for each foreign network.

2.10.2 Range based MObile Multicast - RBMOM

The current multicast protocols on the Internet, DVMRP [27], MOSPF [78], CBT [79], and PIM [80], implicitly assume static hosts when building a multicast delivery tree. They do not consider the dynamic member location. Reconstructing the delivery tree every time a member moves will involve the overhead; yet leaving the tree unchanged can result in inefficient, incorrect, or even failure of multicast datagram delivery. RBMoM (Range-Based Mobile Multicast) was proposed for efficiently supporting multicast for mobile hosts on the Internet.

RBMoM [81] intends to trade off between the shortest delivery path and the frequency of the multicast tree reconfiguration by controlling the service range of the multicast home agent (MHA). The RBMoM is a hybrid protocol of the remote subscription and the bi-directional tunnelling. This protocol intends to trade-off between the tunnelling path and the frequency of multicast tree reconstruction and reduces the cost of tunnelling by employing the idea of the service range and a Multicast Home Agent (MHA). The simulation results show that RBMoM can adapt to the fluctuation of both host movement and the number of mobile group members, and outperforms the current two Mobile IP multicast solutions. However, this protocol does not provide a certain criterion for determining an optimal service range. In addition, when a mobile host also moves to a foreign network which is out of its service range and does not join a multicast group, the problem of multicast packet loss happens during the time required to join the multicast group corresponding to the time of multicast tree reconstruction. The movement of multicast source has not been considered in literature.

2.10.3 Multicast-based Mobility - (M&M)

Performance during handover is a significant factor in evaluating performance of wireless networks. IP-multicast [82, 83] provides efficient location independent packet delivery. The receiver-initiated approach for IP-multicast enables receivers to join to a nearby branch of an already established multicast tree. Multicast-based mobility (M&M) uses this concept to reduce latency and packet loss during handover.

In multicast-based mobility, each mobile node (MN) is assigned a multicast address. The MN, throughout its movement, joins this multicast address through locations it visits. Correspondent nodes (CN) wishing to send to the MN send their packets to its multicast address, instead of unicast. Because the movement will be to a geographical vicinity, it is highly likely that the join from the new location, to which the mobile recently moved, will traverse a small number of hops to reach the already-established multicast distribution tree. Hence, performance during handover improves considerably.

As the MN moves, it joins to the assigned multicast address through the new access router (AR). Once the MN starts receiving packets through the new location, it sends a prune message to the old AR to stop the flow of the packets down that path thus, completing the smooth handover process. In spite of its promise, many issues need to be addressed to realise multicast-based mobility in today's Internet. These issues include scalability, multicast address allocation, multicast deployment and security. The M&M mechanism proposes a paradigm for multicast-based micro mobility, where a visiting mobile is assigned a multicast address while moving within a domain. The multicast address is obtained using algorithmic mapping, and handover is achieved using multicast join/prune mechanisms.

Furthermore, the adjacency can be established based on the adjacency of the radio coverage area of the Serving AR (SAR) in the M&M mechanism, it is similar to the case of cellular wireless networks. The serving AR is called the Head of the Coverage Access Router Set (CAR-set). Thus, there is a unique CAR-set defined for every AR. The memberships of CAR-set are dependent upon a number of factors, such as the handover types and the prediction accuracy, and so on. The membership is determined by the CAR-set prediction algorithm. One major contribution of M&M is the introduction of a handover framework, using the CAR-set protocols, that may be tuned to perform efficient proactive, reactive and gap handoffs especially.

2.10.4 MOBICAST

The proposed multicast scheme known as MobiCast [75], is suitable for mobile hosts in an internet-work environment with small wireless cells. This scheme adopts a hierarchical mobility management approach to isolate the mobility of the mobile hosts from the main multicast delivery tree. To send a multicast packet, the mobile host encapsulates the multicast packet and sends it to the domain foreign agent. The domain foreign agent decapsulate the multicast packet and sends it out on behalf of the mobile host. The main function of the domain foreign agent is to subscribe to the requested multicast group and forward the requested multicast packets to the mobile host using another multicast address known as the translated multicast address. The translated multicast address will be unique within the domain and corresponds to this requested multicast group. The base station receives the multicast packets by subscribing to the translated multicast group and forwards the received multicast packets to the mobile host. As long as the mobile host remains within the domain of the domain foreign agent, the mobility of the mobile host is hidden and no re-computation of the main multicast delivery tree is needed.

This scheme minimises disruptions to the multicast session due to handoffs of mobile group member by organising physically adjacent cells into Dynamic Virtual Macro-cells (DVM). The scheme for wireless networks informs the other member base stations in its DVM to subscribe to the same translated multicast group. While only the serving base station actively forwards multicast data to the mobile host, the other base stations in the same DVM buffers recent packets and quickly forwards those to the mobile host should a handoff occur. This provides short handoff latency, and the use of buffers at the BSs reduces packet loss due to handoff. It also eliminates multicast group join and graft latencies since the new base station has already subscribed to the multicast group prior to the handoff. Hence, the disruptions to the multicast session due to handoffs of mobile host group members are minimised.

2.11 Summary

In this chapter, we have introduced issues associated with mobility in wireless networks. With future wireless networks targeting higher data rates and value added service, there is a need for signal strength prediction for the success of the future wireless networks. In this chapter we have discussed some mobility prediction schemes proposed for both cellular networks as well as ad hoc networks. To efficiently solve the problem of movement detection many prediction algorithms have been proposed based on patterns, movement history, trajectory prediction, location and mobile positioning. However, these schemes tend to have significant weaknesses and degraded performance during handoffs. These assumptions in cellular networks do not apply to MANETs (mobile ad hoc networks), which can have several network topologies over time and multihop communications between mobile devices. So the schemes related to these networks have been discussed in this chapter. Furthermore, even multicast assisted approaches have also been discussed. Unfortunately, all the prior work in the area of mobility prediction and multicasting techniques to improve handoff have not provided a scalable solution that addresses both these issues simultaneously. However, these schemes tend to have significant weaknesses and degraded performance during handoffs.

We recognise that these proposed models do not consider any RSSI measurements in particular to the maximum extent which is important to achieve accurate mobility prediction for future generation networks. Fine tuning of any errors in accurate prediction mechanisms is also important to prediction algorithms. Most of the handoff algorithms are based on the bit-error rate (BER) and signal strength measurements are averaged over time to remove rapid fluctuations due to the multipath nature of the propagation environments. The algorithm proposed in this thesis will utilise the RSSI value to enable handoff decisions by tracking the signal strengths in real-time. The mobility prediction algorithm proposed in this thesis is based on a theory called "Grey theory" which is discussed in the next chapter. Although the literature has proposed several methods for RSSI measurements [30, 56], they are not used for prediction nor do they have appropriate handoff performance. The methods involving hysteresis and threshold values to determine the occurrence of handoff still have some flaws and need considerable improvement to support future generation networks.

In addition, very little work has addressed the problem of reserving the right level of resources in the next cell with the combination of mobility prediction and multicasting during a handoff. The systems that do attempt to perform this combination use a naive approach of using *"hints"* [29] for multicasting during handoff. This method provides unsatisfactory handoff performance as there is no discussion about how these *"hints"* are obtained. In order to support existing protocols and multimedia applications, we need handoffs that complete quickly and do not affect end-to-end performance. As part of exploring support for this mobility prediction handoff supported by multicasting, we present alternative routing update algorithms for future generation networks. In addition, we also describe methods to perform the selection of potential base stations or access points to transfer quickly and in advance of the actual user movement between wireless cells. This chapter has concluded with a literature survey of all the major proposed schemes, which highlights the contributions and drawbacks which further clarifies the contribution made in this thesis. Our purpose is to provide an overview of handoff from the point of view of the performance benefits available from different mobility prediction and multicasting schemes. We present a broad survey of the technical issues involved, with details of each of the prediction schemes proposed. The schemes are highlighted with issues to assist the reader in developing a good understanding of those issues and to understand the place of RSSI measurements in cellular systems.

Chapter 3

Prediction Methodology

3.1 Introduction

Handoff performance in wireless networks is based on the relative importance given to handoff when the mobile user switches from one base station to the other. Upon initial evaluation and literature review, mobility prediction schemes in handoff procedure were found to be very critical in the handoff performance. The handoff procedure is typically based on the received signal strengths from the base station. There exist several models schemes and algorithms for handoff procedure which is based on the RSSI (Received Signal Strength Indicator) values as proposed in the literature [1, 3, 19]. These published methods are regularly based on hysteresis and threshold methods. In this chapter, we propose a methodology which is mobility prediction based on the prediction of received signal strengths from the base station. With a good prediction of the RSSI values it is possible to achieve good handoff performance. Our proposed algorithm aims to predict the received signal strengths and the simulation results prove that accurate prediction can be achieved using our proposed model.

In order to obtain an efficient algorithm for mobility prediction, a suitable model and parameters has to be chosen which will result in superior prediction accuracy to improve handoff performance. In our approach, we use RSSI values not only to predict the movement of the mobile node but also efficiently reserve sufficient resources during handoff. As mentioned before, the parameters chosen for prediction in our methodology use the RSSI values and the methodology incorporates real-time requirements. For efficient handoff, our model should reflect the behaviour and properties of the mobile network as closely as possible. It will be demonstrated that our model suits these requirements very well, by providing sufficient accuracy and also reserving appropriate levels of network resources to meet the forecast demands and required performance objectives. The work related to the network resource reservation is discussed in the subsequent chapters.

In this chapter, our handoff prediction approach is based on Grey theory as it has the benefit of reducing overheads in wireless cellular networks because it requires little data and processing effort. The Grey model will be used to predict RSSI values from the base stations. The decision to handoff is based on predicted RSSI values, the latest RSSI value and the $(n-1)$ past values. To provide a foundation for this prediction methodology, we start this chapter with the construction of our Grey prediction model. The Grey prediction model performs efficiently for a given set of base stations and our simulation results prove that reasonable accuracy in the prediction can be achieved when compared with other alternative approaches in real-time situations as discussed in [56] and references therein.

In the context of good handoff performance, mobility prediction plays a vital role as it provides the means to enable the mobile node select its future serving base station. In section 3.3, we outline the step by step process for construction of the model. We also describe the various assumptions that underpin this construction. In section 3.4, we describe the construction of a simulation example together with its associated parameters. Our example considers two base stations separated by a fixed distance and the design method is used under log-normal fading environment conditions. Section 3.5 compares the performance of the proposed prediction theory model with traditional estimation techniques such as the moving average filter (MAF) and exponential moving average filter (EMAF). The result shows that the Grey model approach not only has excellent performance results but also has a very short calculation time. The processing of the data needs only a few data points to get a prediction and is suitable for use in real-time systems due to its quick response time. The Grey prediction model can be shown to track the original data very closely with only a few errors. However, we propose to eliminate these few errors by proposing a hybrid model in subsequent chapters in order to further improve the accuracy of the method. Finally, we describe the performance results and comparison to other methods followed by the conclusions.

3.2 Motivation and Related work

In recent years, many investigations have addressed handover algorithms for cellular communication systems. Several approaches were proposed based on threshold values, hysteresis margins, dwell timers and the combination of two or more of these methods [3]. The relative signal strength with threshold approach allows a handover only if the current base station signal strength drops below a defined threshold level [84, 85]. The fluctuations of signal strength associated with shadow fading cause a call to be repeatedly handed over back and forth between neighbouring BS's, in what is known as the "ping-pong effect". Instead of using a threshold level, the method involving relative signal strength with hysteresis allows a handover only if the new base station's signal strength is stronger than the current base station by a hysteresis margin. With the hysteresis margin approach, a handoff is allowed to the base station which is stronger than the current one by at least a fixed or time varying hysteresis margin. The dwell timer approach restricts or limits the time between 2 handovers. Each approach is based on the relative signal strength criterion, and induces a handover to a base station whose signal strength is stronger than that of the current base station. This criterion may generate an unnecessary handover when the current base signal is still strong enough. While this approach reduces ping-pongs, it suffers from the initial unnecessary handover problem. To provide a solution for this, a combination of the threshold level with the hysteresis margin was proposed to create a strategy called the "relative signal strength with hysteresis and threshold" approach. While each of these have their advantages and disadvantages, another method is to use prediction of the signal strength.

However, these models, do not lend themselves easily to the development of efficient algorithms for handoff performance due to their associated high levels of complexity. With accurate prediction of signal strength, it is possible to know the exact position of the mobile node and how far is it from the base station. Using prediction, the next base station signal strength can be known; so as to reserve only the required resources for the mobile node. Note that, for efficient reservation of resources during handoff, not many schemes are proposed with a combination of mobility prediction and multicasting. Our research goal is to explore this area of research which also will include mobility prediction to select the right base station, or access point, and apply multicasting techniques to reserve the appropriate level of resources. Therefore, we shall determine the best routes for a mobile node, including resource reservation, with constraints and objectives which will be discussed in detail in chapters 6 and 7. By using prediction, the overheads due to handoff can be reduced in the cellular network. However, it should be noted that the decision to handoff is not only based on the received signal strength RSSI value measurements but also on the amount of available resources in the contending cells which is further discussed in later chapters. Based on the observations made, we suggest a novel method to use Grey prediction methodology to accurately track the signal strengths.

3.3 Grey Theory

Many systems, such as social, economical, agricultural, industrial and biological systems have used Grey theory. In this theory, it is through appropriate organisation of the original sequence of data, a high degree of accuracy can be achieved. The name of Grey systems was chosen based on the colours of the subject or parameter under investigation. For example in control theory the darkness of colours has been commonly used to indicate the degree of clarity of information. One of the most well accepted representations is called a “Black Box", which stands for an object with its internal relations or structure totally unknown to the investigator. Here, the theory uses “black" to represent unknown information and “white" to represent known information and “Grey" for that information which is partially known and partially unknown. According to the theory [2], through reconstruction and different combinations of systems, we can often achieve a higher level of efficiency and accuracy.

The main difference between Grey concepts and fuzzy concepts are described with the properties of internal meanings and external extensions of the subject under investigation. Grey systems emphasise the objects of definite external extensions and vague internal meanings, while fuzzy mathematics mainly studies the objects with definite internal meanings and vague external extensions. However, Grey concepts and stochastic concepts are considered to be the same [2, 86]. Grey systems theory and stochastics have completely different methods of solution and logical thinking. Grey numbers, Grey elements and Grey relations are the main objects of research in Grey systems theory. So, the entire theory of Grey systems relies on the

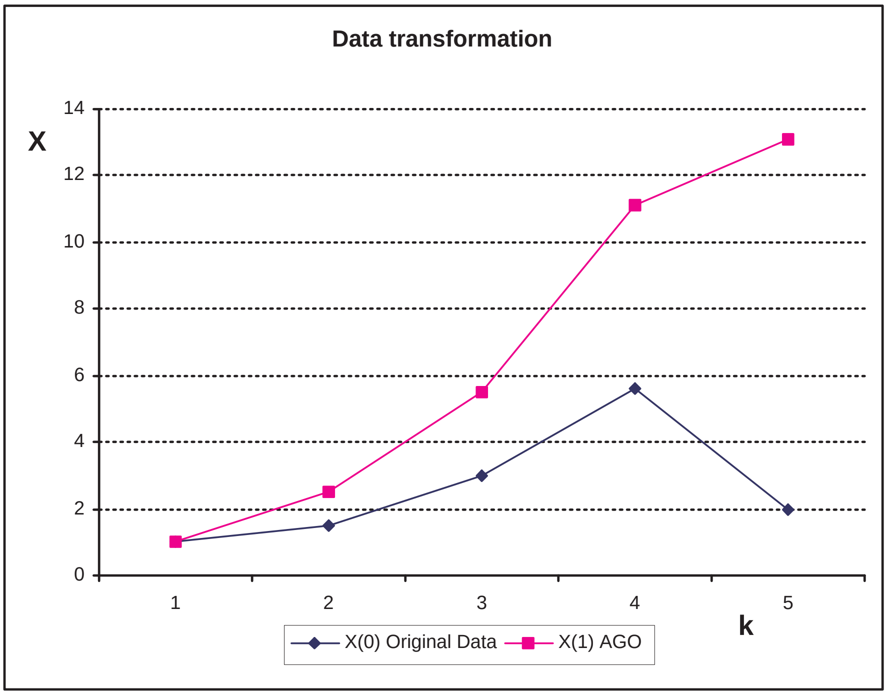

Figure 3.1: Data transformation for AGO operation in the Grey model.

foundation of Grey numbers and their operations: Grey matrices and Grey equations. Some examples of research tasks with Grey systems are: control problems in industry, Grey systems analysis, modelling, forecasting and controlling intrinsic characteristics of systems. In Grey systems theory, by organisation of raw data we can remove randomness, which is followed by series of operations for construction of the Grey model. The organisation and operations helps in building the Grey model for predicting the future data. The prediction method of Grey systems theory has the following features: It is composed of a dynamic model expressed by an ordinary differential equation. The right hand side of the differential equation can have various factors of fluctuation (mainly input factors). The algorithm is relatively easy to understand and requires only a few calculations. The Grey system prediction method can be summarised as follows.

- Observed data are converted into a new data series by a preliminary transformation called AGO (Accumulated Generation Operation).
- Parameters of the Grey Model are determined by the least squares method using a new data series.

The AGO transformation has the following features:

- It transforms the original data into data with a exponential curve which can be illustrated as shown in Fig. 3.1.
- It reduces noise in the input data.
- It is relatively easy to obtain the differential equation for the Grey model.

The theory of Grey systems can deal with incomplete and uncertain problems by using technologies such as the Grey model, Grey prediction, Grey relational analysis and Grey decisions. The Grey system was developed in 1982 and was used for systems that have very little data from which to analyse or predict future data [2]. The system was widely used in weather prediction and control system applications. In general, Grey theory can be applied to any system with incomplete information which is the fundamental reason for choosing this theory in our prediction of signal strength. The theory relies on objects that are known as Grey numbers, together with operations, matrices and equations. An objective of Grey theory is to treat raw data using suitable laws to transform it into meaningful data and this is the first thing addressed by the methodology. The contention is that randomness can be removed by using the Accumulated Generating Operation (AGO) on the raw data to make it meaningful. For example, consider the following sequence which does not follow any specific rule:

$$X^{(0)} = \{1, 1.5, 3, 5.6, 2\} \tag{3.1}$$

After we apply the Accumulated Generating Operation, it can be seen that there is a steady increase in the numbers which are represented.

$$X^{(0)} = \{1, 2.5, 5.5, 11.1, 13.1\} \tag{3.2}$$

The accumulated generating operation is a method to "whiten" the raw data and has a very important role to play in Grey theory. In our model, for mobility prediction, as little as four measurements of the signal strength are required to enable a useful prediction to be made. In the following sections, we describe the construction and modelling of the Grey theory which is used for prediction and provide a simulation for our particular mobility prediction scenario.

3.3.1 Grey Model

In the following section, we present the Grey model that is used in the mobility prediction model. As previously described, the model is used in the prediction of received signal strengths (RSSI) from the base station. The Grey model [2, 4, 87] uses a sequence of raw measurements that are generated by the system under study. The approach is to convert this raw data into a series of meaningful data values, which is achieved using the Accumulating Generating Operation (AGO) that is central to the operation of Grey system theory. The Accumulated Generating Operation is carried out in the following way to create a new series: Let the sum of the first and second elements in the measurement set data be the second element of the new series. Let the sum of the first, second and third elements be the third element of the new series and so on. The derived new series is called the Onetime Accumulated Generating series of the original series. Its mathematical relations are presented in Eqs.(3.3 - 3.6). Suppose that the original series is given by:

$$X^{(0)} = \{X^{(0)}(0), X^{(0)}(1), \cdots\cdots, X^{(0)}(n)\} \tag{3.3}$$

which represent the measurements of the received signal strengths obtained from the system.

Then the Onetime Accumulated Generating series is

$$X^{(1)} = \{X^{(0)}(0), X^{(1)}(1), \cdots\cdots, X^{(1)}(n)\} \tag{3.4}$$

where,

$$X^{(1)}(k) = \sum_{i=0}^{k} X^{(0)}(i) \quad k = 1, 2 \cdots n \tag{3.5}$$

The superscript of (1) in equation (3.5) in $X^{(1)}(k)$ represents the Onetime AGO which is denoted as 1-AGO. If the superscript is (r) then it represents r times AGO and is usually denoted as r-AGO. The elements of the r-AGO series are:

$$X^{(r)}(k) = \sum_{i=0}^{k} X^{(r-1)}(i) \quad k = 1, 2 \cdots n \tag{3.6}$$

The purpose of AGO is to reduce the randomness of the series and increase the smoothness of the series. The following is a first order differential equation model with one variable, which will be denoted by $GM(1,1)$.

$$X^{(0)}(k) + az^{(1)}(k) = b, \quad k = 1, 2 \cdots \tag{3.7}$$

and $X^{(0)}(k)$ is a Grey derivative which maximises the information density for a given series to be modelled.

$$z^{(1)}(k) = \frac{X^{(1)}(k) + X^{(1)}(k-1)}{2}, \quad k = 1, 2 \cdots \tag{3.8}$$

In Grey systems theory, based on the understandings of differential equations the concept of Grey derivatives is introduced to establish models similar to differential equations for a sequence of discrete data. The theory [2], assumes the following differential equation,

$$\frac{dx}{dt} + ax = b \tag{3.9}$$

where, $\frac{dx}{dt}$ is called the derivative of function x, x is the background value of $\frac{dx}{dt}$, and a and b the parameters. Therefore first order differential equation consists of 3 parts: derivative, background value and some parameters.

The whitened differential equation model can be expressed as:

$$\frac{dX^{(1)}(t)}{dt} + aX^{(1)}(t) = b \tag{3.10}$$

Where a and b are constants to be determined. a is known as the developing coefficient and b is known as the Grey input. Based on the ordinary least squares method, we have

$$\hat{a}^T \equiv \begin{bmatrix} a & b \end{bmatrix}^T \tag{3.11}$$

$$\begin{bmatrix} a & b \end{bmatrix}^T = (B^T B)^{-1} B^T Y_n \tag{3.12}$$

where B is known as the accumulated data matrix and Y_n is a constant vector.

$$\mathbf{B} = \begin{bmatrix} -\frac{1}{2}\left[X^{(1)}(1), X^{(1)}(2)\right], & 1 \\ \vdots & \vdots \\ -\frac{1}{2}\left[X^{(2)}(1), X^{(3)}(2)\right], & 1 \\ & \\ -\frac{1}{2}\left[X^{(1)}(r-1), X^{(1)}(r)\right], & 1 \end{bmatrix}$$

$$Y_n = [X^{(0)}(2), X^{(0)}(3) \cdots X^{(0)}(r),]^T \tag{3.13}$$

By solving a, b and the differential equation, the time response sequence of the

$GM(1,1)$ is given by

$$\hat{X}^{(1)}(k+1) = \Big(X^{(1)}(0) - \frac{b}{a}\Big)e^{-a(k)} + \frac{b}{a}, \quad k = 1, 2 \cdots n \tag{3.14}$$

let $X^{(1)}(0) = X^{(0)}(1)$ then,

$$\hat{X}^{(1)}(k+1) = \Big(X^{(0)}(1) - \frac{b}{a}\Big)e^{-a(k)} + \frac{b}{a}, \quad k = 1, 2 \cdots n \tag{3.15}$$

Then the value of $\hat{X}^{(0)}(k+1)$ can be given by:

$$\hat{X}^{(0)}(k+1) = \hat{X}^{(1)}(k+1) - \hat{X}^{(1)}(k) = (1 - e^{-a})\Big(X^{(0)}(1) - \frac{b}{a}\Big)e^{-a(k)} \tag{3.16}$$

where, $\hat{X}^{(0)}(k+1)$ denotes the prediction of $X^{(0)}(k+1)$ at time $k+1$.

3.4 Modelling Assumptions

In this section, we present the background to a simulation study that aims to check the validity of the predictive method that we presented in the previous section. The simulation study is underpinned by a series of assumptions that we shall now describe in some detail. The discussion includes the model chosen and the performance dependent parameters.

3.4.1 Simulation Parameters and Modelling

In this simulation study, two base stations A and B were selected which were separated by D metres. The mobile device moves from one cell to another with a constant velocity and the received signal strength is sampled at a constant distance (in metres). Here, the mobile node moves from one base station to the other with constant speed. The signals from the base stations are affected by two major factors: path loss and log-normal fading. Rayleigh fading is not taken into account as it is assumed that any rapid fluctuations are being averaged out. The received signal strengths from the current base stations to the target base station are sampled at distances Kd_s. The simulation also includes slow fading. The received signal strengths a_t and b_t (in dB) when the mobile is at a given distance kd_s are given by:

$$a_t = K_1 - K_2 \log kd_s + u_t \tag{3.17}$$

$$b_t = K_1 - K_2 \log (N - k)\, d_s + u_t \tag{3.18}$$

where $N = D/d_s$. The parameters $K_1 = 0$ and $K_2 = 30$ in dB which are typical of an urban environment accounting for path loss. K_1 is the signal strength at distance $d = 1$, and K_2 is the path loss component. Since threshold levels are not considered, the received signal strengths depend on the difference in received signal strengths. The simulation parameters used for the movement detection are as shown in Table 3.1. In the following sections, we shall briefly describe some alternative

Number of Base Stations	2
Trajectory	Straight Path
Sampling distance	10 m
Distance between base stations	2000 m
Path loss (K)	30 db
Transmitter power	0 dB
Fading Process	Lognormal fading
Standard Deviation (u_k)	8dB

Table 3.1: The simulation parameters used for the prediction model

methods of predicting the signal strength in the above scenario and make some detailed comparisons of these methods with the proposed prediction approach based on Grey theory.

3.5 Comparison with the Moving Average Filter

In this section, we present a brief overview of moving average filters considered for comparison with the Grey model. The choice of these moving average filters is due to the fact that they are widely used for forecasting. It is a simple mathematical technique used to eliminate irregularities in data and give the real trends in a collection of data points.

Moving average filters have their historical roots in the exponential smoothing methods of the 1950's [Brown (1959), Holt (1957), and Winters (1960)] [88, 89, 90]. The most basic of these methods is simple exponential smoothing. They were suggested procedures for both stationary and non-stationary data. Forecasts are usually made using models that give the most weight to recent observations, and negligible weight to the distant past. According to [89] the smoothing parameter should not be too big for forecasting purposes. In principle, the main difference in the simple moving average and weighted moving average methods are as follows: In a simple moving average, an *n-period* moving average is the average value over the previous n time periods. As we move forward in time, the old time period is dropped from the analysis. On the other hand, in weighted moving averages, an *n-period* weighted moving average allows you to place more weight on more recent time periods by weighting those time periods more heavily. The above concepts are detailed in several papers [91, 92].

Filtering is a procedure used to reduce the noise of a measured signal [93]. There are many different ways to reduce noise, one of them being "averaging". Moving average techniques have been used extensively in a wide variety of applications, and continues to be used today. It was first suggested by C.C. Holt in 1957 and was used for non-seasonal time series having no trend. The moving average filter can be considered as a simple Low Pass filter commonly used to smooth an array of sampled data [88]. The data points are considered to be an array of elements. Moving average models generate a new series by computing moving averages of the original series. They are oriented primarily by removing the irregular components or isolating the trend-cycle components of a time series. The newly generated series is a “smoothed" version of the original series. The procedure adapted by the moving average filter to determine the amount of filtering involves a smoothing factor. The smoothing factor is controlled in such a way that it can be increased or decreased to specify the number of data points or samples that will span the filter. The moving average filter involves a window of a certain size moving along the entire data set with one element at a time [90]. The middle element of the window is replaced with the average of all elements in the window. However, it is important to remember the value of new elements and not make the replacement until the window has passed. This must be done since all averages shall be based on the original data in the data set.

The algorithm involves taking two or more data points, adding them and dividing their sum by the total number of data points added. Then the first element is replaced with the average computed and the steps are further repeated till the end of the data set is reached. Depending on specific requirements, we use different types of averaging filters, with one of them being a moving average filter which considers all the data points to be equally important. The average at the k^{th} instant is based on the most recent set of n values which is given by:

$$\bar{x_k} = \bar{x}_{k-1} + \frac{1}{n}\Big(x_k - x_{(k-n)}\Big) \tag{3.19}$$

Hence, from the above equation we note that we need to store the values of $x_{(k-n)}$

data points which require up to n storage locations. Similarly, we can obtain the equation for weighted moving average filter which is given by:

$$\bar{x_k} = \bar{x}_{k-n} + \frac{1}{n}\Big(x_k w_k - x_{(k-n)} w_{(k-n)}\Big) \tag{3.20}$$

where w_k and $w_{(k-n)}$ are the weighting factors applied for data points x_k and $x_{(k-n)}$. In a weighted moving average filter it places importance to the data on which there is more weight, usually the recent data.

3.6 Exponential Weighted Moving Average Filter

In this section, we present the exponential moving average filter that will also be used in a comparison with the Grey model for tracking of the signal strengths. The exponential weighted moving average filter is arguably the most commonly used noise reduction algorithm in the process industries. The exponential moving average has the advantage of reduced data storage over moving average methods which stimulated the adoption of the exponential smoothing technique. They have become very popular because of their (relative) simplicity and their good overall performance. However, it has its roots in electrical circuitry and are used to produce smooth electrical signals. Among the simplest methods is ordinary exponential smoothing, which assumes no trends and no seasonality. There are many variants proposed for the exponential smoothing among which are, Holt's linear trend method and Winters extensions which add a trend and a seasonal component. Some of the other techniques proposed are by Brown (1963), who set up forecasting procedures in a regression framework and used the method of discounted least squares [89]. Others include forecasting by Kalman filter techniques, Binomial filter, Gaussian filter etc. Although many newer models based on an underlying exponential smoothing approach have been proposed, we shall just consider the simple exponential moving average filter for our study.

To estimate the current level of a series of observations one of the simplest approaches is to use the sample mean. The purpose of estimating the level is to use this as the basis for forecasting future observations as it is more sensible to put more weight on the most recent observations. For the exponential moving average in its simplest form, only one old value has to be remembered. The difficulty lies in the proper choice of α, the exponential weight. The filter estimates the data set quite well but the estimation follows real changes too slowly. The moving average models described are satisfactory for treatment of a time series which are divided into data sets. These provide approximately the same mean for each data set and the series. By changing means from one data set to the next has the presence of a gradual trend factor. But, if there is an outside change to the data set, the data set for the series will exhibit significantly different “means" which appear to have little or no relationship to each other. Exponentially weighted moving average (EWMA) models provide some ability to adapt to the changing-mean phenomenon [90]. Each newly computed mean for EWMA is based on the previously computed mean and the current observation. The smoothing constant is chosen with value $0 < \alpha < 1$. This permits division of the total weight between the current observation and the mean of all previous observations.

The EWMA process will smooth seasonal and irregular variations of an original series but may result in loss of some of the trend. For forecasting purposes, a series with a trend pattern tends to lag below the trend and it is important to alter the EWMA model to avoid this trend loss. The exponential moving average filter places greater importance on more recent data by discounting the older data in an exponential manner. The form of the exponential moving average filter is as follows:

$$\bar{x_k} = \alpha \bar{x}_{k-1} + (1 - \alpha) x_k \qquad (3.21)$$

The value of the filter constant α monitors the degree of the filtering action. We can also notice that the calculation of $\bar{x_k}$ does not require the storage of values of x. It can be shown that as we keep increasing the RHS of the equation, we can see that the older values x_i are weighted by increasing powers of α. The graphs show the weighted moving average and the exponential weighted moving average plotted in section 3.7.

3.7 Results

All the algorithms discussed above were implemented using Matlab Version 6. We performed various tests and derived the expression for each of the Grey model, weighted moving average filter and the exponential weighted moving average filter. For each test that was performed, the parameters shown for the simulations were the same as shown in the section 3.4. The simulations included both lognormal fading and slow fading as required by any mobility model. The simulation was custom built and kept simple for the prediction model. To be able to easily understand the behaviour of the mobility model, the signal strengths generated followed the standards proposed in [94]. As previously mentioned, the model consisted of two base stations separated by a distance of D metres.

For the evaluation of accuracy, we obtained a set of graphs which include results obtained by using the Grey theory, the moving average filter and the exponential moving average filter used for forecasting. The preliminary results of the Grey prediction in Fig. 3.2 show a plot of the actual values of received signal strength and corresponding predicted values. It can be seen that the Grey model tracks the curve – but with some error. The Grey model does not predict large variations in the input data. The plotted absolute errors show the deviation of the actual and the predicted values in Fig. 3.3. To compare the performance of the Grey prediction model, we have compared it with a moving averages filter and the exponential moving average filters in Figs. 3.4, 3.5, 3.6 and 3.7. The moving average filter largely depends on the selection of appropriate weights as it averages the value. Large variations in the Grey prediction values are shown in the graph displayed in Fig. 3.3 by plotting the absolute error.

From the simulation results, we can find that the Grey model provides better prediction accuracy and tracks the curve more accurately than the traditional moving average filters. From the graphs it is clear that the prediction accuracy is less than one decibel in value. Simulation has been done under log-normal fading of standard deviation ranges of $(2 - 10dB)$. The Fig. 3.1 shows the relation between the actual and the predicted data. The Grey prediction data can also be seen to track the actual data very closely.

Based on our results, it is very clear that the Grey model outperforms the traditional weighted moving average filters (MAF) and the exponential moving average filters (EMAF). If we observe the two curves namely, actual curve and the predicted curve in Figs. 3.4, 3.5, 3.6, it is clear that although the curves have the same shape, they lag slightly behind the actual trend. The reason for this is that the weighted moving average filter places equal emphasis on all data points unless specified. Thus a value in the past will have the same influence as more current measurements. This may be a desirable feature when the mean value of the measurement is almost constant but not when the measurement has a trend. In dynamic systems however, the most current values tend to reflect better on the state of the process. A filter that places more emphasis on the recent data such as the EMAF will be more useful.

Although the EWMA process smoothes seasonal and irregular variation out of an original series, it causes loss of some of the trend variation. The close observation of the graph reveals that the exact trend is not followed by the EMAF filter. The EMAF has a much better performance than the weighted moving average filter since it provides some ability to adapt to a changing-mean phenomenon. When we observe the Figs. 3.4, 3.5, 3.6, 3.7 the EMAF definitely shows a much better performance than the MAF. The plot of the absolute error for EMAF in Fig. 3.8 shows that the performance is not as good as the basic Grey prediction model.

Overall, the prediction based on Grey methodology is much better than any of the averaging filters chosen for this study. This is due to the fact that the procedure employed by the Grey technique converts the initial random series into a series that is meaningful data. Another important point to mention is that the main requirements for using Grey prediction is that it needs considerable and good behaviour of distribution of data. The characteristic of the Grey model is that we can utilise only a few data by the Accumulated generation operation (AGO) to establish the prediction model. Furthermore, the Grey model is seen to track the peaks consistently as shown in Fig. 3.2.

3.8 Summary

In this chapter, we proposed the use of the Grey model for prediction of signal strengths since it has better prediction accuracy. The proposed technique requires minimal data points and the errors obtained are comparatively much less than other methods tested, such as the moving average filters and exponential moving average filter. This chapter addressed the problem of tracking data i.e., the signal strength from a base station which is useful to improve the performance in wireless networks. The Grey model is modelled according to the mobility requirements of the wireless

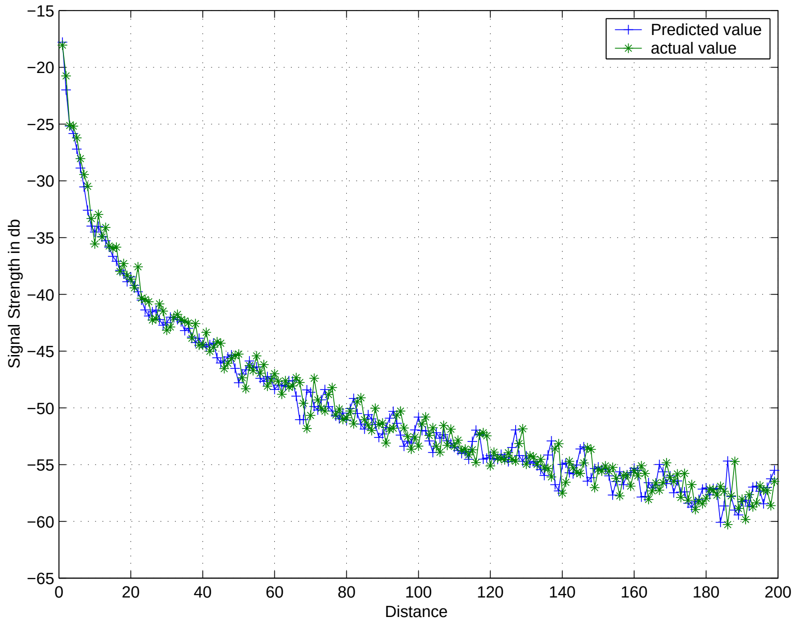

Figure 3.2: The received signal strength tracked by the Grey model.

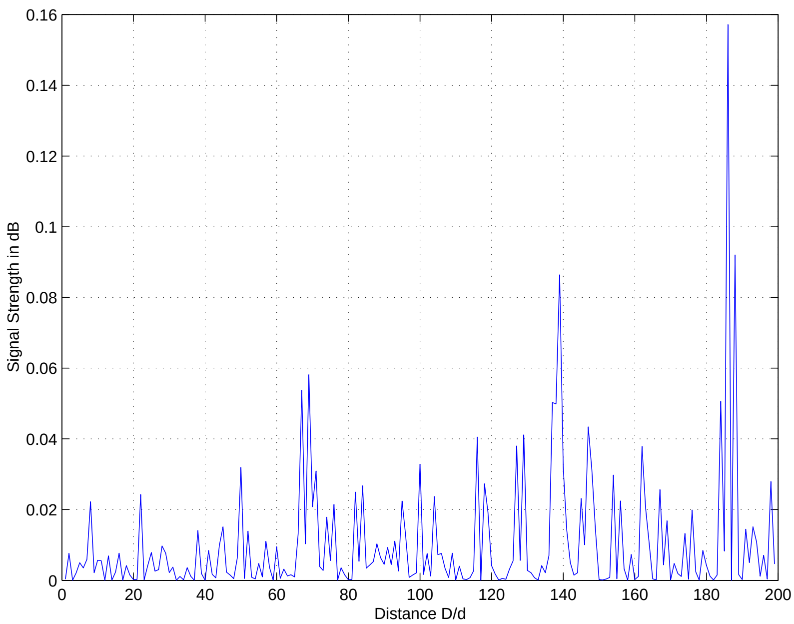

Figure 3.3: The absolute error from the Grey model.

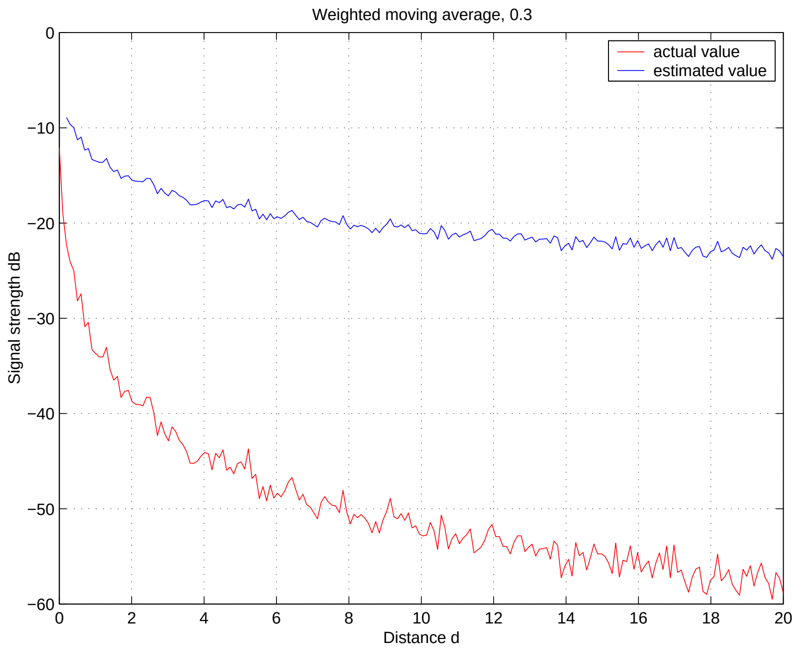

Figure 3.4: Weighted moving average filter with weight of 0.3 model.

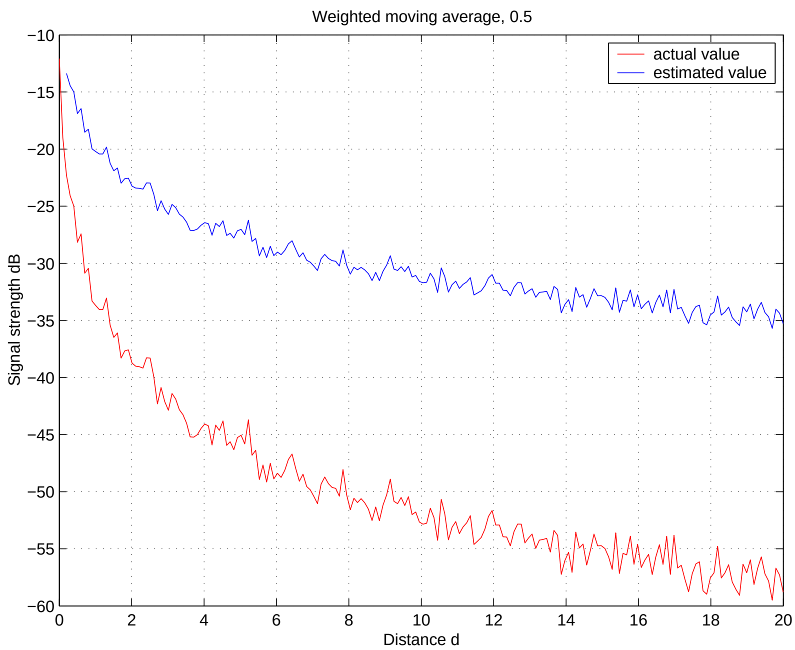

Figure 3.5: Weighted moving average filter with weight of 0.5 model.

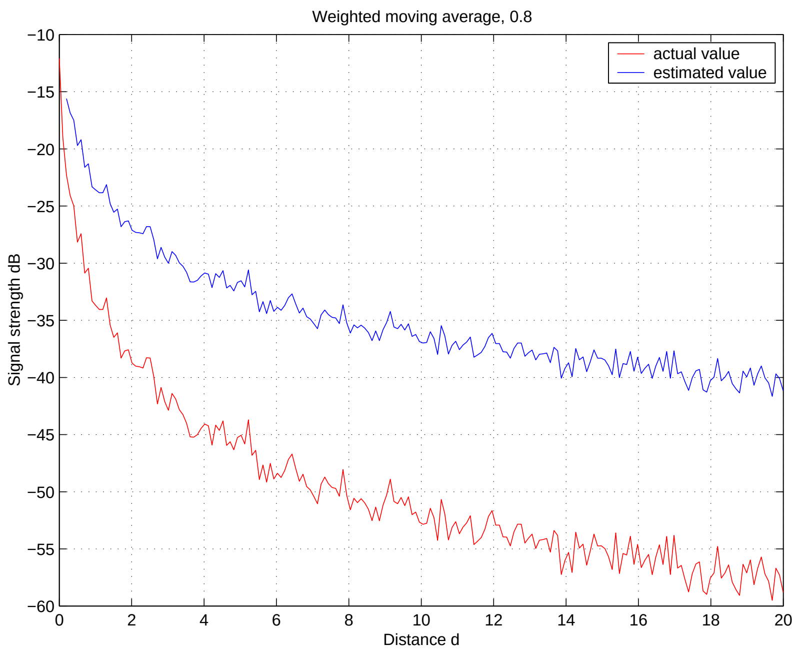

Figure 3.6: Weighted moving average filter with weight of 0.8 model.

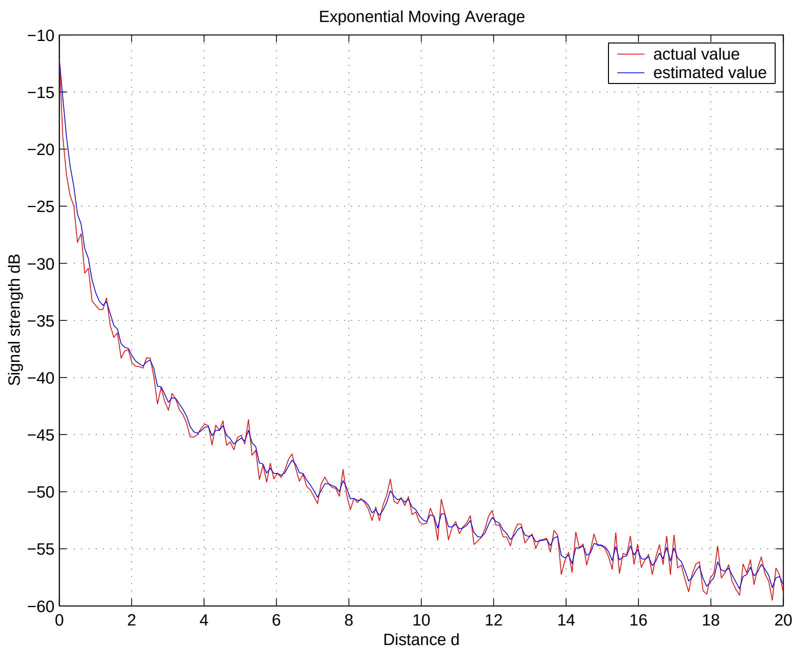

Figure 3.7: Exponential weighted moving average filter model.

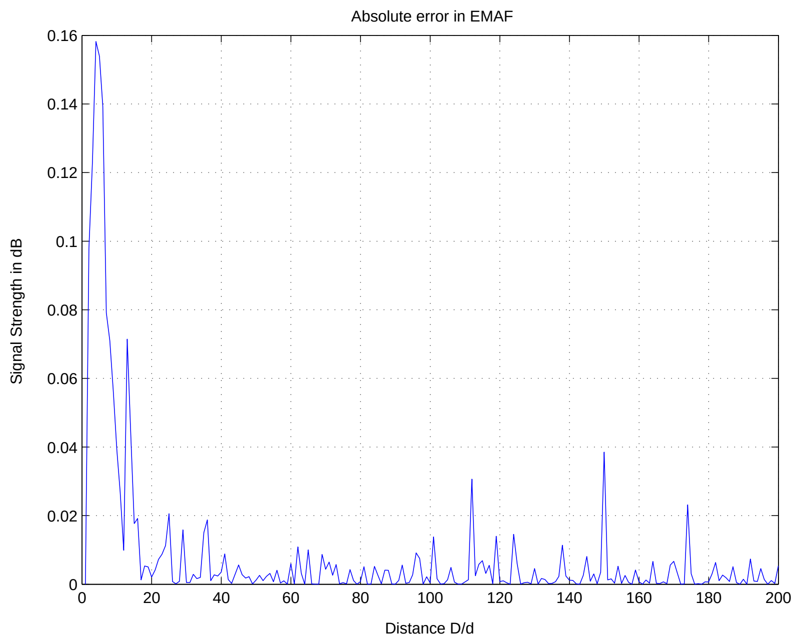

Figure 3.8: The absolute error from exponential weighted moving average filter model.

networks. The model also works well in a real-time environment as demonstrated in our simulation. From the model it is possible to detect movement and predict the exact location of the mobile node. The advantage of this model is that it needs only a small amount of data for prediction and requires a very short calculation time. This also means that the Grey prediction model provides a very fast response time for handoff decisions. As a result of the dynamic nature of the Grey prediction algorithm, it is perfectly suited as a mobility model.

Thus, we have introduced a methodology that has some unique characteristics which enable it to efficiently solve the mobility prediction problem, by predicting the received signal strength. To support our choice of the Grey theory approach, we developed a simulation model which includes slow fading and other practical network considerations. Specifically, the simulation used log-normal fading environment conditions. This provides us with the ability to relate the conditions of the obstructions that are present for a mobile node in the urban environment. Finally, we have made a series of comparisons with traditional moving average filters which gives an averaging of the values but not the exact prediction.

Although we have shown that the model has good prediction accuracy and has a very small prediction error, it still has some room for improvement. To improve this error, we propose a hybrid prediction model with additional algorithms which will be described in more detail in following chapters. This chapter has also included a literature survey of the some of the handoff methodologies that have been proposed.

Chapter 4

Hybrid Mobility Prediction Model

4.1 Introduction

The major goal of wireless communications is to allow a user to access the capabilities of global networks at anytime without encountering problems of location and mobility. Handoff performance is mainly dependent on the algorithm performing the handoff operation and this is where the challenge arises in performing accurate mobility prediction in order to achieve the QoS (Quality of Service) in wireless networks. For efficient handoff performance without any packet losses, an accurate model for mobility prediction is required. Accuracy in mobility prediction holds the key to such a capability while handing off a call. The main problems involved are handoff latency and unnecessary handoffs. As discussed in chapter 3, our approach to mobility prediction, which is the focus of this chapter, involves the use of a Grey model to make handoff decisions. As we have noted previously, the Grey model provides the necessary requirements that allow it to be used in realtime. In this chapter, we propose a hybrid model for the solution of mobility prediction which has significantly fewer errors and good prediction accuracy. There are a number of potential candidates for hybrid models but our choice here is to use a combination of fuzzy inference rules and Particle Swarm Optimisation (PSO). A framework for the hybrid model is provided in Fig. 4.1, where each block represents an algorithm that is used in the hybrid prediction model. Each algorithm has been selected to optimise prediction accuracy. The purpose of developing a hybrid approach is to improve on errors introduced by the Grey model in such a way as to meet QoS standards required for wireless networks. The model operates such that it can meet realtime requirements while ensuring that prediction accuracy is maximised. Based on the solution obtained, the hybrid model provides the prediction accuracy required to determine the exact position of the mobile node and also give the direction it is moving so as enable appropriate reservation of resources during handoff. Errors that occur in the basic prediction model will be fine-tuned by the fuzzy controller which is optimised by using an optimisation technique called "Particle Swarm Optimisation". The operation of the hybrid model involves taking the errors from the prediction model and feeding them into the fuzzy controller which uses fuzzy inference rules (based on appropriate parameters). The parameters from the fuzzy inference rules have to be adjusted such that the signal strength error is minimised. Prediction in the hybrid model is achieved by using a two step process which involves, 1) the construction of the fuzzy controller 2) the implementation of an optimisation technique used to determine parameters of the membership functions. Note that for determining the membership functions required, there are a number of existing methods, one of which is the "self-tuning algorithm" which is discussed in the chapter to follow. The errors in the Grey prediction model can be further corrected to provide accurate mobility prediction and improve handoff performance so as to reserve appropriate network resources. A prediction model for mobility assistance, based on Grey theory was discussed in sections 3.3 and 3.3.1. This model, although very efficient in prediction, still has some errors.

The model allows us to determine the exact position of the mobile node from the

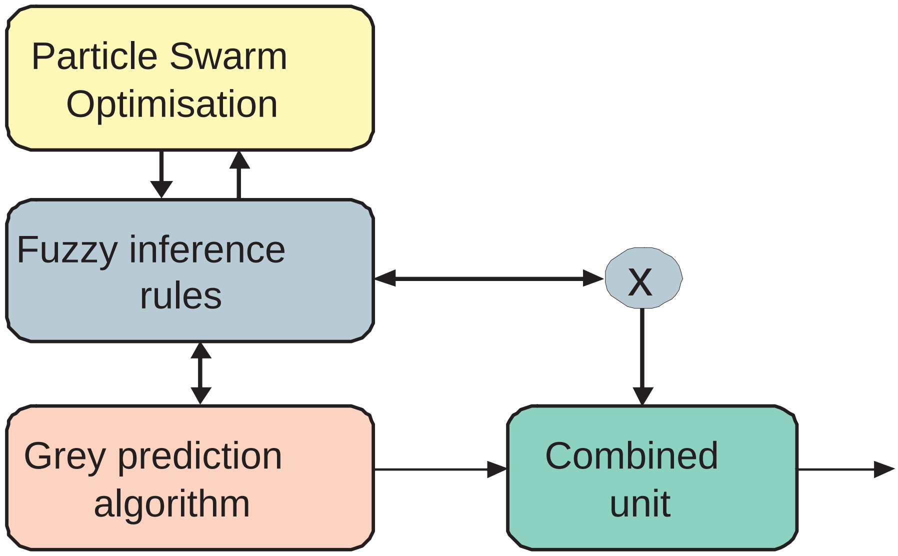

Figure 4.1: The hybrid model proposed model.

base station and also has a very quick response time. Fig. 4.1 shows the hybrid prediction model which sets out to correct the errors that occur in the basic prediction model previously described. The approach adopted by the hybrid model involves the following steps:

- Process the input data so as to construct the mobility model according to the Grey theory to predict the next signal strength values.
- Calculate the prediction errors and feed these errors as input into the fuzzy controller and fine tune the fuzzy parameters using appropriate optimisation techniques.
- The compensation value obtained is then added to the predicted value to get the optimal prediction.

The algorithm for the solution of mobility prediction using Grey prediction was discussed in chapter 3. When applied to the prediction of signal strength, it will enable us to determine the potential base station to which the call should be handed over. We model the hybrid model with the same simulation parameters as described for the simple Grey prediction model. In section 3.3.1, we modelled the situation for the simple Grey prediction model which can predict received signal strength. But in this chapter, we include two more operations to improve the error using the hybrid model. The procedure can be summarised into 3 major stages as depicted in the Fig. 4.1 using shaded boxes. The stages are:

- Perform Grey prediction
- Apply Fuzzy Inference rules
- Perform an appropriate optimisation technique (This will be the PSO method in this case, but genetic algorithms and self-tuning algorithm will be discussed in later chapters).

It is important to emphasise here that overall improvement obtained using the hybrid prediction model involves determining accurate predictions of signal strength. After compensating for errors this method will lead to excellent prediction accuracy. The major benefit of this outcome is that it is important to know the position of the mobile node so as to determine the potential base station and consequently improve handoff performance. The work in this chapter led to publications [4, 5]. The contributions of this chapter can be summarised as follows:

- Development of a hybrid model which includes the Grey model, fuzzy inference rules and particle swarm optimisation.
- Problems with errors arising from use of the Grey model are fine tuned with the hybrid model. This enables accurate mobility prediction leading to improved the handoff performance in wireless networks.
- A simulation was carried out based on a simple mobility scenario where the mobile node moves from one base station to another and results were plotted. A fitness function was also chosen to optimise the values from the fuzzy inference rules.

The rest of the chapter is organised as follows. In section 4.3, the fuzzy logic and the inference rules are reviewed in detail. A detailed description of the procedures that combine the algorithms in order to achieve improved handoff performance by using signal strengths is given in section 4.4. The simplified fuzzy reasoning with the details of the membership function is given in section 4.5. The algorithm used for optimisation (PSO) of fuzzy parameters is described in section 4.6. Finally, this chapter concludes with a simulation study that validates the hybrid model used for prediction of signal strength in the mobility environment.

4.2 Related work

In this chapter, we describe the hybrid model that uses the Grey model in combination with fuzzy logic [13, 95] and particle swarm optimisation algorithms. Our mobility prediction technique is designed to help improve handoff capability. The proposed technique minimises the number of handoffs and it is shown to have a very short calculation time and better prediction accuracy compared with hysteresis based decisions. Although our model has a very good tracking capability, prediction error is inevitable and we propose that this can be compensated for by the use of a fuzzy controller and then fine-tuned using PSO algorithms. In the past, many papers have discussed reducing errors by learning techniques, which is a very tedious process.

As previously noted, our hybrid prediction model makes use of search algorithms to fine-tune the fuzzy membership functions. To date, most of the available algorithms for tuning fuzzy parameters have been based on learning techniques and, specifically, genetic algorithms rather than any other evolutionary technique. Some of these methods include learning methods by gradient descent, neural networks, clustering methods etc. See [96, 97] for examples. Among these methods the gradient descent method has been widely exploited to adjust such systems.

Nomura et. al. [13] used a systemic approach or tuning methodology to self-adjust the function. The main advantage of this was faster speed in comparison with the neural network approach. The only disadvantage of this approach was that the tuning phase crucially depended on an ad hoc set of selected learning rates and the number of rules has to be fixed initially. But without knowing the optimum set of initial learning rate values in advance, there is unlikely to be an optimal result. However, since the method was very popular at that time and was the first of its kind to tune fuzzy parameters we decided to simulate and apply this to our hybrid prediction model.

Huang et.al [98] proposed the method of tuning the fuzzy parameters with an approach that combined the symmetric approach of Nomura et. al. It included a symmetric and an asymmetric adjusting method for the membership functions. The results that they found proved to be better than conventional neural network approaches and symmetric methods. The main advantage of using this approach was that it could solve to a local minimum as well as improve the convergence speed. Although better than the method proposed by Nomura, the computational methods were not very simple to apply.

Shinya et.al [99] proposed a learning approach based on a destructive method for the fuzzy rules as constructive methods and self-tuning did not have the capability to express the acquired knowledge efficiently. The proposed algorithm for self-tuning was also based on the gradient descent method – but with some changes to the construction of the fuzzy rules. In this method, they solved the rule base by deleting unnecessary rules based on the gradient descent method. The method proved to be superior in that the number of rules being constructed and convergence was achieved more rapidly. However, it was inferior in comparison with the learning speeds. The method of destructive learning was based on the inference error generated by the rule. If the rule generates more errors than a given threshold then it would be deleted. The details of the algorithm and its construction are presented in [99].

In this chapter, we consider only the method proposed by Nomura et.al [13] as our main focus was only to compare the convergence of the optimisation techniques with respect to our hybrid model based on PSO. Although many variations of the self-tuning algorithm exist there was no significant improvement in any of the methods. In the following sections, we discuss the details of the fuzzy rules, their construction and move on to the proposed hybrid model which is based on particle swarm optimisation.

4.3 Basics of Fuzzy Logic

This section begins with some of the fundamental concepts of fuzzy logic sets and their application with reference to our problem. In this section, we discuss how to design the fuzzy model to compensate for the predicted output errors from the Grey model. This section deals with the formation of membership functions which are crucially dependent on the acquired rules. Grey theory and fuzzy logic have been applied to many fields recently [100, 101]. Fuzzy logic is a fast growing technology which was first introduced by Professor Lotfi A. Zadeh in 1965 from the University of California at Berkeley. Fuzziness primarily describes uncertainty or partial truth. According to the theory, the traditional idealistic mathematical approach has been improved to accommodate partial truth by the introduction of fuzzy sets. Once the fuzziness is characterised to a reasonable level, fuzzy systems can perform well within the expected precision range. The concept of partial truth is characterised by fuzziness which yields a more accurate mathematical representation of perception of partial truth than so-called *crisp sets*. A crisp set may be defined as the collection of distinct precise defined values. In classical set theory, a crisp set can be a superset containing many other crisp sets. A superset can be represented as the "universe of discourse" if it defines the boundaries of all the elements that reside in it. We can determine whether the element belongs to the set and it can be represented by a characteristic function. If the set under investigation is A and testing of element x using the characteristic function χ is given by

$$\chi_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \in A \end{cases}$$

The numbers 0 and 1 constitute a valuation set. Using the notations above, the crisp set can be represented as

$$A = \{(x, \chi_A(x))\} \tag{4.1}$$

where $(x, \chi_A(x))$ is called a singleton. A fuzzy set is a set of collection of distinct elements with a varying degree of relevance. The characteristic function plays an important role, as it determines the degree of relevance. The characteristic function – also known as the *membership function* takes on values between 0 and 1. Using the same notation, a fuzzy set A can be represented as

$$A = \{(x, \mu_A(x))\}, \quad x \in X \tag{4.2}$$

where μ represents the membership function and $(x, \mu_A(x))$ is a singleton. The difference between crisp sets and fuzzy sets can be summarised below in Fig. 4.2. If set A is a crisp set, it is determined by the membership function that precisely identifies the boundaries of (a, b) on the universe of discourse. If the set A is a fuzzy set then the membership function determines the distribution of degrees of relevance across the universe of discourse. The following section discusses the construction of

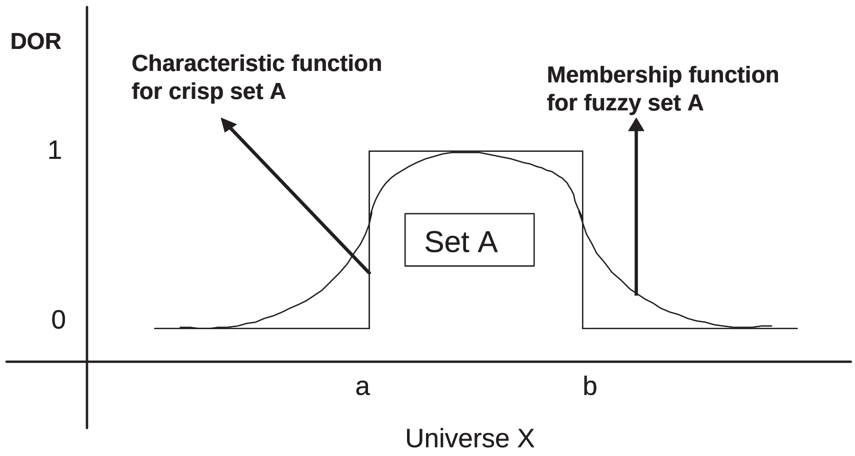

Figure 4.2: Degree of relevance for crisp sets and fuzzy sets using the membership function representation.

the basic IF-THEN rules and their use identified in our problem to compensate the error.

4.4 Fuzzy IF-THEN rules

According to Zadeh [102, 103, 104, 105], using fuzzy measures and possibility theory it is possible to apply classical methods of inference to a wider set of decision making problems. For example, "*the coffee is hot, drink slowly*" can be modelled in mathematical terms using membership functions concept characterising *"hot coffee"*. The capability of representing the *hot coffee* mathematically and similar such events generates ideas of linguistic variables. i.e., coffee is a linguistic variable like a variable in calculus which can take a value such as *very hot, hot, medium hot, cold* and *very cold* etc. Depending on the type of problem, we can indicate the category/region or get a suitable inference for the problem. This inference provides a medium for defining the linguistic variables or fuzzy variables which lead to the formation of fuzzy logic. We can also conclude that inference allows a heuristic interpretation to be part of the mathematical representation that leads us to the fuzzy variable concept. According to the fuzzy logic concept [106], a fuzzy variable has the following hierarchical information structure:

- Fuzzy variable
- Predicates linguistically identifying different regions of the universe of discourse.
- Membership function of each fuzzy set labelled by one predicate.
- Universe of discourse.

In fuzzy theory, the classical operations of the set theory such as the subset and union sets can also be applied. In comparison, fuzzy set theory offers a set of operations due to the nature of fuzzy sets such as partial truth and membership functions. The problem of fuzzy sets is to determine who are in the intersection of *"hot"* and *"cold"* which is a matter of degree expressed in the form of membership functions. Based on the degree of expressions of fuzzy sets and operators used, they derive what are known as the "Fuzzy IF-THEN inference rules". This theory also supports the use of conditional statements or rules. The most commonly used are the AND, OR and THEN operations. Specifically, the 'THEN' operation performs like a mapping function which is modelled by an implication process. By using such operations along with fuzzy sets it is possible to make an IF-THEN type inference rule as shown in the example below. For example, consider the typical case of body temperature, where it is possible to categorise different regions of temperature using predicates such as *normal, fever and high fever*.

Rule 1: IF temperature is 98.6 THEN person is normal

Rule 2: IF temperature is $98.6 - 101$ THEN person has fever

Rule 3: IF temperature is $101 - 104$ THEN person has high fever

In the above rules, the IF condition is termed as the *antecedent part* and the THEN part is considered to be the *consequent part*. In a conditional case, the first event constitutes a requirement for the second event to occur. The fuzzy implication process is suitable for a variety of linguistic process including conditional, deductive and inductive reasoning. A membership function that defines the implication relation can be expressed in a number of ways. A simple conditional proposition can be derived using the implication operators as follows:

```
IF X is A THEN Y is B
```

The implication relation can be defined by

$$R(x, y) = \bigcup_{x,y} \mu(x, y)/(x, y) \tag{4.3}$$

where the fuzzy variables x and y take the values of A and B respectively and $\mu(x, y)$ is the membership function of the implication relation. The membership function can be denoted as

$$\mu(x, y) = \phi[\mu_A(x), \mu_A(x), \mu_B(y)] \cdots x \in X, y \in Y \tag{4.4}$$

A fuzzy system is mainly based on fuzzy information processing and decision making. Thus designing a fuzzy algorithm usually refers to developing mechanisms which employ relational, compositional or implicational inference methods. The implicational reference in the form of conditional IF-THEN includes all levels of complexity. Based on these IF-THEN rules, a large domain of problems are solved which include control, classification, pattern recognition, modelling, prediction and forecasting [101]. The three main basic domains of information in a fuzzy algorithm are the input data, output data and design data. The input and output data are dynamic from the system whereas the design data is stored in memory. This can be represented as follows,

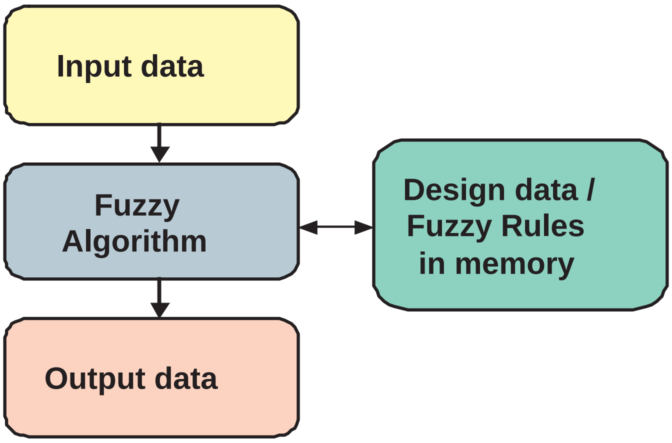

Figure 4.3: Figure showing the basic domains of the fuzzy algorithm.

In the Fig. 4.3, the design data consists of IF-THEN rules composed of fuzzy statements such as *the temperature is high*, *logic operators AND/OR* and the *implication operator THEN*. The *THEN* operator *(consequent part)* can be varied, depending on the specific solution they represent. The operations before the THEN operator which are IF statements can be the *left hand side* (LHS) of the operation of the fuzzy rule also called the '*antecedent part*'. In addition, there could be multiple statements on both sides of the rule depending on the type of rule. After the antecedent part and the consequent parts of the membership functions are determined, we can integrate them into the design of the fuzzy inference engine to process the output. The design of the membership function is a quality control step that will determine the output of the system. Therefore, identifying the antecedent part and the consequent part of the fuzzy rule plays an important role in design data. One of the features of these fuzzy rules is that it could also be used to interpret simple analytic expressions which are translated into fuzzy rules by mapping the input/output data. By appropriate membership function design, a fuzzy set can be used to model an algebraic operation such as $y = 2x + 4z$.

Understanding of membership functions depends on fuzzy system design. There can be triangular membership functions, bell shaped curve or trapezoidal membership functions. Usually between the triangular and bell shaped membership functions, the triangular membership function is the most preferred due to its ease of computation. Similar considerations also apply to the choice of implication operators, input data processing and logic design operations.

Amongst the fuzzy systems designed, adaptability is one of the criteria that is one of the most widely researched topics [102]. Adaptive fuzzy systems have the ability to change their internal design structure in accordance with environment changes that effect the performance of the original fuzzy system. One of them used in our proposed hybrid model is the adaptation of a membership function shape using a dedicated set of fuzzy rules. In adaptive fuzzy systems, they may consist of a set of fuzzy rules controlling the properties of another set. Adaptability has been used to design fuzzy system by learning from past experience. The topic has been under intense research and development in the form of learning algorithms and neural networks for some time. Some of the existing works discussed in [107] focus on using neural networks using training data to determine the membership function. Another method uses self-tuning methods which apply the notion of iterative learning. Some of the other techniques that could be used are evolutionary techniques which are discussed in the chapter 5. Finally, the most important point for any fuzzy inference engine is the design of the membership function and the rules of the inference engine.

In our problem, we have used a triangular membership function. The details of the input data, the inference engine and output data are discussed in the following section 4.5. At this point, it is important to state that a fuzzy inference engine is used to process data from the Grey Prediction model. The input data considered for the inference engine is the error obtained from the actual value and the predicted value from the prediction model. After the input data has been processed, we use an optimisation technique called PSO to determine the membership function.

4.5 Simplified Fuzzy Reasoning

In this section, we define the fuzzy rules and the membership function that has been used in our fuzzy inference engine. As stated in the sections above, the error from the Grey model is treated as the input to the fuzzy modelling which is compensated for by fuzzy inference rules and Particle Swarm Optimisation. It is important state that, the tuning process for the membership function is done by an optimisation technique. There are several methods for automatic adjustment of the membership function and one of such methods is the self-tuning algorithm based on the gradient descent method. The details of the optimisation techniques used in our hybrid model is discussed in the sections below. To determine the quantity of compensation for the predicted outputs, the inference rule that is used is as follows:

The input is expressed by $x_1, x_2, \cdots x_m$ and the output is expressed by y. The inference rule for the simplified fuzzy reasoning can be expressed as follows:

Rule i: IF x_1 is A_{i1} and $x_{(m)}$ is A_{im} THEN y is w_i (where $i = 1, 2, \cdots n$)

where, i is a rule number, $A_{i1}, \cdots A_{im}$ are the membership functions of the antecedent part, and w_i is the real number of the consequent part. The membership function, A_{i1} of the antecedent part is expressed by an isosceles triangle. The parameters that determine the triangle are the values of a_{ij} and b_{ij} shown in Fig. 4.4. The output of the fuzzy reasoning can be given as:

$$A_{ij}\left(x_j\right) = 1 - \frac{2 \cdot \mid x_j - a_{ij} \mid}{b_{ij}} \tag{4.5}$$

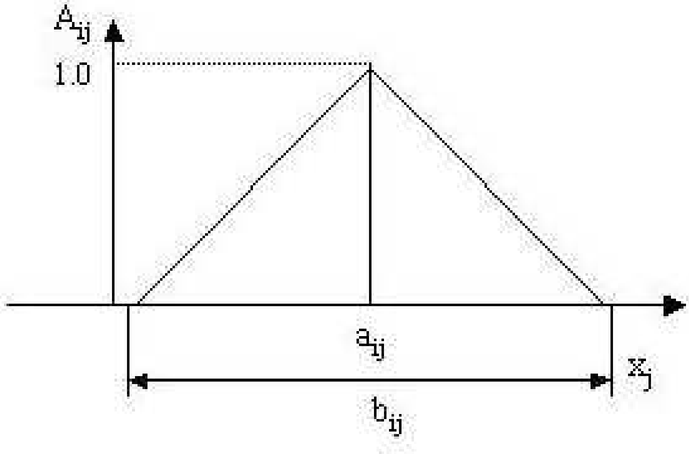

Figure 4.4: Membership function

where $(j = 1, 2, \cdots m)$ and i is a rule number.

$$\mu_i = A_{i1}(x_1).A_{i2}(x_2). \cdots A_{ij}(x_m). \tag{4.6}$$

$$y = \frac{\sum_{i=1}^{n} \mu_i.w_i}{\sum_{i=1}^{n} \mu_i} \tag{4.7}$$

μ_i is the membership function value of the antecedent part. The function which shows the shape of the membership function is adjusted by the central value a_{ij} , the width b_{ij} and the actual value w_i in the consequent part. The inference rules are tuned so as to minimise the objective function E that can be expressed by the following

$$E = \frac{1}{2}(y - y_r)^2 \tag{4.8}$$

where y_r is the desirable output data. The objective function E is interpreted as the inference error between the desirable output y_r and the output of the fuzzy reasoning scheme y [13].

4.6 Particle Swarm Optimisation

This section we describe the optimisation technique that has been used in our hybrid model. After the fuzzy rules have been defined for the inference engine, the next step is to determine the membership function. The parameters of the membership function viz: a_{ij}, b_{ij} and w_i are fine tuned by using the optimisation technique which in turn checks the fitness function which has been defined in the previous section 4.5. To begin with, this section introduces the topic of evolutionary techniques and this is followed by a detailed description of particle swarm optimisation.

4.6.1 Swarm Intelligence

Evolutionary techniques have their roots in artificial intelligence [108, 109]. As the research progressed, researchers started to explore ideas of intelligence arising from social contexts. The ideas were inferred by observing the interactions among social insects, birds and fish. For example, schools of fish have the advantage of escaping from their predators as each fish acts as a kind of watchful eye. Another example is that of birds which are on the lookout for food. If one of the birds sights some food, the others tend to follow. This idea of social behaviour is at the heart of the principles that define particle swarm optimisation. The general field of evolutionary optimisation is often considered to comprise of genetic algorithms, genetic programming, evolutionary techniques, programming and strategies. All of these evolutionary algorithms have a basic concept of population size. Particle swarm optimisation, like other techniques, utilises a population of candidate solutions to evolve an optimal or near optimal solution to a problem. The degree of any optimisation technique is measured by a fitness measure that must be defined.

Evolutionary algorithms such as genetic algorithms were first introduced in the 1970's by Holland and were used for numerical optimisation which utilise searching points in a solution space. PSO was first proposed by Eberhart and Kennedy in 1995 and is based on the swarming behaviour that guides individuals or particles searching the parameter space of possible solutions [110]. PSO was originally designed for continuous optimisation problems while GA can handle combinatorial optimisation problems. Recently, increased research on PSO has led to the handling of discrete, continuous and mixed-integer nonlinear optimisation problems. Thus, the main concept of swarm optimisation is based on research into how natural creatures behave as a swarm and model this idea with a computational algorithm.

PSO differs from evolutionary computational techniques by the use of population members which are called “particles" that are flown through a problem space. When the population is initialised, in addition to the variables that are given random values they are also stochastically assigned velocities. With each iteration, each particle's velocity is stochastically accelerated towards its previous best position (when it had its highest value) and towards a neighbourhood best position (the position of the highest fitness by any particle in its neighbourhood). The flow chart (Fig. 4.5) below represents a simple flow diagram of the PSO [111]. Particle swarm optimisation is powerful, easy to understand, implement and efficient in computation. The advantage of this technique is that it is faster than any other evolutionary optimisation technique and extremely resistant to being trapped in local optima. Particle swarm optimisation has been applied to diverse fields for optimisation in applications involving engineering and computer science.

Particle swarm optimisation (PSO) [112, 111, 113, 114] is a population based stochastic optimisation technique and like any other evolutionary technique it requires a population size to be defined. PSO was developed through simulation of bird flocking defined in a two dimensional space which is the problem space defined by the user. In PSO, each single solution is a “bird" in the search space and is called a “particle". The position of each particle is represented by its X-Y axis position in

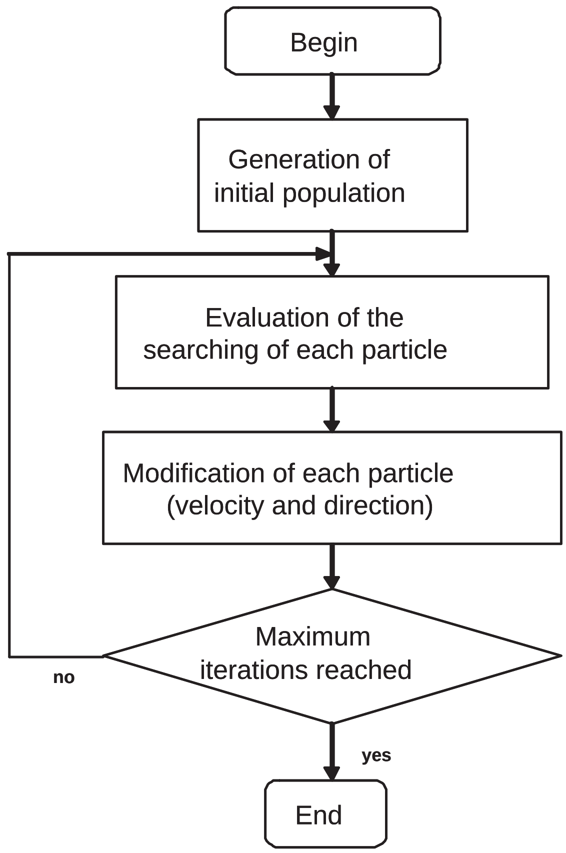

Figure 4.5: The flow chart of basic particle swarm optimisation technique

the problem space and also the velocity as represented by V_i^k (velocity of the particle i at iteration k). Modifications to the particle can be realised through the velocity and position information. The overall optimisation is governed by a prescribed *objective function* or *fitness function.*

In our problem definition, the fitness function used is defined in by equation 4.8. The membership function to be tuned depends on the parameters a_{ij}, b_{ij} and w_i. Each particle has three dimensions which carry the values of the membership parameters. Each of the particles knows its best value so far, called *pbest* and its X-Y position. This information is the best position of each particle. The particle also will know the best value so far in the group which is the *gbest*(global best) among the *pbest*. With each iteration, every particle tries to modify its position using the information of the current position, current velocity, the distance between the current position and *pbest*, the distance between the current position and *gbest*. As the particles have velocities and position it is updated with each iteration. The equations for the velocity and the positions are given by equations 4.9 and 4.10 respectively.

$$\begin{aligned} V_i^{k+1} &= wv_i^k + c_1 \mathrm{rand}_1 \times (pbest_i - s_i^k) \\ &\quad + c_2 \mathrm{rand}_2 \times (gbest - s_i^k) \end{aligned} \tag{4.9}$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \tag{4.10}$$

where,

v_i^k	velocity of the particle i at iteration k
v_i^{k+1}	velocity of the particle i at iteration $k+1$
w	inertia weight
c_j	acceleration coefficients
$\mathrm{rand}_{1,2}$	random numbers between 0 and 1
s_i^k	current position of i at iteration k
$pbest_i$	pbest of the particle i
$gbest$	*gbest* of the group
x^{k+1}	position of the particle at iteration $k+1$

s^k	current searching point
s^{k+1}	modified searching point
v^k	current velocity
v^{k+1}	modified velocity
V_{pbest}	velocity based on *pbest*
V_{gbest}	velocity based on *gbest*

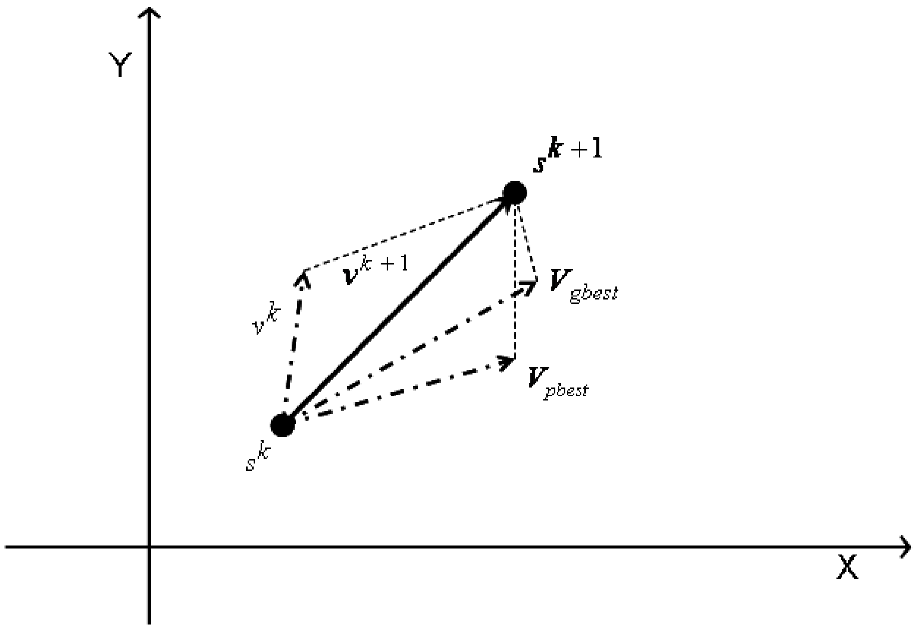

Figure 4.6: The concept of modification of a searching point by PSO

Modification of the searching point by PSO can be represented as shown in Fig. 4.6. The general flow of the PSO model can be explained as follows: The initial searching points s_i^k and velocities v_i^k of each of the particles is randomly generated within the problem space. The current searching point is set to *pbest* for each particle. The best value obtained so far with the *pbest* is set to *gbest* and the value is stored in memory. In the next step, it is checked to see if the objective function value for each particle is good. If the value is better than the current *pbest*, then the *pbest* value is replaced with the newer value. Then, if the current *pbest* is better than the *gbest*, it is also replaced and stored. In the next step, the velocity and the position of the particles are modified. These steps are repeated until a maximum number of iterations has been reached or if the objective fitness function is satisfied. In equation 4.9, the right hand side (RHS) of the equation has 3 terms. The first term is to provide diversification in the search procedure, the second and the third terms provide the intensification needed for the search. Basically, the PSO utilises searching points like GA and the searching points get close to the optimal values using the *pbest* and *gbest* values.

There are two acceleration coefficients mentioned in the equation. The acceleration coefficients closer to 1 will result in the lesser exploration and faster convergence. Setting the acceleration greater than 1 will result in the particle to possibly overstep *pbest* and *gbest*, resulting in a broader search. By setting to 2 or greater will cause the algorithm to go unstable because it may go out of the defined range. The random components in the equation are typically selected using the uniform distribution to provide high degree of randomness to the search. This can also be further given a higher density near to 1 for a faster and more direct search. The inertia weight w controls the influence of the velocity of the particle. It is typical to set the decay from $w < 1$ in order to allow the algorithm to converge on *gbest*. Any greater value of w makes the algorithm unstable because of the inappropriate emphasis on the previous velocity.

4.7 Simulation Parameters and Modelling Assumptions

In this section, we introduce the collection of assumptions that define our simulation model. These include the model chosen and the performance dependent parameters. For the proposed model, the parameters chosen are the best ones for optimal performance. As required for the hybrid model, we predict the signal strengths according to the simulation parameters shown in the Table 4.1 below. However, to improve the basic prediction model, the errors from the Grey model are fine-tuned using the fuzzy inference rules which are optimised using the PSO technique. The following section describes the settings for the Grey model and the PSO.

4.7.1 Experimental Setup for Proposed Hybrid Model

In this model, two base stations A and B were selected which were separated by D metres. The mobile device moves from one cell to another with a constant velocity and the received signal strength is sampled at a constant distance (in metres). Here, the mobile node moves from one base station to the other with constant speed. The signals from the base stations are affected by two major factors: path loss and log-normal fading. Rayleigh fading is not taken into account as it is assumed that any rapid fluctuations are averaged out. The received signal strengths from the current base stations to the target base station are sampled at distances Kd_s. The model considered also includes slow fading. The received signal strengths a_t and b_t (in dB) when the mobile is at a given distance kd_s are given by:

$$a_t = K_1 - K_2 \log kd_s + u_t \tag{4.11}$$

$$b_t = K_1 - K_2 \log (N - k) \, d_s + u_t \tag{4.12}$$

where $N = D/d_s$. The parameters $K_1 = 0$ and $K_2 = 30$ in dB which are typical of an urban environment accounting for path loss. K_1 is the signal strength at distance $d = 1$, and K_2 is the path loss component. Since threshold levels are not considered, the received signal strengths depend on the difference in received signal strengths. The simulation parameters used for the movement detection are as shown in Table 4.1.

Number of Base Stations	2
Trajectory	Straight Path
Sampling distance	10 m
Distance between base stations	2000 m
Path loss (K)	30 db
Transmitter power	0 dB
Fading Process	Lognormal fading
Standard Deviation (u_k)	8 dB

Table 4.1: The simulation parameters used for the prediction model

4.7.2 Experimental Setup for PSO

In the hybrid model for prediction of signal strengths, PSO was used as an optimisation technique to adjust the membership functions of the fuzzy inference rules. In our experiments, PSO was set to typical values which resulted in the optimum performance. According to the findings of many researchers, the scalability of standard PSO is not sensitive to population size [110, 115]. However, in our tests we examined the effects of population size which included values of 30, 60 and 100. Although the effects of inertia were not tried and this was set to a value near to 0.5 and decaying to zero, as the inertia weight is not problem dependent. Similar to the population size, the acceleration coefficients c_1 and c_2 were set to a value greater than 1 for optimal performance; since a smaller value lead to reduced accuracy since it was found that the particles were unable to converge quickly. For each test scenario, the simulation experiment was set up based on the errors from the prediction model, the standard fuzzy rules in the inference engine and different population sizes for the PSO. In the following section, we present our simulation results for the basic prediction model and the hybrid model and compare the performance in terms of improvements in the error. The results of convergence for the PSO with different population sizes is also presented.

4.8 Results

In this chapter, we have provided a methodology for the development of a hybrid prediction model. All the required building blocks of such methodology were assembled together into a unified algorithm for prediction of the received signal strengths. In the following, we first present results for the basic mobility prediction based on Grey theory as well as the hybrid model and then assess the accuracy of the model via simulation. The results of the Grey prediction in Fig. 4.7 show a plot of the actual values of the signal strength and the corresponding predicted values. The basic prediction model tracks the signal strengths but with some error. The Grey model does not predict large variations in the input data. These variations in prediction values are as shown in Fig. 4.8.

To improve this variation, we focussed our research on using PSO to identify the fuzzy rules in the proposed model. These variations in prediction, motivated our research to use fuzzy rules to compensate for the output from the Grey model. The fuzzy rules are built based on the error from the actual value and the predicted value that is assigned as the antecedent part and the quality of compensation constitutes the consequent part. The rules from the fuzzy inference engine have the membership functions which are identified by the PSO. By identification, we mean that we adjust the triangular membership function. Since the triangular membership function is being used here, we take into account the centre point a_{ij} and the width b_{ij} while fine-tuning it with the PSO. In the hybrid model, the inference rules are tuned so as to minimise the objective function E as given before in equation 4.8.

However, the proposed model improves the performance of the basic prediction model with very good accuracy. A close matching of the predicted values when the hybrid model is used is shown in Fig. 4.9 and 4.11. When the performance is compared with the basic prediction model which has an accuracy of $\pm 0.02dB$. In comparison when the hybrid model is used, it improves the performance to the order of 10^{-5} which is found to have an excellent prediction accuracy and is shown in Fig. 4.10 by plotting the absolute errors.

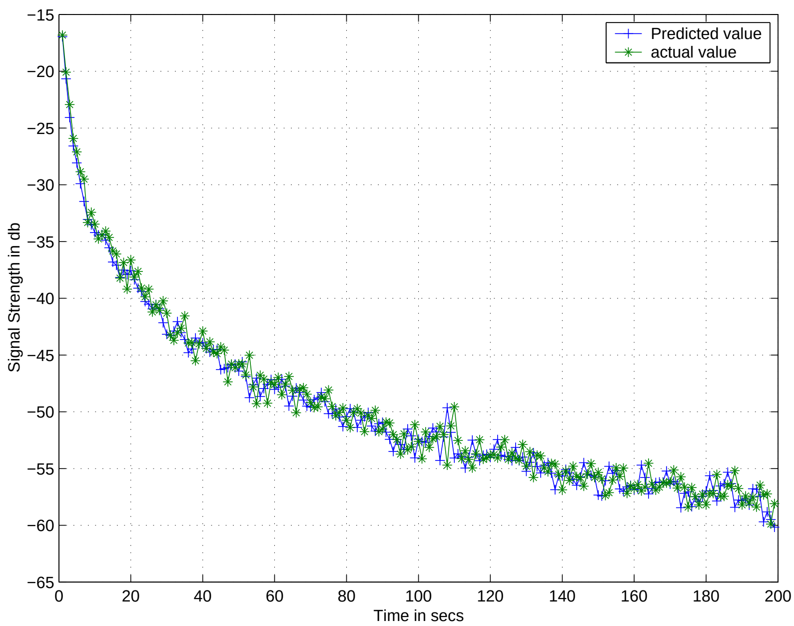

Figure 4.7: The output received signal strength from the basic Grey prediction model

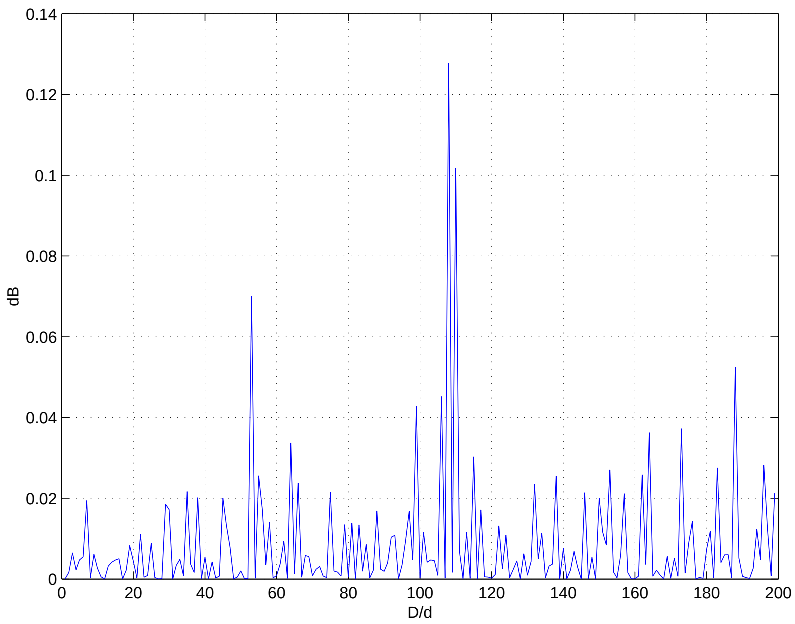

Figure 4.8: The predicted errors from the $GM(1,1)$ based model

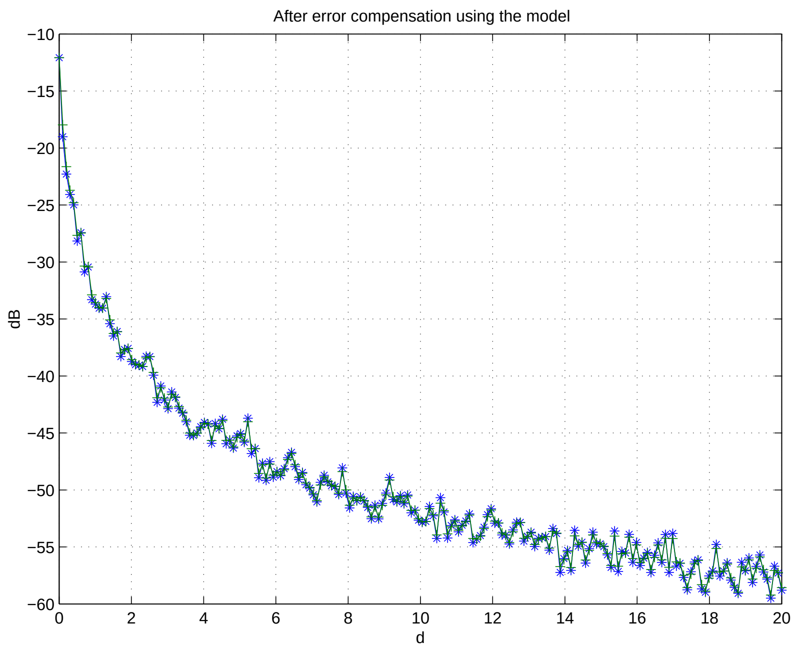

Figure 4.9: The output received signal strength from the proposed hybrid model

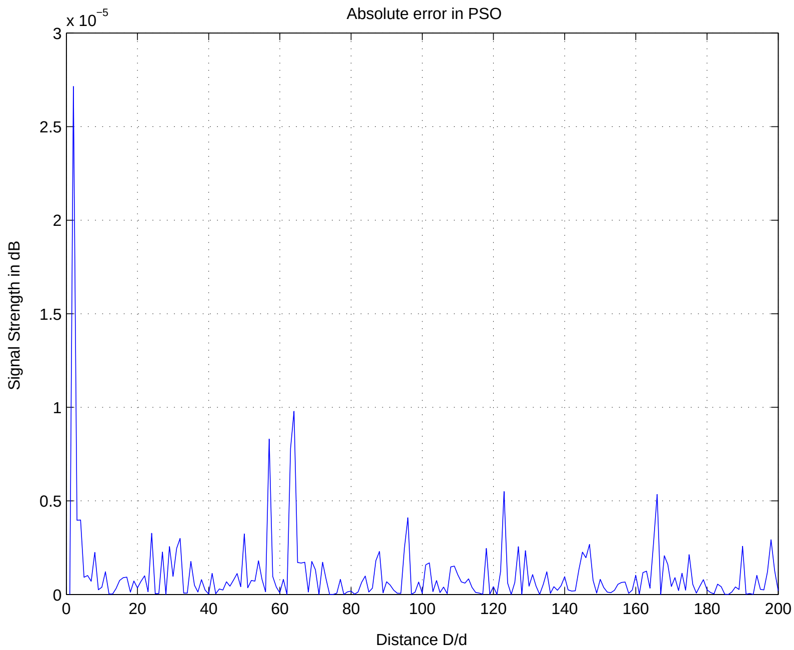

Figure 4.10: The predicted absolute errors from the proposed hybrid model

4.8.1 Discussion

Our goal in this section was to test the optimisation technique with different settings of the PSO as it is known to be governed by 3 principal parameters, viz: population size, acceleration coefficients and the inertia weight. As previously stated, we tried to determine typical values for the above-mentioned parameters. Although we did not test the hybrid model for sensitivity to the acceleration coefficients and inertia weight, we focused on the influence of population size. Although the effect of population size was not "sufficiently or significantly large" we found that different population sizes did have some effect, in terms of convergence. The Fig. 4.12 shows the convergence of the PSO for different population sizes.

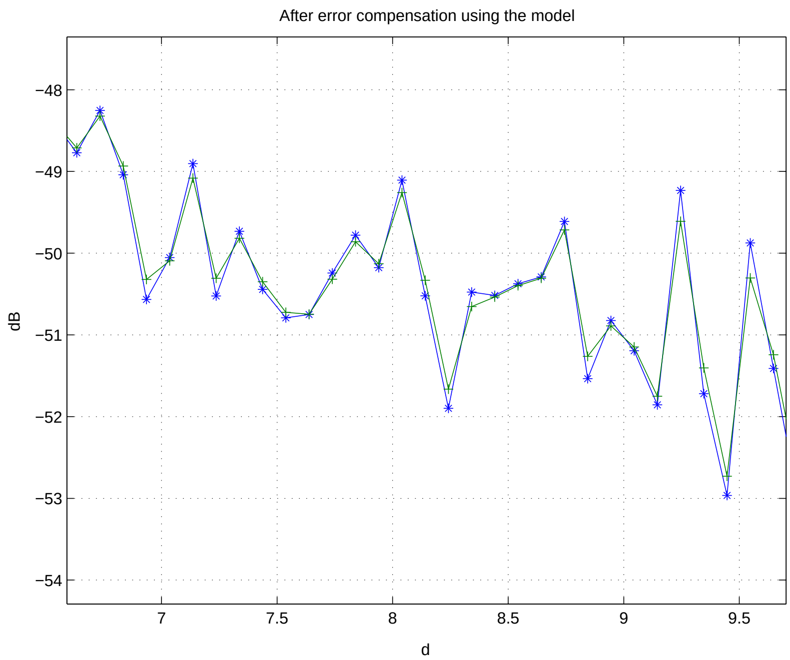

Figure 4.11: A closer look at the prediction accuracy of the proposed hybrid model using the PSO algorithm

Problems with PSO

The above results obtained by the PSO were based on the 3 parameters mentioned earlier. As in most learning algorithms – depending on the nature of the problem – the PSO can be modified to be more efficient for a specific problem. In order to obtain the optimal values for a specific problem like ours; especially in terms of efficiency and reliability, its weaknesses must be explored. There is at least one problem that we found to be addressed with our problem formulation: When a particle is found to be the new *gbest* all the particles tend to move towards it. If this new *gbest* is a particle lying outside with respect to the swarm, the particles may tend to move towards it and may leave some critical region around the *gbest* to be excluded from the search. A solution to this problem may be to increase the acceleration coefficients c_1 and c_2 so that they are greater than 1. But this may again

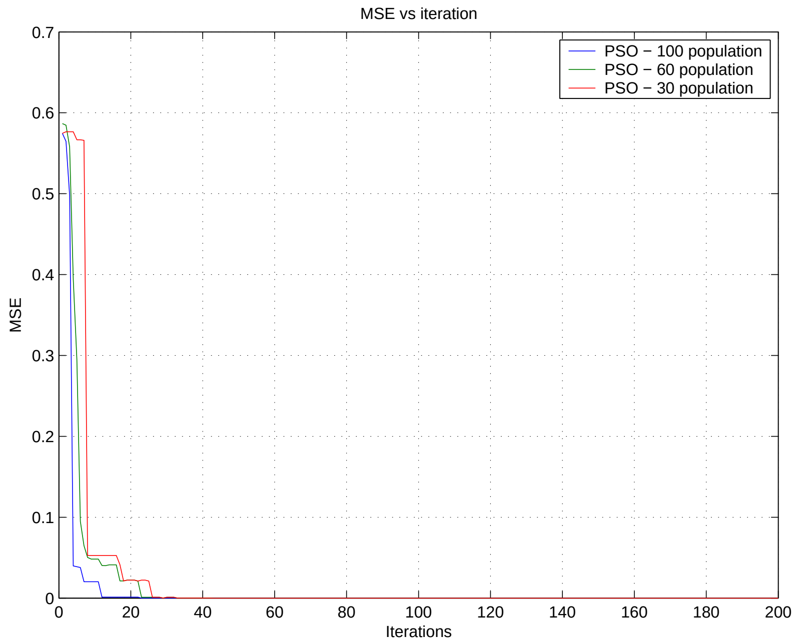

Figure 4.12: The convergence of the PSO algorithm with different population size

lead to convergence issues. To avoid this, some particles were randomly generated around the new *gbest* so that it still keeps the search somewhat global. By doing this, we are creating greater exploration regions for the particles which will result in better fitness. Another point to be mentioned is that the convergence of the PSO can be enhanced. According to Shi et. al. [114], a distribution such as Gaussian could be used which replaces the uniform distribution for selecting the random direction components. This will increase the randomness, and make it more directed towards *gbest*.

4.9 Summary

In this chapter, we have presented a technique for prediction of received signal strength values, which aids in providing efficient handoffs in wireless networks. We have evaluated the Grey model and further work has been done to perform error compensation. To improve prediction accuracy, we took the output and calculated the error between the predicted output and actual output. This is treated using fuzzy rules by actually compensating the error and fine-tuning it with a PSO algorithm. Grey prediction uses very little data (as little as four measurements) to predict the next signal strength. Even though the prediction accuracy of the Grey model is accurate it still deviates from accurately predicting values that have large variations. However, our simulation model shows considerable improvements by using fuzzy inference rules and the PSO algorithm. Our simulation results show that this hybrid model can improve prediction performance. By using the PSO approach, we find that it is far better than any other search technique that we have tried - particularly as it converges faster than other techniques.

The proposed hybrid model has a very good prediction accuracy which allows us to use it for the mobility environment. The choice of using the PSO algorithm as an optimisation technique to fine-tune the fuzzy parameters yields very good results. Since our model is highly dependent on the optimisation technique, we conducted several experiments to get the correct configuration of the parameters to be used in the PSO. The chapter also presents the results based on the different population sizes, acceleration coefficients and inertia weights for the PSO, that have been used in developing the model. We have also evaluated the accuracy of the model to predict the signal strengths very closely with the actual values. The simulation study validated the modelling approach and confirmed that, by using the hybrid model we can achieve good prediction accuracy. In future wireless networks, prediction techniques will be increasingly important because handoff will become more frequent in small cells and resources will be limited for various applications.

Finally, we recognise the dependence of the model on the optimisation technique and this motivated us to use other optimisation techniques such as genetic algorithms and self-tuning algorithms to compare their performance. Therefore, in the next chapter we shall modify the hybrid model by installing different combinations of the optimisation techniques to test the resulting performance via further simulation studies.

Chapter 5

Optimisation Algorithms for Hybrid Model

5.1 Introduction

The handoff performance in wireless networks poses a challenge to transfer the call from one base station/access point to another. Mobility prediction and its accuracy play an important role in achieving this performance. In the previous chapter, we discussed the hybrid prediction model which consisted of three building blocks, namely: the Grey prediction model, fuzzy inference engine and an optimisation technique (namely PSO). However, as discussed in chapter 4, we would like to extend the work on the selection of an appropriate optimisation technique wherein PSO was chosen for fine tuning of the fuzzy parameters. We need to determine whether the PSO method is the best approach to this optimisation problem. In this chapter, we propose the same hybrid model but with different optimisation techniques for compensating the error from the basic Grey prediction model. We have selected two other popular algorithms for this study and they are the Genetic Algorithm (GA) and the Self-Tuning Algorithm (STA) – which works on the gradient descent method. The hybrid model is modelled with these techniques and compared with the performance involving PSO. We describe the optimisation criteria involved and compare the algorithms in terms of accuracy and convergence times. The accuracy improvements obtained by the various approaches is shown by comparing results from simulation experiments.

As discussed in chapter 4, mobility prediction is done by determining the signal strengths from the base station using our hybrid model. Based on the hybrid approach model we construct new versions by substituting the PSO model with two existing optimisation techniques, viz: a Genetic Algorithm (GA) and the self-tuning algorithm (STA). Fig. 5.1 shows the proposed hybrid model considered for predicting the signal strengths from the base station. Each optimisation technique is substituted one at a time and the performance is tested. The steps defining the variants of the model can be summarised as follows:

- Received signals strengths from the base station
- Fuzzy inference engine with the defined IF-THEN rules
- Optimisation technique to identify the fuzzy rules

As hinted previously, it is very important that the chosen optimisation technique exhibits the necessary prediction accuracy when applied to the hybrid model. Over the years, many researchers have developed optimisation techniques based on evolutionary techniques (e.g., GA, PSO etc.) and also based on gradient descent methods (e.g., self-tuning algorithm) which are specified to a particular type of problem. Therefore, it is important to state that each of the optimisation techniques suits a particular problem and not every optimisation technique will give the same results.

This chapter will show the performance of these optimisation techniques when substituted as part of our hybrid model. A selection of the various optimisation algorithms has been combined together to produce three different hybrid models. In the framework that has been constructed (Fig. 5.1), each block represents an

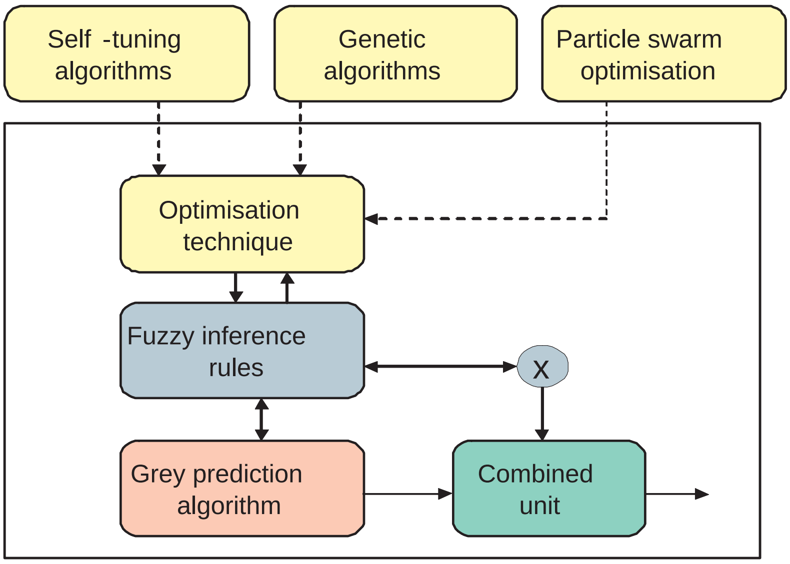

Figure 5.1: The hybrid model proposed model with different optimisation techniques.

algorithm used in the hybrid model. It will be seen that each optimisation technique has been applied one at a time. As per the hybrid model mentioned in chapter 4, the errors from the Grey prediction model are fed into the fuzzy inference engine whose membership functions are determined by an optimisation technique. The information required by all the models can be summarised as follows:

- Process the input data so as to construct predictions according to the Grey theory in order to predict the next data values.
- Calculate the prediction errors and feed these errors as input into the fuzzy controller and fine tune the fuzzy parameters using the selected optimisation technique.
- The compensation value obtained from the optimisation is then added to the predicted value to get our best prediction value.

The work in this chapter led to publications [6, 10]. The contributions of this chapter can be summarised as follows:

1. Development of variations to the hybrid model to select the best evolutionary

optimisation technique for fine tuning the fuzzy parameters.

2. Two algorithms, the particle swarm optimisation and the genetic algorithm were compared in terms of their convergence speed to the optimal value.
3. The hybrid models, viz: the genetic algorithm based mobility prediction and the particle swarm optimisation were tested and simulated.
4. A self-tuning algorithm, proposed by [13], was also compared.

The rest of the chapter is organised as follows. In section 5.2 and 5.2.1, we discuss the motivation and related work that led to further investigations on suitable optimisation techniques for our hybrid model. We describe the fuzzy rules in the inference engine which have been used in the hybrid model with two optimisation techniques which are the self-tuning algorithm and the genetic algorithm. In the next section 5.3, we describe the hybrid model with the self-tuning algorithm based on gradient descent method. The genetic algorithm approach is discussed in section 5.3.2. Some results and conclusions are discussed in sections 5.5 and 5.6 respectively.

5.2 Related Work

Over the last two decades, arguably a major advance in telecommunication networks has been the deployment of wireless access technologies. In order to achieve seamless mobility, the problem that needs to be addressed is changing the network point of attachment transparently as the user moves around. When a Mobile Node (MN) moves away from its current point of attachment, handoff is invoked to choose another point of attachment. Implementation of a mobility prediction technique is a promising approach that helps to improve this handoff capability [56, 61, 116]. In [56], the authors discuss mobility prediction based on moving patterns of mobile nodes. Here, their aim is to reduce the number of control packets needed to reconstruct the routes and thus minimise overhead. Their paper also uses GPS tracking systems to assist in their prediction method. There are also some papers that use a sector concept where the cell of a particular base station is divided into defined regions or zones. Depending on the position of the mobile node, it predicts the next likely cell that would be visited by the user. All the methods that were proposed for mobility prediction has been discussed in detail in literature review in chapter 2.

In this chapter, the technique proposed is a combination of Grey prediction, fuzzy logic [13] and evolutionary algorithms such as Genetic Algorithms (GAs) or Particle Swarm Optimisation (PSO) [112]. The parameters considered in our problem, utilise the Received Signal Strength Indicator (RSSI) values from the base station. In [13, 117] some roughly determined membership functions from fuzzy rules have been fine-tuned by using a gradient descent method. The gradient descent method has been widely used for tuning in many similar systems. However, the self-tuning algorithm depends heavily on the choice of initial settings and is often very tedious or complicated. Arslan et. al [118, 119] were the first to propose the tuning of fuzzy membership functions using genetic algorithms. According to them, the approach adopted for acquiring the shape of the particular membership function depends on the type of application. Primarily, they proposed that the problems associated with fuzzy logic often have membership functions which are linear and usually triangular in shape. Although there have been many proposals involving the gradient descent method that report good results, their method used the popular genetic algorithm for determining the shape of the membership function. Similar to Arslan et. al., a GA was proposed by Karr et. al. for the same purpose. Karr [120] applied a GA to design the Fuzzy Logic Controller (FLC) for the cart pole problem. He presented two examples: a non-adaptive GA designed for the FLC and a GA designed adaptive FLC where membership functions were adapted in real-time. The membership functions were Gaussian in nature and the objective was to minimise a given fitness function. Meredith et. al. [121] also applied a similar technique for memberships functions in the FLC for a helicopter. Lee and Takagi [122] applied a similar technique, but the objective was to minimise the number of rules in the fuzzy system rather than adjusting the membership function.

In our hybrid model, we applied the above technique, based on a GA, to our problem of determining the membership functions in the proposed hybrid model. The reason behind trying out both the self-tuning algorithm (STA) and GA on our model was that they were very popular and used in many of FLCs [123, 124]. We also extended our work by comparing these various techniques and applying them to our particular research problem.

As computational power grew, simple search techniques became more elaborate and population based searches gained more importance. In [108, 109, 110], comparisons were made on evolutionary and GA techniques showing their relative advantages. These comparisons were made using standard benchmark problems such as the generalised sphere function, Rosenbrock function, Griewank function and the Rastringin function. Evolutionary techniques such as GA, Hill Climbing and PSO were compared on these functions - including tests on speed of convergence. With respect to all of these, GA and PSO faired better than the Hill Climbing method. This was a motivating factor for us to investigate alternative approaches to PSO for our problem. Thus, in this chapter, two optimisation techniques for fine-tuning the fuzzy parameters are proposed and a comparison of these 2 methods is carried out together with the gradient descent method as proposed in [13]. The following sections discuss the hybrid model with different optimisation techniques, their construction and parameter selection.

5.2.1 Simplified Fuzzy Reasoning

In this section, we shall review how to design the fuzzy model to compensate the predicted output from the Grey system. Among the methods proposed for automatic adjustment is the self-tuning algorithm which has been proposed in [99, 125] and will be explained in section 5.3.1. This self-tuning algorithm is applied which is aimed at drastically reducing the processing by the use of learning functions. Similar to PSO used in the previous chapter, a membership function can be developed for both the self-tuning algorithm and the popular genetic algorithm. To determine the quality of compensation for the predicted outputs, the inference rule that is used is as follows: The input is expressed by $x_1, x_2, \cdots x_m$ and the output is expressed

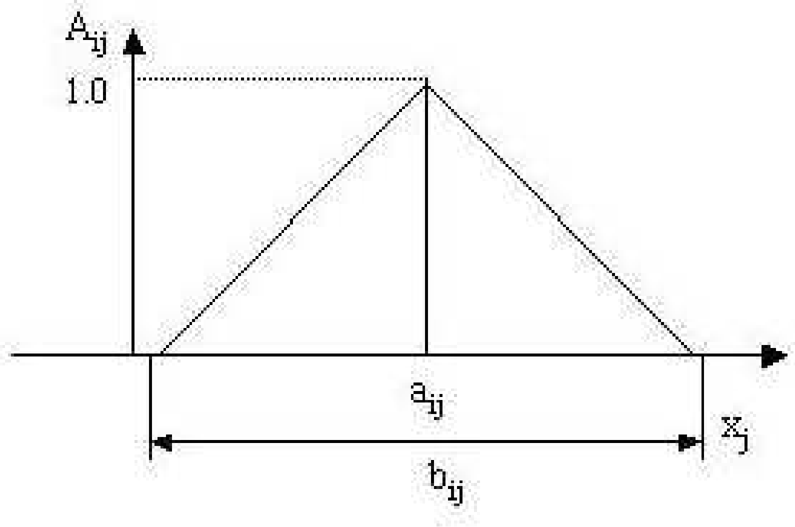

Figure 5.2: Membership function.

by y, the inference rule of simplified fuzzy reasoning that can be expressed by the following :

Rule i: IF x_1 is A_{i1} and $x_{(m)}$ is A_{im} THEN y is w_i, (where $i = 1, 2, \cdots n$)

where, i is a rule number, $A_{i1}, \cdots A_{im}$ are the membership functions of the antecedent part, and w_i is the real number of the consequent part. The membership function, A_{i1} of the antecedent part is expressed by an isosceles triangle (Fig. 5.2). The parameters that determine the triangle are the values of a_{ij} and b_{ij}. The output of the fuzzy reasoning can be given as

$$A_{ij}\left(x_j\right) = 1 - \frac{2\cdot \mid x_j - a_{ij} \mid}{b_{ij}} \tag{5.1}$$

where $(j = 1, 2, \cdots m)$ and i is a rule number.

$$\mu_i = A_{i1}(x_1).A_{i2}(x_2). \cdots A_{ij}(x_m). \tag{5.2}$$

$$y = \frac{\sum_{i=1}^{n} \mu_i.w_i}{\sum_{i=1}^{n} \mu_i} \tag{5.3}$$

μ_i is the membership function value of the antecedent part. The function which shows the shape of the membership function is adjusted by the central value a_{ij} , the width b_{ij} and the actual value w_i in the consequent part. The inference rules are tuned so as to minimise the objective function E that can be expressed by the following

$$E = \frac{1}{2}(y - y_r)^2 \tag{5.4}$$

where y_r is the desirable output data. The objective function E is interpreted as the inference error between the desirable output y_r and the output of the fuzzy reasoning scheme y [13].

5.3 Optimisation Algorithms for Hybrid Model

In this section, we present the two models namely the hybrid mobility prediction model optimised by self-tuning algorithm and genetic algorithm. The main purpose of this investigation was to find an appropriate optimisation technique suitable for our hybrid model. The above-mentioned fuzzy parameters are used in both of these models. In this section, we investigate the performance of the three optimisation techniques with respect to our hybrid model. This investigation will include a comparison based on the convergence of the fitness value. We also discuss the performance of the GA and PSO, based on the simulation results obtained from our prediction of signal strengths. The results also include comparison with different population sizes used in both GA and PSO based hybrid models.

5.3.1 Self-tuning algorithm

We shall now consider the self-tuning algorithm based on the gradient descent method. Nomura et. al. [13] were the first to propose the gradient descent method to tune the fuzzy parameters. According to their proposed method it uses the "Takagi Sugeno" model [126] with constant outputs and triangular membership functions. For a comprehensive study of the model, see [126]. Tuning, the modification of one or more of the design parameters [127], is usually initiated by modifying the membership functions. Tuning methods can be divided into on-line and off-line methods, depending on whether the design parameters are tuned while the model is running or afterwards. These methods include variations of least-square methods, gradient descent algorithms, fuzzy clustering and neural net approaches. Among some of the methods proposed [107, 127, 128], for learning of fuzzy rules the gradient descent method was a promising approach. It is simple and can accurately learn fuzzy parameters as it allows them to build realtime learning algorithms and converge to a minimum quicker than those proposed at that time.

The algorithm proposed by them was a self-tuning or a self learning algorithm which was based on learning constants. The idea was to tune the membership functions in the antecedent part and real number in the consequent part of the inference rules by means of gradient descent method. The motivation was to improve the learning speed and generalisation capability over that proposed by the neural networks method [97]. This method of providing fuzzy reasoning using a learning function was done using neural networks. But these methods had problems achieving the desired result over short periods of time. In the self-tuning method, the triangular membership function of the antecedent part and the real number of the consequent part are assumed to be given. As previously done by PSO in the hybrid model proposed in chapter 4, we use self-tuning algorithm for the same. In the self-tuning algorithm the centre value, width of the triangular membership function and a real number of the consequent part are tuned by means of descent method.

According to the algorithm [13], the descent method is to seek the vector Z which minimises the objective function $E(Z)$ where Z is a p-directional vector $(Z_1, Z_2 \cdots Z_p)$ of the tuning parameters. In this method, the vector which decreases the value of an objective function $E(Z)$ is expressed by $(-\delta E/\delta Z_1, -\delta E/\delta Z_2 \cdots, -\delta E/\delta Z_p)$ and the learning rule is expressed by

$$Z_{t+1} = Z_i(t) - K.\frac{\delta E(Z)}{\delta Z_i}, \qquad (i = 1, 2 \cdots p) \tag{5.5}$$

where, t is the number of iterations of learning and K is a learning rate. Altering Z according to this learning rule, the objective function E converges to a minimum. Since the shape of the membership function A_{ij} is defined by the centre value a_{ij} and width b_{ij}, the objective function consists of tuning parameters which are a_{ij}, b_{ij} and w_i. In this method, the inference rules are tuned so as to minimise the objective function as defined in section 5.2.1. From equation 5.5, the learning rules of the simplified fuzzy reasoning are expressed by following equations 5.6, 5.7 and 5.8:

$$a_{ij}(t+1) = a_{ij}(t) - K_a.\frac{\delta E}{\delta a_{ij}} \tag{5.6}$$

$$b_{ij}(t+1) = b_{ij}(t) - K_b.\frac{\delta E}{\delta b_{ij}} \tag{5.7}$$

$$w_i(t+1) = w_i(t) - K_w.\frac{\delta E}{\delta w_i} \tag{5.8}$$

The above equations show the respective $(t+1)$ values of the tuning process. K_a, K_b and K_w are constants. The gradients of the objective function viz:, $-\delta E/\delta a_{ij}$, $-\delta E/\delta b_{ij}$ and $-\delta E/\delta w_i$ in equations 5.6, 5.7, 5.8 can be derived from the equations 5.1, 5.2, 5.3 and equation 5.4 which is as follows:

$$\frac{\delta E}{\delta a_{ij}} = \frac{\mu_i}{\sum_{i=1}^{n}\mu_i}.(y - y_r).(w_i - y).sgn(x_j - a_{ij})\frac{2}{b_{ij}.A_{ij}(x_j)} \tag{5.9}$$

$$\frac{\delta E}{\delta b_{ij}} = \frac{\mu_i}{\sum_{i=1}^{n}\mu_i}.(y - y_r).(w_i - y).\frac{1 - A_{ij}(x_j)}{A_{ij}(x_j)}.\frac{1}{b_{ij}} \tag{5.10}$$

$$\frac{\delta E}{\delta w_i} = \frac{\mu_i}{\sum_{i=1}^{n}\mu_i}.(y - y_r) \tag{5.11}$$

where sgn is the positive or negative sign from z. Substituting equations 5.9, 5.10, 5.11, in equations 5.6, 5.7, 5.8, the learning rules of the simplified fuzzy reasoning can be expressed by the following equations:

$$a_{ij}(t+1) = a_{ij}(t) - \frac{K_a.\mu_i}{\sum_{i=1}^{n}\mu_i}.(y - y_r).(w_i - y).sgn(x_j - a_{ij}).\quad \frac{2}{b_{ij}.A_{ij}(x_j)} \tag{5.12}$$

$$b_{ij}(t+1) = b_{ij}(t) - \frac{K_b.\mu_i}{\sum_{i=1}^{n}\mu_i}.(y - y_r).(w_i - y).\frac{1 - A_{ij}(x_j)}{A_{ij}(x_j)}.\frac{1}{b_{ij}} \tag{5.13}$$

$$w_i(t+1) = w_i(t) - \frac{K_w.\mu_i}{\sum_{i=1}^{n}\mu_i}.(y - y_r) \tag{5.14}$$

The learning rules of equations 5.12, 5.13, 5.14 adaptively change the tuning parameters to minimise the objective function E. Thus, using the learning rules the tuning parameters of the inference rules are optimised to minimise error between the desirable output and the output of the fuzzy reasoning.

As previously noted, we have assumed a triangular membership function. The details of the input data, the inference engine and output data have been discussed in section 5.2.1. In summary, to increase the performance of a fuzzy model, the tuning algorithm requires careful analysis of the parameters. In tuning of fuzzy membership functions, the shape, number and width of the membership functions, the length of the data series for membership function generation and the tuning period are all considered here as tuning parameters. The input data that is considered for the inference engine is the error obtained from the actual value and the predicted value from the basic Grey prediction model. After the input data has been processed, we use an optimisation techniques such as PSO, GA and self-tuning algorithm separately (one at a time) to determine the membership function. In the following section 5.3.2, we discuss the genetic algorithm approach to our problem.

5.3.2 Genetic Algorithms

We shall now consider the details of the GA approach to solving our optimisation problem. As before, the fuzzy rules are determined by the optimisation scheme - in this case, the GA algorithm. The parameters of the membership function to be tuned by the GA algorithm are a_{ij}, b_{ij} and w_i. To begin, we shall briefly describe the popular genetic algorithm methodology and then discuss its application to our hybrid prediction model.

5.3.3 Simple Genetic Algorithm

A Genetic Algorithm is a general search technique [24, 124, 129] that was introduced by Holland in 1970's, not to solve a particular problem, but to investigate the effects of natural adaptation in stochastic search algorithms. A GA model consists of possible solutions which can be refined through selections of parameters, crossovers and mutations. An objective function (also called the "fitness function") is chosen in such a way that good points in the search space possess a high fitness value.

Most of the optimisation paradigms move from one point in the decision space to another and use a deterministic value. The drawback of this approach is that it is easy to get stuck on a local optimum value. But, with evolutionary techniques, it always starts with a population and the sub-populations or newer populations also have the same number of members in each generation. As a consequence of this, all the populations are explored lowering the probability of becoming stuck on a local optimum. With the help of operations such as crossovers and mutations, population size can be controlled and the search can be more effective. Although there exist many variations of the basic GA approach, the fundamental idea revolves around a mechanism that operates on a population of chromosomes or individuals. GA has been proven to provide a robust search in a complex search space, giving a valid results that can only be confirmed by requiring an exhaustive search. The process of optimisation or the general cycle of the GA can be summarised as follows:

1. Initialisation and generation of a population of chromosomes which are selected at random.
2. Evaluate the individual fitness of the existing population and select the best individuals to reproduce.
3. Perform the crossover operation to produce new offsprings or chromosomes.
4. Perform the mutation operation on the offsprings.

5. Decode the offsprings and evaluate the objective function fitness values. If the size of the current population has not exceeded the maximum population size, return to step 2.
6. If the termination condition is not satisfied return to step 1 otherwise; stop the operation.

A typical GA process consists the three basic operations. The first step needed is to start the GA process with a initial population. This step is usually done randomly. Note that, in the actual application this population could range from 20 to 100 individuals or more. Initialisation of this population is done using a random number generator depending on the type of problem. The parameter values, however, are chosen in such a way that they are close to possible solutions. After the initialisation step, the individuals are evaluated and then the next set of individuals are needed which are known as the sub-populations. In many cases, the solutions may exist in the initial set of individuals. However, the process is repeated with mutation and crossover until the fitness value is reached. In the following subsections, we describe the key GA operations known as selection, crossover and mutation that occur in a typical GA process.

Selection

The selection operator plays a key role for GA individuals as it drives them towards optimality. It also determines how individuals compete in gene survival. Each individual represents a possible solution to the given problem. The selection process removes the bad solutions and keeps the good ones. In this process, the individual with the best fitness value is selected to be part of the next generation. The selection criteria is usually done on the whole population and is repeated for individuals which results in the loss of diversity. In a GA, population is altered by crossover and mutation (see below).

Crossover

Crossover is done to investigate the performance of the new individuals that resemble existing ones. This is done on individuals and leads to the construction of new intermediate solutions. The notion of generations arises as parents crossover to create new offsprings. The crossover operator used in our GA is known as a one-point crossover. Crossover does not always take place between two selected genomes but with a given crossover probability. A population losing diversity often converges faster before the global optimum and is described as premature convergence.

Mutation

After the crossover operation, a genome is subject to mutation. In GA's, the mutation operator is the source of random variations. Mutation is done to alter the population slightly. The operator iterates through each gene in the genome altering it slightly. Altering the genes in this way can be vital to provide the diversity which is needed. The probability of mutation is usually a variable GA parameter.

These processes continue for a prescribed number of iterations or generations. The performance of the GA depends significantly on the population size. Increasing the population will increase the computation time. There should be a balance in choosing the population size and the number of chromosomes. In our problem, a GA has been used for optimising (minimising) the error by fine tuning the parameters based on fuzzy reasoning. A simple GA algorithm with a single point crossover was used and selection was based on a roulette wheel process. The GA was primarily used to compute the membership functions from fuzzy reasoning and to compute the fitness functions as suggested in equation 5.4. For our experiment, we used 100 chromosomes in the population. The maximum number of generations allowed was 1000. The criterion was to find the best solution so that the fitness value was kept to a minimum. In the following section, we once again describe the other optimisation technique described earlier in chapter 4 which we used for comparison. After the basic definitions, we proceed with simulations in order to compare them with the three variations of the hybrid model based on our three optimisation techniques.

5.3.4 Particle Swarm Optimisation

Details concerning particle swarm optimisation have already been presented in chapter 4. In every iteration, each particle is updated using the following two "best" values. The first one is the best solution (fitness) that has achieved so far. The best value is stored. This value is called the *pbest*. Another ŚŚbest" value that is tracked by the optimiser is the global best and this is called *gbest*. The particle will have velocities, which direct the flying of the particle. In each generation, the velocity and the position of the particle are updated. The equations for the velocity and the positions are given by equations (5.15) and (5.16) respectively.

$$\begin{aligned} V_i^{k+1} &= wv_i^k + c_1 \mathrm{rand}_1 \times (pbest_i - s_i^k) \\ &\quad + c_2 \mathrm{rand}_2 \times (gbest - s_i^k) \end{aligned} \tag{5.15}$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \tag{5.16}$$

where,

v_i^k	velocity of the particle i at iteration k
v_i^{k+1}	velocity of the particle i at iteration $k+1$
w	inertia weight
c_j	acceleration coefficients
$\text{rand}_{1,2}$	random numbers between 0 and 1
s_i^k	current position of i at iteration k
$pbest_i$	pbest of the particle i
$gbest$	*gbest* of the group
x^{k+1}	position of the particle at iteration $k+1$

In our experiment, there were 100 particles used and the number of generations was limited to 1000 generations. The maximum velocity of the particle was limited to the search space and any particle moving away from the problem space was moved back so that the range of the particle did not go beyond the boundary of the problem space.

5.4 Simulation Model and Parameters

5.4.1 Experimental Setup

In this model, two base stations A and B were selected which were separated by D metres. The mobile device moved from one cell to another with a constant velocity and the received signal strength is sampled at a constant distance (in metres). Here, the mobile node moves from one base station to the other with constant speed. The signals from the base stations are affected by two major factors: path loss and log-normal fading. Rayleigh fading is not taken into account as it is assumed that the rapid fluctuations are averaged out. The received signal strengths from the current base stations to the target base station are sampled at distances Kd_s. The model considered also includes slow fading. The received signal strengths a_t and b_t (in dB) when the mobile is at a given distance kd_s are given by

$$a_t = K_1 - K_2 \log kd_s + u_t \tag{5.17}$$

$$b_t = K_1 - K_2 \log (N - k)\, d_s + u_t \tag{5.18}$$

where $N = D/d_s$. The parameters $K_1 = 0$ and $K_2 = 30$ in dB which are typical of an urban environment accounting for path loss. K_1 is the signal strength at distance $d = 1$, and K_2 is the path loss component. Since the threshold levels are not considered, the received signal strengths depends on the difference in received signal strengths. The simulation parameters used for the movement detection are as shown below in the Table 5.1.

Number of Base Stations	2
Trajectory	Straight Path
Sampling distance	10 m
Distance between base stations	2000 m
Path loss (K)	30 db
Transmitter power	0 dB
Fading Process	Lognormal fading
Standard Deviation (u_k)	8dB

Table 5.1: The simulation parameters used for the prediction model

5.4.2 Experimental Setup for Self-Tuning Algorithm (STA)

In the hybrid model for prediction of signal strengths, the self-tuning algorithm was used as an optimisation technique to adjust the membership functions of the fuzzy inference rules. Recall that the error from the actual value and the predicted values from the basic prediction model are the errors that are treated as the input for the fuzzy model. In our experiments, the self-tuning algorithm was set to typical values which resulted in the optimum performance. According to the findings by many researchers, the initial conditions of the parameter a_{ij} are set so that the domain of the input x_{ij} is divided equally. The initial value of width b_{ij} is set to allow overlaps of membership functions. Fuzzy rules are performed for the input data $(x_1, x_2 \cdots x_m, y_r)$ using equations 5.1 to 5.3 and the membership functions μ_i of each inference rule and the output of the fuzzy reasoning are derived. Tuning of the real value w_i of the consequent part is performed by substituting the output of the fuzzy reasoning y, membership functions μ_i and output data y_r into equation 5.14. The tuning process of a_{ij} and b_{ij} is done using equations 5.12 and 5.13. The error is calculated using the error equation and until the value is less than a threshold value which, in our case, was less than 0.002. For experimental purposes, fixed values of a_{ij}, b_{ij} and w_i are chosen in such a way that the maximum error is within the range of the membership function. The simulation results of the STA based hybrid model are presented to show the number of iterations it takes for learning. Although the STA based hybrid model produced some good results with respect to improving the error from the prediction model it was very slow and the convergence achieved took many iterations. Details of the results and convergence are discussed in section 5.5.

5.4.3 Experimental Setup for Genetic Algorithm

In the hybrid model for prediction of signal strengths, the genetic algorithm was used as one of the optimisation technique to adjust the membership functions for the fuzzy inference rules. In our experiments, GA was set up with typical values which resulted in optimal performance. The initial population of GA was uniformly distributed over the entire search space. The life cycle of the GA model alters the population by crossover and mutation. Crossover recombines several possible solutions to form new intermediate possibilities. The selection and recombination keeps the population size constant. Mutation is done to alter the population slightly. In all the GA experiments for our GA based hybrid prediction model, we used a fixed population size of 100 individuals. The crossover operator used was a one point crossover with a crossover probability of 0.5 and the mutation probability was set to 0.3. No crossover or selection was performed if there were two or fewer individuals. In order to get a fair comparison between the models, with regards to the total number of evaluations, a population size of 100 individuals was selected. Although, we tried with 30 individuals in the PSO, due to the fact that GA requires a fairly large population size, we used 100 individuals as the GA performed below average with fewer individuals. In all our experiments with GA, we tried out the best possibilities for parameters to produce the best results. Similarly to the PSO based model, the GA model was set up with typical values to improve the prediction error from our basic Grey prediction model proposed in the chapter 4. In the following section we present our simulation results for the basic prediction model, the self-tuning algorithm based hybrid model, the PSO based prediction model and the GA based prediction model to compare their performances in terms of improvements in the prediction error. The results of comparison between the GA and PSO approaches in terms of convergence with respect to our model was also compared.

5.5 Results

In this chapter, we have investigated the best methodologies that could be used for the optimisation phase in our hybrid prediction model. We have compared three variants of the hybrid model for prediction of signal strengths for the mobility environment. The results of the Grey prediction are shown in Fig. 5.3 which is a plot of the actual values of received signal strength and their corresponding predicted values. The Grey model tracks the curve with some error which is shown in Fig. 5.4. The Grey model does not predict large variations in the input data. The error varied from as little as 0.02 to 0.6 db with the basic prediction model. However, to compensate, we used fuzzy parameters and fine-tuned them with the three optimisation algorithms. The parameters tuned by the optimisation algorithms are a_{ij}, b_{ij} and w_i to minimise the objective function shown in the equation 5.4. Further, the simulation serves two purposes: firstly, to help us decide which optimisation algorithm best suits our problem and secondly, to see the performance of our prediction methodology with the two evolutionary algorithms and the self-tuning algorithm proposed in [13].

5.5.1 Comparison of the Algorithms

The prediction methodologies explained in the previous sections used fuzzy parameters which were fine-tuned by optimisation algorithms. Using the Grey model for prediction of signal strength caused some errors. The compensated models for the genetic algorithm and PSO are plotted in Fig. 5.5 and Fig. 5.7 respectively. We also plotted the absolute errors for both the models as shown in Fig. 5.6 and Fig. 5.8 respectively. A closer look at the absolute errors shows that the errors with PSO were smaller than those obtained from the GA approach. This also means that the performance of PSO in achieving the result was much better overall. Since the self-tuning algorithm, is fairly complex in operation and due to the fact that it consumes more time we have not plotted the compensated output for the self-tuning algorithm. Our main focus was on the PSO and GA based models as they faired better than the self-tuning algorithm. Although STA has the capability to provide a good result, it showed poor results with our problem. The fuzzy parameters tuned using the self-tuning algorithm work with a learning constant set to the parameters initially, which reduces the error with every iteration. The self-tuning algorithm takes a very long time to converge to the minimum value set. Fig. 5.9 shows the convergence of the self-tuning algorithm, PSO and genetic algorithms. With our hybrid model, it is observed that the PSO has a better performance than the GA and the self-tuning algorithms.

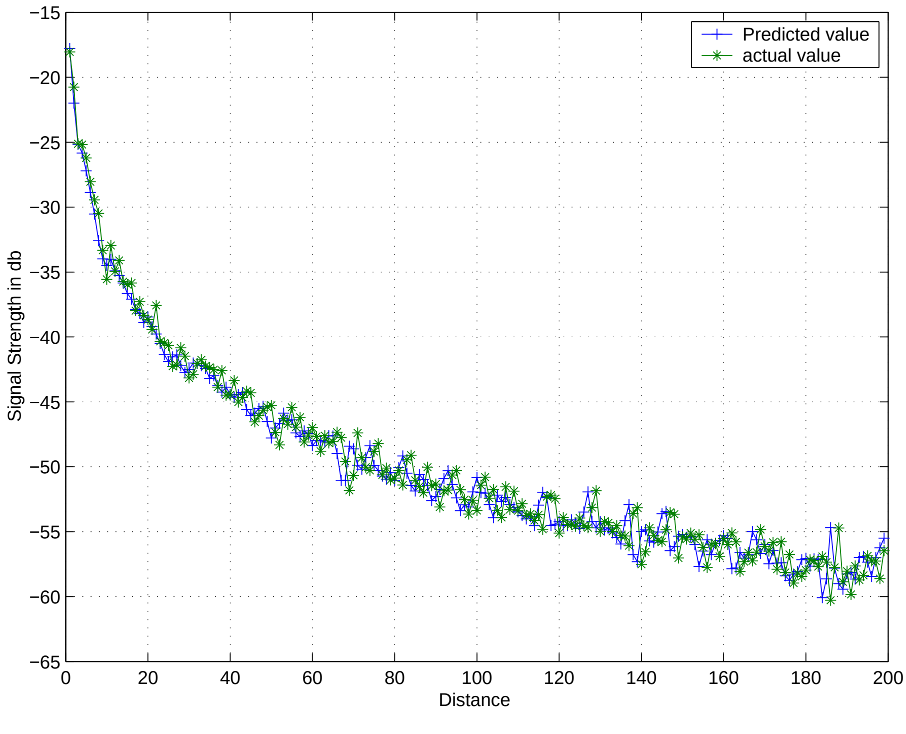

Figure 5.3: The received signal strength tracked by the Grey model.

The algorithms were run several times and in 90% of the cases the PSO converged

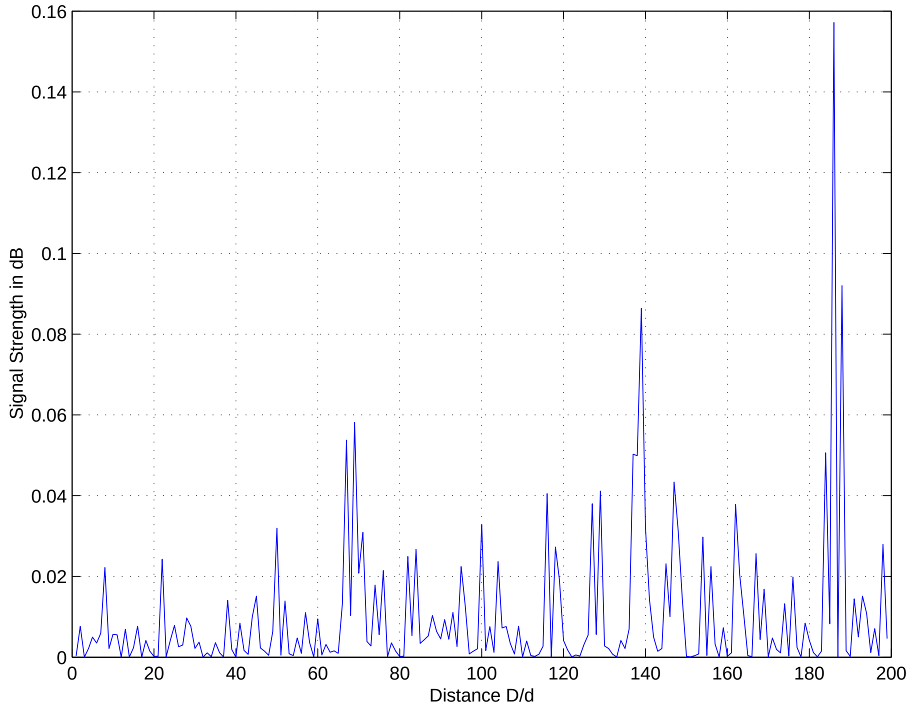

Figure 5.4: The absolute error from the Grey model.

faster than the genetic algorithm. The PSO reaches the desired fitness value in fewer iterations than the GA. This is mainly due to the population size that has been chosen initially. The Fig. 5.11 shows the convergence for the two evolutionary algorithms across several runs. In our experiment, for both PSO and GA based algorithms, the hybrid technique performs very well giving very small errors. The GA was not able to reach the optimum in any of the experiments as compared to the performance of the PSO method. This is probably due to the fairly small population size used in the GA. On the other hand, settings of the velocity factors mainly determine the performance of the PSO. Also, previous research by authors of [112] shows that PSO is not sensitive to population size. The following section provides a detailed discussion of the results and our choice of optimisation technique.

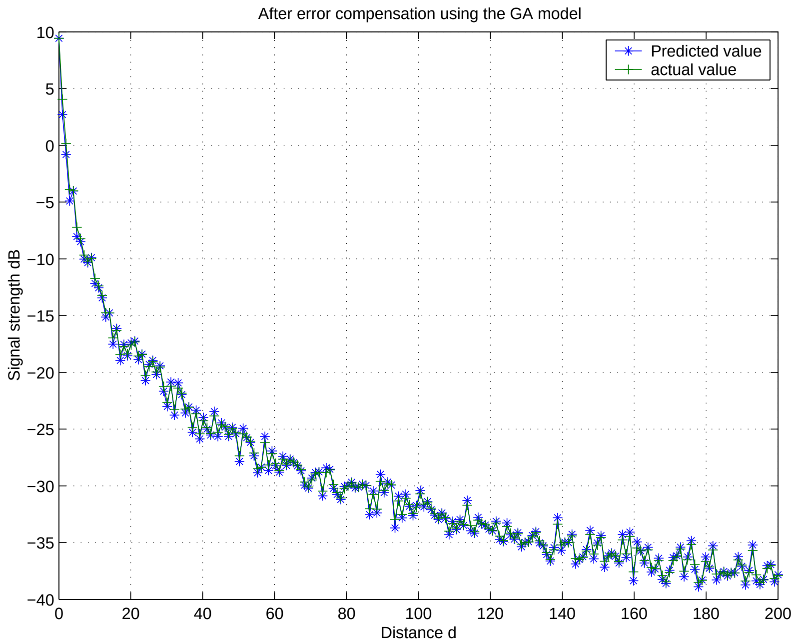

Figure 5.5: The received signal strength tracked by the genetic algorithm model.

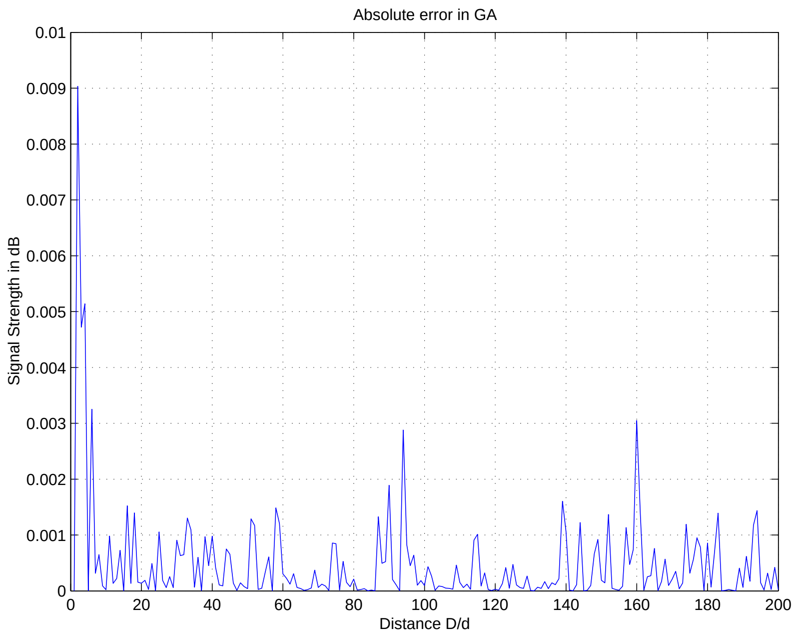

Figure 5.6: The absolute error after compensation by the genetic algorithm model.

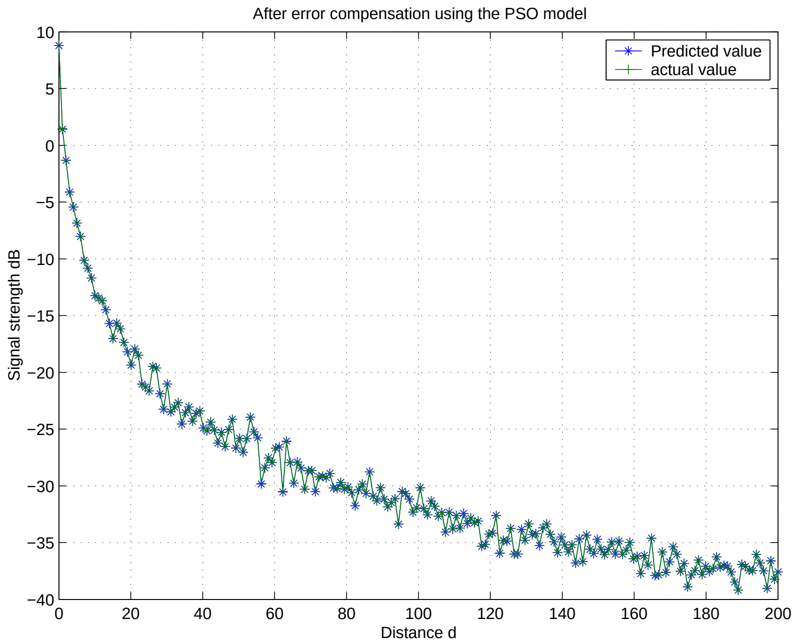

Figure 5.7: The received signal strength tracked by the PSO model.

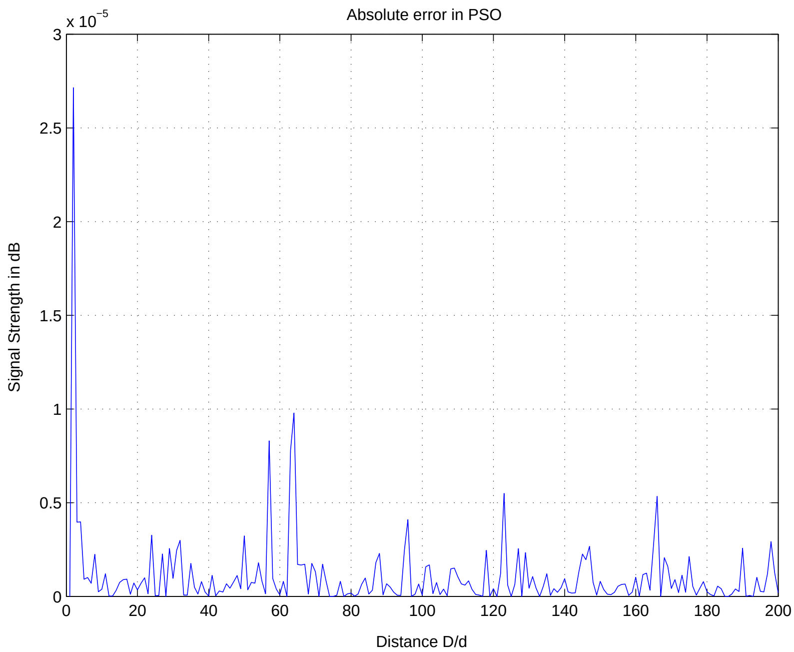

Figure 5.8: The absolute error after compensation by the PSO model.

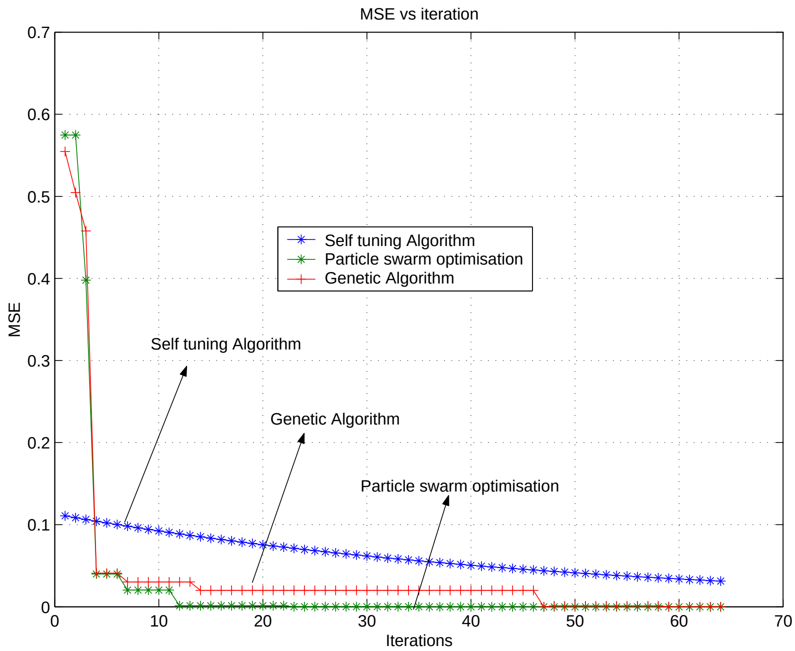

Figure 5.9: The outputs of the GA, PSO and the self-tuning algorithm.

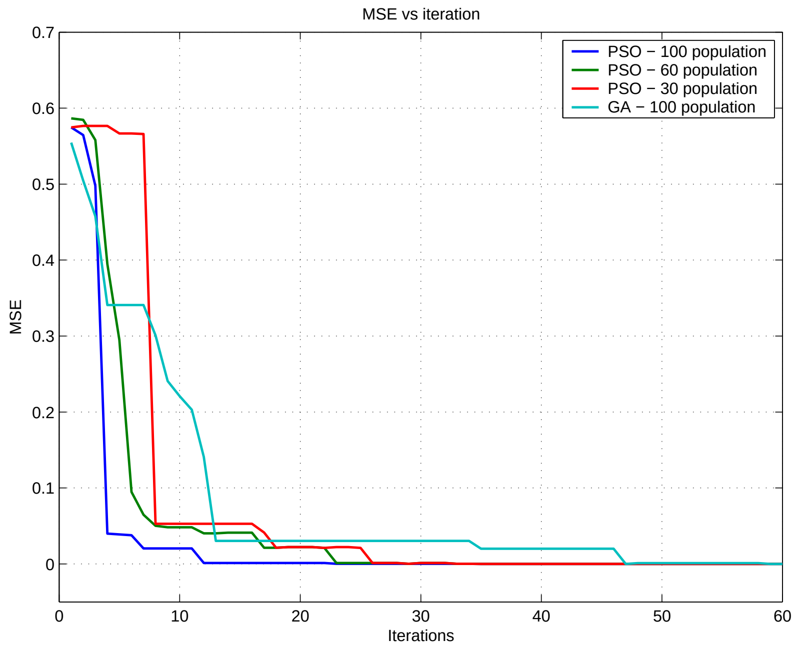

Figure 5.10: The outputs of the GA and PSO with population sizes 30, 60 and 100.

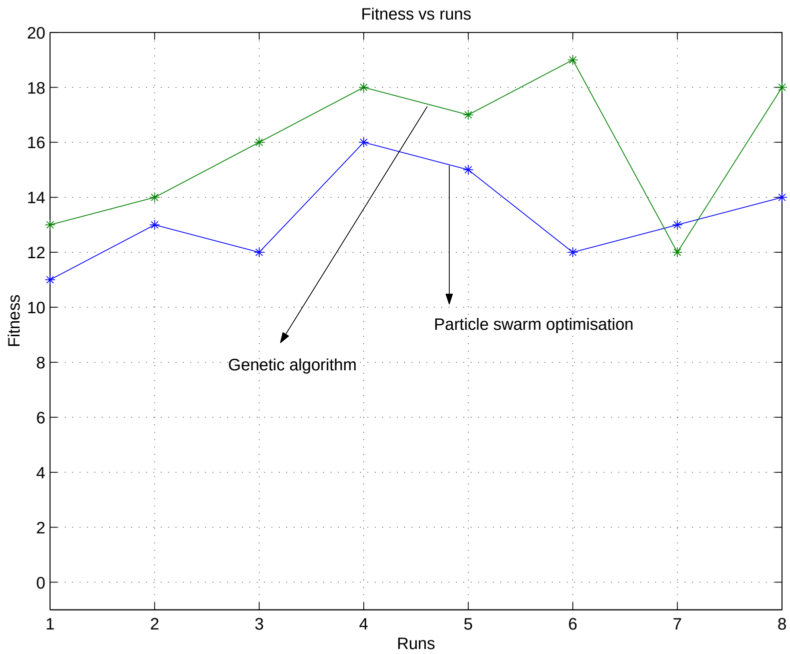

Figure 5.11: The comparison on fitness values for several runs for PSO and GA.

5.5.2 Genetic Algorithm vs. Particle Swarm Optimisation

Figs. 5.9 and 5.11 show the best average fitness for each generation, using both the standard genetic algorithm and the PSO model. The graphs illustrate the fitness achieved by the same population sizes for both the GA and the PSO. Initially, the PSO model was tested for different population sizes such as 30, 60 and 100. The Fig. 5.10 shows the average fitness for the models for different population sizes of PSO and the GA. With the GA model, when the population size was fixed at 100, the performance in achieving the desired result was much lower in comparison to PSO with a population size of 100. Therefore, any population size below 100 for GA was not performed. In most cases, the convergence of PSO with population size of 30 was found to be better than that for the GA with a population size of 100. Another important point to mention is that, in the PSO approach, for the purpose of adding diversity we cleared memory so that the new particles were unaware of the previously found optima. By adding diversity, we can avoid the problem of suboptimal solutions. It is to be noted that, there is also a risk involved in doing this as the particles may take longer to converge for a given fitness. According to our experiments, we concluded that GA was not able to deliver a better performance – mainly because of the smaller population size. Our motivation for trying out different combinations on the hybrid model was to show that each of these search techniques on its own has its specific problem-dependent strengths and weaknesses. However, their performance is highly dependent on their starting point and often results in premature convergence.

Although there are advantages of using evolutionary algorithms in the optimisation process, there are some drawbacks as well. The reason for using a particular type of algorithm is always problem dependent. This means that some problems are solved better by one technique and other problems can be solved with another technique. The selection of an evolutionary technique or PSO is again dependent on the problem and the parameters involved in the problem formulation. We could say that no optimal parameter setting applies to all the problems and tuning of these parameters can alter the performances of the system. Another important disadvantage of using an evolutionary technique or a PSO is the problem of premature convergence as they lose population diversity. In both PSO and GA approaches, good solutions that attract attention early may result in not searching the entire problem space which may also lead premature convergence. Population diversity is one important feature to be taken into account as it is vital for searching the problem space. It is also to be noted that too much population diversity can lead to enormous search spaces, too much computation time or individuals being unable to reach a desired fitness value.

We conclude that the choice of the appropriate optimisation technique is highly dependent on the problem and the objective of the search i.e. finding the solution to a given degree of precision or just finding a good solution.

5.6 Summary

In this thesis, a hybrid prediction model that integrated particle swarm optimisation, self-tuning algorithm or genetic algorithm was proposed. Here, we compared the self-tuning algorithm along with the PSO and GA. We have discussed the application of evolutionary computing techniques to find the optimum way of reducing the error by fine tuning it with fuzzy parameters and optimisation algorithms. The Grey model was used as the prediction methodology and errors were compensated for by the use of evolutionary techniques. Our algorithm incorporating PSO showed better accuracy in comparison with any of the existing prediction methodologies. We have compared the two evolutionary techniques, namely, the genetic algorithms and particle swarm optimisation, in terms of convergence. Each of these search techniques on its own has its specific problem dependent strengths and weaknesses. GA's, for instance, are widely applicable and particularly powerful when domain knowledge can be incorporated into the operator design. However, particle swarm optimisation (PSO) can achieve clearly superior results in many instances of numerical optimisation, but there is no general superiority compared to GA's.

Although there exists a huge literature base on evolutionary techniques and optimisation algorithms, we can broadly classify them into deterministic techniques and stochastic techniques. We also know that deterministic optimisation techniques are very simple and are often able to give accurate and fast results just as with stochastic techniques. However, deterministic techniques depend a lot on the initial values and are suboptimal when compared to stochastic techniques. The reason behind our choice of the PSO approach is that it is a stochastic technique. One of its key advantages is that it is easy to implement in both coding and the selection of the parameters. The standard PSO proposed by Drs Eberhart and Kennedy consists of the straight-forward vector equations for the update of velocity and position and there is a simple loop to make decisions on the *pbest* and *gbest* values. By comparison with GA, or any other evolutionary technique, it is simpler to implement as it does not have any mutation operators or crossover operators. In addition, the variation of PSO presented in our thesis is not a major variation as we have only changed it according to the requirements of the problem formulation. Another advantage of using PSO is that it is easier to control the search properties and convergence by variation in parameter values. Due to the fact that it has update equations it is easier for us to track the particles in each loop. It is to be noted that PSO gives the luxury of parameter assignment and is very flexible. Also the direction of each particle can be adjusted and kept within the limit of the search space which is not the case of genetic algorithms or any other evolutionary technique. Finally to summarise, the PSO model that we chose has proven to be a much better optimisation technology to our problem than the traditional GA method.

We can conclude that, in the context of our problem, the GA did not perform as well as the PSO because a GA needs a bigger population size. But bigger population sizes need more computation time and do not serve the purpose of using it in realtime. The GA algorithm works better for more individuals (increased population size) to find a good solution that it can mutate. The PSO, on the other hand, has particles which are there ‘forever’ and can locate better results in the search space. Thus, our proposed prediction technique performs best with particle swarm optimisation rather than the traditional genetic algorithm. In future work, we shall attempt to improve the performance of the evolutionary algorithms so that they converge at a faster rate.

Chapter 6

Handoff and Multicasting in Wireless Networks

6.1 Introduction

In previous chapters, we showed that we can have accurate mobility prediction which can detect the position of a mobile exactly from the base station which will help us to predict the future direction that the mobile node is likely to move. By doing this, we not only get details of the mobile node's position but also we can reserve resources for the handoff at the best locations. Here we have used the hybrid prediction model as the detection technique for finding the target cell so that we only reserve resources in that particular cell. But, reserving unwanted resources will reduce system performance. Therefore, in order to achieve the best system performance we have to accurately reserve the minimum resource levels. Some of the research done earlier on reserving resources proposed that resources be permanently reserved for handoff. By doing this, we can achieve only marginal gains in system performance. Therefore, such permanent reservation schemes are not a good idea. Moreover, very low blocking probabilities are required in future generation networks which can only be achieved by lowering the operating utilisation of the network. A lower utilisation means that the number of users that can be supported by a given amount of bandwidth is reduced, resulting in a higher cost for the user. Thus, lower utilisation does not seem to be an effective solution for future mobile networks. Furthermore, due to the growing demand for more bandwidth for broadband services, this becomes a dilemma for network designers in dimensioning mobile networks. So far no detailed study has been done on the interaction of mobility prediction and multicasting techniques to improve performance of handover in wireless networks.

Achieving seamless mobility is a significant challenge for wireless networking today. In this chapter, we shall try to achieve efficient handover by using our mobility prediction approach and, in addition, propose an approach based on simple multicasting techniques which can improve system performance and achieve higher utilisation. We propose two algorithms that will use mobility information and optimise the resource reservation in cellular networks. Improving handover performance in mobile networks has been the subject of numerous technical publications. In this research, we propose an innovative approach to improve the handover performance of cellular networks, by using mobility prediction combined with multicasting techniques to reserve only the necessary resources. This extra mobility prediction information can be used to optimise the resource reservation process for new and handover users.

Effective handoff holds the key to defining optimal performance of wireless networks since if this is not performed efficiently, there could be packet losses during handoff as the mobile node moves from one point of attachment to another. A new method of determining a multicast tree routing scheme with specific performance objectives is presented in this chapter. A situation is modelled where a multicast tree is defined covering multiple access routers (AR) to maintain connectivity with the mobile node using mobility prediction (by selecting the least number of access routers) whilst ensuring guarantees of bandwidth and minimum hop count such that packet loss can be avoided. To simultaneously solve the above two problem formulations gives rise to a multi-objective optimisation problem. Discovering optimal routing is an NP hard problem where network state information is not accurate – this is a common feature in wireless networks.

In this chapter, we provide a solution for the problem by proposing a method to predict the signal strengths using the hybrid model proposed previously in chapter 5 and allocate the right level of resources in the target cell for a given time depending on application demands. The two objectives considered are the residual bandwidth and the minimum number of hops. We propose a handover strategy which uses the signal strength prediction and multicasting algorithms for the scenario depicted in Fig. 6.3. The proposed strategy consists of two main components, viz:

- **Hybrid mobility prediction model:** A mobility prediction algorithm to predict the probable target cell based on the received signal strengths from the base station.
- **Multicasting algorithms:** Two multicasting algorithms namely the MMP algorithm and K-Minhop algorithm proposed on the objectives, residual bandwidth and number of hops respectively.

These two essential components of the system are explained in detail in the following sections of this chapter and they are compared to the CAR-set algorithm proposed by Helmy et. al. It is important to emphasise here that the mobility prediction algorithm used, is the hybrid mobility prediction model which is based on the Grey prediction, fuzzy inference rules and particle swarm optimisation as discussed in earlier chapters. After describing the problem, an algorithm that satisfies the constraints and objectives with a near optimal cost is presented. The work in this chapter led to publications [9, 11]. The contributions of this chapter can be summarised as follows:

1. Development of a mathematical model for optimisation of building a minimum cost multicast tree. A situation is modelled where a multicast tree is defined covering multiple access routers (AR) to maintain connectivity with the mobile node using mobility prediction (by selecting the least number of access routers) whilst ensuring guarantees of bandwidth and minimum hop count. To simultaneously solve the above two problem formulations gives rise to a multi-objective optimisation problem.
2. To solve the problem as formulated, two algorithms are proposed based on the residual bandwidth and minimum hop count, viz: The MMP and the K-MinHop algorithms respectively.
3. Numerical results are presented for various network sizes and topologies to determine the accuracy of the proposed algorithms.

The rest of the chapter is organised as follows: Section 6.2 presents the related work on multicasting techniques that have been proposed to improve handoff performance in cellular networks. Section 6.3 introduces the CAR-set algorithm proposed by Helmy et. al. in the M&M protocol. Section 6.3.2 presents the concept of mobility prediction as discussed in chapters 3, 4 and 5. This section gives an overview of how it is used in prediction of signal strengths that can help in improving handover performance. The hybrid model required for prediction of received signal strengths is discussed to determine the position of the mobile node. The parameters that can be used to characterise mobility are presented in Table 6.1. Section 6.4 describes the spanning tree algorithm and the basic definition of this problem formulation. Section 6.5 describes the Framework/Architecture for the given scenario. This section also describes details of multicasting techniques that are used in our framework. In section 6.6, we describe the simulation model that we used in our experiments. Section 6.7 describes the problem formulation. The section 6.8 describes the two proposed algorithms, pseudocode is presented together with a detailed explanation.

Finally, this chapter concludes with a study that validates the proposed algorithm via simulation.

6.2 Related work

Good mobility prediction [4] holds the key to improving performance when calls are being handed-off between cells. When handoff is being performed, there can potentially be a loss of packets. In order to improve handoff performance, mobility prediction is considered an important technique [56]. Mobility prediction is used to identify the minimum number of access routers that need to be readied on order to provide connectivity to the source router servicing the mobile node. This minimum set of nodes lie on the leaves of a multicast tree that is to be constructed by our algorithm.

Multicasting is an efficient means of group communication. It has been used for video-conferencing and many other real-time applications, the advantage being bandwidth savings. Routing packets from the source to the destination in an optimal way not only decreases end-to-end delay but it also saves on network resources. These factors also influence handoff performance in terms of handover delay and packet loss. There are many protocols that have been proposed to overcome these performance limitations in the literature. Methods for alleviating these problems have been described in several studies [78]. Several protocols have been proposed for wireless networks [26, 52, 130, 131]. Although they solve several important problems, some are quite complicated and still have problems in achieving an optimal multicast tree and do not meet the necessary QoS requirements. Some of the schemes claim advantages through using a shortest route path algorithm to minimise resources and to achieve QoS. However, whenever the mobility factors involved are high, building the multicast tree may lead to significant overhead requirements. Furthermore, minimum join and prune operations are required so that there is minimum disruption to the multicast tree.

In the MMA protocol [132], the authors propose a way to reduce the number of joins by calculating the time that the mobile node spends in the previous cell area. Thus, if the expected time for a visiting mobile node is not long enough, it does not invoke the join process. In a further extension to the MMA protocol, it not only reduces unnecessary joins but it also reduces the duplication of packets. The MOM protocol [69, 70], proposed by Harrison et. al. involves tunnelling of multicast datagrams. This scheme reduces the number of duplicate multicast datagrams and any additional load on the wireless links, which are of low bandwidth. The approach primarily focuses on scalability with respect to group size, number of multicast groups and mobile hosts.

In Seshan et. al., the authors propose a multicast-based algorithm which claims to use *"hints"* from the wireless network to improve the handoff performance. These hints provide a tradeoff to wired network bandwidth for reduced handoff latency [29]. However, the proposed protocol does not suggest any technique for obtaining these hints.

In multicast based mobility (M&M) [28] proposed by Helmy et. al. they define a set of protocol suites to enable multiple access routers to receive traffic for the mobile host. Here, a group called the CAR-set (Coverage Access Router set) of access routers is highlighted or selected to provide the necessary coverage to the mobile node so that there is no loss of packets during handoff. However, this method utilises more network resources and there are more access routers selected into the multicast tree. This, in turn, can result in more unnecessary join and prune operations. As a mobile node moves from one coverage area to the next, the access routers are selected in such a way that there are no losses. For example, in a cellular network, all the cells surrounding the current cell or the active cell is selected. Doing so leads to multicast overheads in terms of bandwidth usage and delay as they use more network resources.

These issues along with low bandwidth and higher bit-error rates in wireless networks make efficient multicasting a challenging task in a mobile environment. An approach to IP mobility using standard multicasting has also been proposed in the literature [52, 133]. In this approach, the mobile node is assigned a multicast address through which it joins the access routers that it visits during its movement. Handover is performed by standard IP-multicast join and prune mechanisms [26]. Further, dynamic algorithms can be designed to identify probable new access routers [AR]. If there is replication of packets, there should be a heuristic, that will reduce this overhead. When the old AR sees that the signal from a mobile node is fading (and this is an indication of the onset of a handover condition), it triggers the AR's in the vicinity to join the multicast group. To avoid packet losses, handover must be detected early enough to provide an adequate time margin before actual handover takes place [25]. Once the mobile node is connected to the new AR, the remaining set of AR's will be removed from the group. However, this chapter proposes a solution that reduces such overheads by performing accurate mobility prediction which can select a potential AR. In this chapter, the formation of the near optimal multicast tree problem is considered. The main idea is to establish a multicast session from the source to these potential AR's to compute a minimum cost tree with specific constraints. The chapter discusses the details of how mobility prediction can help multicast routing that will improve handoff performance.

6.3 Multicast based Handover Algorithms

In this section we compare our mobility prediction based multicast algorithms and the M&M protocol which has been proposed to improve the performance of handoff in wireless networks. In both techniques, the assumption is that the Corresponding Node [CN] wishing to send information to the Mobile Node has to send its packets via these access routers. A number of access points [AP] can be connected to the access routers [AR]. Each AP covers a region called a cell area.

6.3.1 CAR-set Algorithm

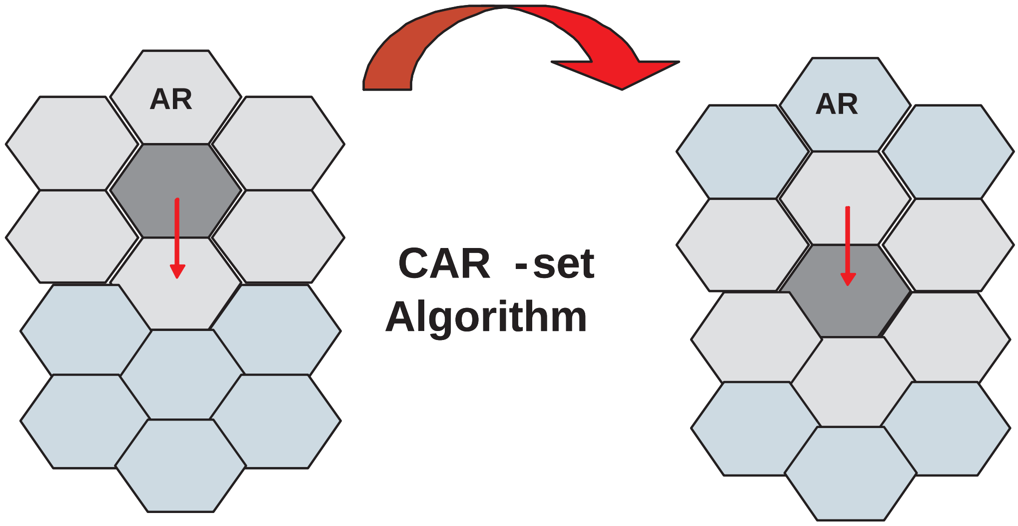

Figure 6.1: Figure showing the selection criteria in the CAR-set algorithm

The main goal of the work done by Helmy et. al. was to design the multicast based micro-mobility protocol. The work by Helmy et. al. proposes to reserve resources using multicasting techniques such that there is no loss of packets. Some of the issues addressed while designing the protocol are choice of underlying protocol, mobility detection and handoff procedures. Figure 6.1 shows the 7-segment cell structure which is used by the CAR-set algorithm proposed in the protocol.

In the framework of the CAR-set algorithm, a set of potential ARs called the coverage access router set (CAR-set) are defined. The routers in the CAR-set are adjacent to the serving AR. The serving AR is also called the head of the CAR-set.

The adjacency is established based on the adjacency of the radio coverage of the serving AR as shown in the Fig. 6.1. In Fig. 6.1, the dark hexagonal cell is the head and the surrounding cells are the members of the CAR-set. As the mobile node moves, the serving AR changes, by which the members join and leave the CAR-set thus maintaining connectivity for the mobile node.

In this algorithm, all the cells surrounding the current cell are selected for the multicast group. This also means that the algorithm reserves resources in all the cells in anticipation for an imminent handoff that the mobile node may enter the cell. The Fig. 6.1 clearly depicts the procedure of reserving resources by the CAR-set algorithm. The CAR-set algorithm described in this section accomplishes a proactive route by initiating an multicast join from the source to all the members of the CAR-set. They remain joined to the tree as long as the MN is not connected to the new AR or the new base station. It is important to state here that, the route is established from the source AR to all the destinations in the CAR-set. The prune process is initiated by the new AR after the MN achieves a stable network connectivity in its coverage area. The CAR-set algorithm also assumes that the route established from the source to all the destinations in the CAR-set is the optimal path. In the example Fig. 6.1, all the cells surrounding the current cell are highlighted. The CAR-set also makes sure that there are no losses during handoff because it reserves unnecessary resources as it does not have a detection procedure to accurately determine the next potential target cell. In our research, we particularly explore this disadvantage by using the proposed hybrid model to detect the next potential cell as well as reserving the right amount of resources by establishing the least cost path from the source to the destination AR.

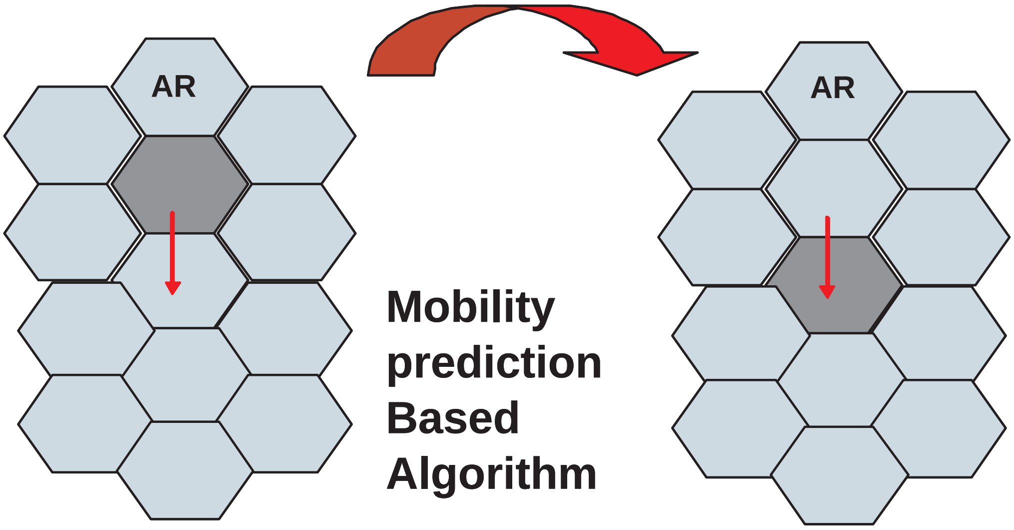

Figure 6.2: Figure showing the selection criteria in the proposed technique

6.3.2 Mobility Prediction Based Model

In this section, we present the mobility prediction based multicasting technique to improve handoff performance. In comparison to the CAR-set algorithm, this technique does not have the overheads associated with replicating the packets to all the surrounding cells of the potential cell. Multicasting packets to the CAR-set improves handoff performance at the cost of added overheads over wired links that lead to ARs that belong to the CAR-set. These overheads depend on the size of the group whose members are highlighted and the duration for which the member routers are connected to the multicast tree. It is important to state that the highlighted cells are connected via ARs which are, in turn, the part of the wired network. The idea behind this is that the multicast tree is formed from all the selected ARs to the source node. In our technique, this overhead is reduced with the help of the prediction methodology as we highlight only the potential ARs to be connected to that cell. In our proposed technique, we use the hybrid prediction model based on the PSO model. The selection of the potential cell is as shown in Fig. 6.2. Now, after we have selected only the potential AR, we construct a multicast tree from the destination node to the source node (CN) based on the constraints of bandwidths and the minimum number of hops. It is to be noted that our model does not use any bandwidth in the wireless link from the AR to the mobile node. The algorithms for establishing a path from the selected potential AR to the source node is based on two algorithms that are discussed in the sections below. We also explain the basics for the spanning tree algorithm, our problem formulations and proposed framework for multicasting in a wireless environment.

6.4 Spanning Tree Algorithm

In graph theory, we associate weights or costs on each edge as adopted in many applications [134]. In traditional graph theory problems, two problems are considered which are 1) to find the lowest cost path between two points and 2) the way to connect these points. Researchers have been using the minimum spanning tree (MST) for numerous applications in telecommunications. By definition, a minimum spanning tree of a weighted graph is a spanning tree whose total weight is no larger than that of any other spanning tree. There are situations when the edges may have equal weight which may lead to examples where spanning trees are not unique. Therefore, it is necessary to consider equal weights carefully as they are not unusual in many applications. Many of these problems are addressed in MST with unidirectional and bidirectional graphs [134, 135]. Although we are not generally interested in these special problems in the MST algorithm, we have implemented the MST in our problem to construct to find a suitable multicast tree. The issues concerning MST have been addressed in numerous papers and these problems are not within the scope of this research. However, there have been many algorithms proposed that compute the MST and these include Prim's algorithm, Kruskal's algorithm and

Boruvka's algorithm. Although each of these algorithms perform the function of computing a minimum spanning tree, they all differ slightly.

In our research problem, after we have found the potential base station or AR (destinations), we construct the lowest cost path to the source from the destination and generate the multicast tree for which we apply the MST. These two low cost paths found are based on the objectives of bandwidth available or the minimum number of hops depending of the requirements of the problem. In our problem, consider the network topology shown in Fig. 6.3. For any multicast connection, the source is the corresponding node and the receiver is a set of candidate routers which are serving the mobile node. There are many algorithms that can be used to choose the minimum number of hops to these destinations in wired and wireless networks. One such example is the well-known Dijkstra algorithm – used to find the shortest path from a given source node to one or many destinations. Depending on the different optimisation goals, a multicast tree can be constructed with a minimum cost such as delay and bandwidth optimisations on the link. The techniques for multicasting that perform cost optimisation for a tree can be regarded as a minimum Steiner tree problem and finding such a tree is known to be an NP-complete problem [136, 137]. Therefore, the problem of computing the minimum cost tree for a given multicast tree with a source and a set of destinations D can be modelled as a Steiner tree problem [134, 138]. When there are additional constraints such as the need for available bandwidth limits on a directed link, the problem becomes the directed Steiner problem which has an objective of finding the minimum cost rooted at s and spanning all the nodes in D, which can be defined as follows:

Given a directed graph $G = (V, E)$ with a specific source node $s \in V$, and a set of destinations $D \subseteq V$, the objective is to find the minimum spanning tree rooted at s and spanning all the nodes in D.

6.5 Framework/Architecture

The network model that we consider is a wireline/wireless network with a number of access routers connected together and is shown in the Fig. 6.3. The Corresponding Nodes [CN] wishing to send information to the Mobile Node [MN] have to send their packets via these access routers. A number of access points [AP] can be connected to the access routers [AR]. Each AP covers a region called a cell area. When a mobile node moves from one AP to the other without changing the AR it is called an intra-AR handoff and when it changes from one AR to another it is called an inter-AR handoff. An access point that is connected to the access router serves a mobile node. A mobile node, throughout its movement join and leave these access points. The access point acts as the radio point of contact to the mobile node. An AR considers that each AP is on a separate subnet [28].

Consider a mobile user in a cell using some mobile communication service (call this cell '0' connected to AR0). Assume that there are n neighbouring cells surrounding this cell '0'. When the user makes a handover, any one of the neighbouring cells may be the potential cell. Therefore when we want to reserve resources, a path or a tree must be built to support this handover. In this scenario, we have a mobile node moving in a straight line. It was intended at the beginning of this research to obtain some mobility data from the existing mobile communication network to create the architecture. However, due to the commercial nature of the data, we are not successful in obtaining the data. Therefore it was decided to create a model as shown in the Fig. 6.3. Furthermore, all the experiments done hold equally good for any realistic mobility model as it is a simple model.

Traditionally in cellular networks, cells are clustered into a cluster of 7 due to the frequency reuse factor. In this model, we have used the 7 cell model and assume that each cell is under one access router. In this framework, the mobile node periodically measures the signal strength from the base station or access point. Our prediction algorithm predicts the future received signal strength and selects the potential cell or the AR. Once the selection process is complete, the resources necessary for the handover are reserved (by multicasting techniques) taking into consideration, the bandwidth requirements or the minimum number of hops depending on the type of application running on the mobile node.

Multicasting has been chosen as the technique for reserving resources as we can construct a minimum spanning tree from the source to the destination predicted by the mobility prediction algorithm. A minimum cost tree will reduce the overall transmission time and will reduce the required bandwidth. Obtaining the network topology graph will be vital and so is the computation of the minimum cost tree. The construction of the network topology graph G requires the selection of a suitable subset of nodes. Once the source has guaranteed the topology graph G, a multi-communication tree i.e., a minimal set of routes to the destination D is computed. If there is only one destination node, a single source shortest path algorithm such as the Dijkstra algorithm can be used on G with source s. Given a graph G constructing a minimum cost tree that covers a specific number of nodes is also called the Steiner tree problem for a given set of nodes in a network. This can be classified as an NP hard optimisation problem. There are a number of heuristic algorithms that have been proposed for the above problem [24, 139].

6.6 Simulation Model and Parameters

Terminologies used to describe the hybrid prediction model were presented in chapters 3, 4 and 5 are maintained throughout this chapter. In this chapter, we investigate the performance of handoff which is assisted by mobility prediction and multicasting techniques. As discussed previously, we know that by mobility predic-

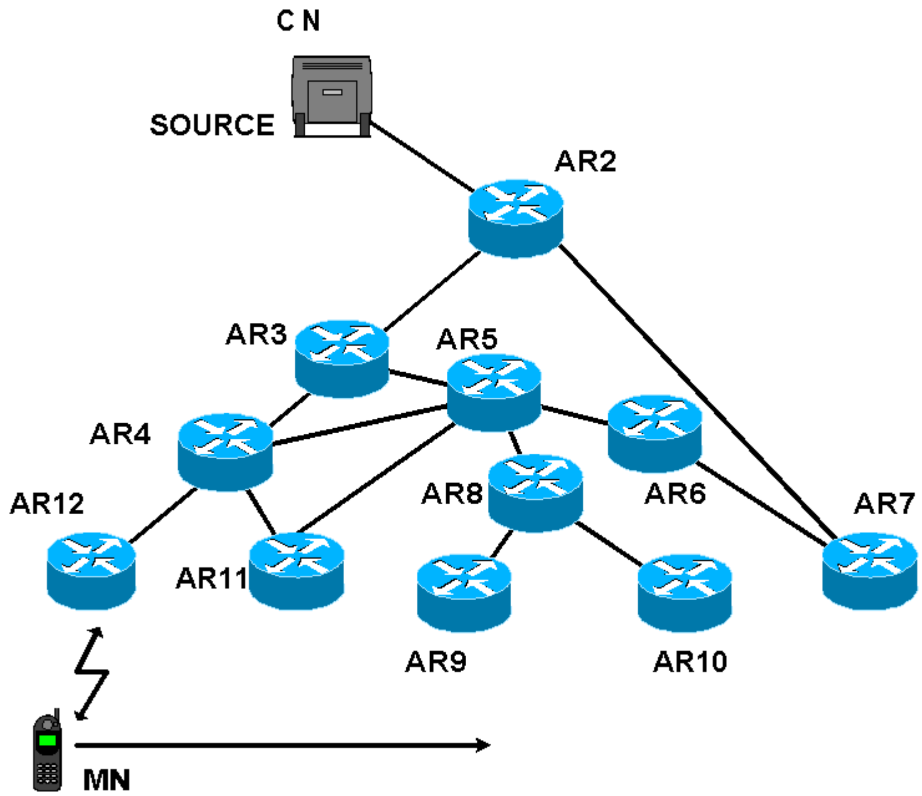

Figure 6.3: 10 node access router network and problem formulation

tion we could find the potential base station or access point to which the call can be handed over. By selecting the potential target cell, we can reserve resources that could help in handoff performance. Both the speed and direction of users can be easily derived from position measurements taken over a period of time. Therefore, to characterise the mobility of users we need to periodically measure the position of the mobile users. Systems like the Global Positioning Service (GPS) can be used to accurately determine the position of a mobile user. However, due to the additional cost in providing GPS to mobile users this is currently not a very attractive solution for a mass market. Instead of an additional system like GPS, the strengths of signals from various base stations received by a mobile terminal can be used to estimate the position of a mobile user – typically by using triangulation or similar methodologies.

Modelling the radio channel has historically been one of the most difficult areas of mobile radio systems. Multipath fading, shadowing from obstructions in the line of sight propagation path, doppler effects due to the mobility of users and reflections from the ground and other obstructions are just a few of the effects that contribute to the complexity of a mobile channel model. As our interest is on the processing of the signal strengths, we use these methods to convert the signal strength measurements to position measurements and decided to use a hybrid prediction model. The model that we used to convert the signal strength measurements to distance measurements is given in equations 6.1 and 6.2.

In this model, we have selected two base stations A and B, which are separated by D metres. The mobile device moves from one cell to another with a constant velocity and the received signal strength is sampled at a constant distance d_s in metres. Our model includes slow fading. The received signal strengths a_t and b_t (in dB) when the mobile is at a given distance kd_s are given by

$$a_t = K_1 - K_2 \log kd_s + u_t \tag{6.1}$$

$$b_t = K_1 - K_2 \log (N - k)\, d_s + u_t \tag{6.2}$$

where $N = D/d_s$. The parameters $K_1 = 0$ and $K_2 = 30$ in dB are typical of an urban environment accounting for path loss. The simulation parameters used for the movement detection are as shown below in the Table 6.1.

6.7 Problem Formulation

The Multicast Routing problem with bandwidth and number of hops as constraints can be formally stated as follows: ***Problem:*** Given a network $G = (V, E), \{c_l = 1/b_l\}_{l \in E}$, a source node $s \in V$, multicast group $M \subseteq V - s$, find a tree $\mathbf{T}$ rooted at s and

Number of Base Stations	2
Trajectory	Straight Path
Sampling distance	10 m
Distance between base stations	2000 m
Path loss (K)	30 db
Transmitter power	0 dB
Fading Process	Lognormal fading
Standard Deviation (u_k)	8dB

Table 6.1: The simulation parameters used for the prediction model

spanning all of the nodes in M such that $c(T)$ and the total number of hops from s to all the nodes in M is minimised. $c(T)$ is defined as

$$w_1 \sum_{l \in E} c_l(b_l) + w_2 \sum_{l \in E} c_l(d_l) \tag{6.3}$$

where $w_1 + w_2 = 1$ and w_1, w_2 are weighting factors, b_l is the available bandwidth on the link l and d_l is hop count.

The above routing problem involves finding the optimal multicast tree when the source and all destination nodes are given. For some delay sensitive applications such as VoIP, one will put a higher weight for w_2 to ensure a tree with a minimum hop count is chosen to minimise the end-to-end delay.

6.8 Proposed Algorithms

6.8.1 Tree Construction and Coding

Using an initial evaluation and simulation study, mobility detection (mainly prediction) and handoff performance were found to be very critical in multicast based handoff performance. Hence more emphasis was given to the design of mobility prediction (as discussed in chapters 3, 4 and 5) and multicasting techniques. For multicasting to be useful, it is important to construct a minimum cost tree for efficient handover performance. The assumption is that the MN would handover to only one base station/access point under the AR even though it is covered by many base stations/access points as directed by the mobility prediction algorithm. An implicit assumption is that the link layer is uniform everywhere and MN can communicate with any base station/AP within the range of MN.

Prim's or Kruskal's algorithm are perhaps the simplest Minimum Spanning Tree (MST) algorithms and represent the method of choice for dense graphs. Trees are the minimal graphs that connect any set of nodes, thus permitting all nodes to communicate with each other without any redundancies. According to [134], MST algorithms involve certain abstract operations which are to:

- Find a minimal edge connecting two subtrees.
- Determine if the edge would create a cycle.
- Delete the longest edge on a cycle.

The basic difference between Prim's and Kruskal's is that, Prim's builds MST one edge at a time finding the new edge to attach to a single growing tree at each step; whereas Kruskal's also builds the MST one edge at a time but, in contrast, it finds an edge that connects two trees in the forest of growing MST subtrees. In our tree construction, we have used Kruskal's algorithm to build the MST and this was coded according to [134]. Some of the disadvantages that need to be taken care of before implementing our multicasting algorithms are as follows:

- The MN should not know about the mobility scheme that is being used.
- Suboptimal routes can waste bandwidth.

In summary, the key here is to find the shortest distance from each non-tree vertex to the tree. In this problem, (Fig. 6.3) the MST algorithm is modified and used to find the tree for all the destination ARs specified by the mobility prediction algorithm. The topology used here is a very simple topology with a maximum depth of 3 (intermediate nodes between source and destination) from the source node. Packets are assumed to be generated at the source node ("correspondent node") and are destined to reach the mobile node via intermediate access routers. Although the ARs are near each other, as seen in the topology figure, their coordinates may differ and any AR could be the neighbour to any other AR. The main idea is to simulate the behaviour of a mobile network and show that the routers are interconnected. In this section, two algorithms for evaluating the minimal cost spanning tree (MMP algorithm-Multicast Mobility Prediction) and the least hop count approach are proposed separately. The pseudo-code and a short description of the step by step process is given in the following subsection.

6.8.2 MMP Algorithm

Input: signal strength values, a set of minimum AR's as predicted by the mobility prediction algorithm.

Output: A minimum multicast tree that satisfies the objectives from source to destination.

Method:

---

1 Run the Grey prediction algorithm to select the potential AR's
2 Run the MST algorithm to find the routes to all the selected AR's
3 Start with source s to all the nodes in $M \subseteq V - s$.
4 Do, $d_i = d_j + c_{ij}$, where i is the current source, until the distance for all nodes has been calculated.
5 For every destination; backtrack all the links that will be used in the spanning tree.
6 Mark the link (i, j) where $(d_i > d_j)$
7 Remove unmarked links from the spanning tree
8 Result is a spanning tree from s to nodes in M with minimum cost

Steps involved: Minimum Cost:

Step 1: Start with the source s, compute the distances d_i for the nodes spanned from the current source i to be $d_i = d_j + c_{ij}$ once the distance to all the nodes has been found, go to step 2.

Step 2: For every destination; back-track and mark all the links that will be used in the spanning tree. Mark the link (i, j) if $(d_i > d_j)$ (numbering to track).

Step 3: Remove the unmarked link from the spanning tree or prune the other links that are not supposed to be in the tree. The remaining links form the multicast tree with the minimum cost.

6.8.3 K-Minhop Algorithm

The general idea behind our algorithm is to determine the optimal feasible route in the multicast tree from source to destination. This section describes the algorithm and discusses its operation.

Input: A graph $G = (V, E)$ with minimum hop as constraints from a source s to a set of destinations.

Output: A minimum hop count that satisfies the constraints from source to the destination.

Method:

1 Set the costs of the edges to unity.

2 For each destination D in M.

3 Run the K-shortest path algorithm from the source s until the destination D is reached.

4 The result is the total number of paths for a given source to all destinations.

Minimum hop count:

Take the graph G and set all $c_{ij} = 1$ and then find the shortest path from source node s to all destinations D.

In principle, there are two alternative solutions, one involving minimum cost and the other requiring a minimum number of hops. Solving the minimum cost and minimum hop count problems gives two extreme solutions which may differ significantly. It is good to find a near optimal tree with respect to these two objectives by exploring all the other trees "between" these two extreme solutions. A set of candidate tree solution can be found by exploring a *"good"* set of solutions from the source s to all destinations D. This set of *"good"* solutions can be obtained by applying a k-shortest path algorithm. For every solution provided, we can get possible routes based on residual bandwidth as well as minimum hop count. Further, the final solution depends entirely on the type of application. To provide flexibility in selecting an appropriate outcome, we assign "weights" to enable us to bias the solution towards one outcome over the other. The cost is evaluated according to equation 6.3 and can be summarised as:

$$\text{Total Cost} = w_1.c_1 + w_2.c_2, \tag{6.4}$$

where, $w_1 + w_2 = 1$ and,

c_1	cost from the MMP algorithm *w.r.t* the bandwidth,
c_2	cost from the K-Minhop algorithm *w.r.t* the total number of hops.
w_1,w_2	weight factors.

For a delay sensitive application, such as VoIP we assign a higher weight for w_2 to ensure a tree to bias the solution towards the case with the minimum hop count and, possibly, minimise the end-to-end delay.

6.9 Results

The prediction model used in this chapter is the same as that discussed in previous chapters. The results of the PSO-based Grey prediction model are given in chapter 4, Fig. 4.9 and they show a plot of the actual values of received signal strength and the corresponding predicted values. The PSO-based Grey model tracks the curve very accurately. The variations in the prediction values are shown in Fig. 4.10 by plotting the absolute error. Using the above results which provide accurate mobility prediction, tree selection will be minimised and this, in turn, reduces the use of network resources during handoff.

6.9.1 Numerical Results

The algorithms presented in the following section were implemented in the C++ language according to [134]. We performed the tests on a 10 node network and a 20 node network. We have compared the performance of the CAR-set algorithm against our proposed MMP algorithm. For testing our algorithm, we have considered the 20 node network which is a wired network of AR's as shown in the Fig. 6.4. We consider the selection of nodes from the prediction algorithm as being far better than the CAR-set as it can reduce the number of ARs and thus reduces the total bandwidth required for the multicast tree. Here, we considered the source node 1 as the corresponding node (e.g. video streaming server) and the remaining nodes could be the access routers sending information to the wireless network. We tested our

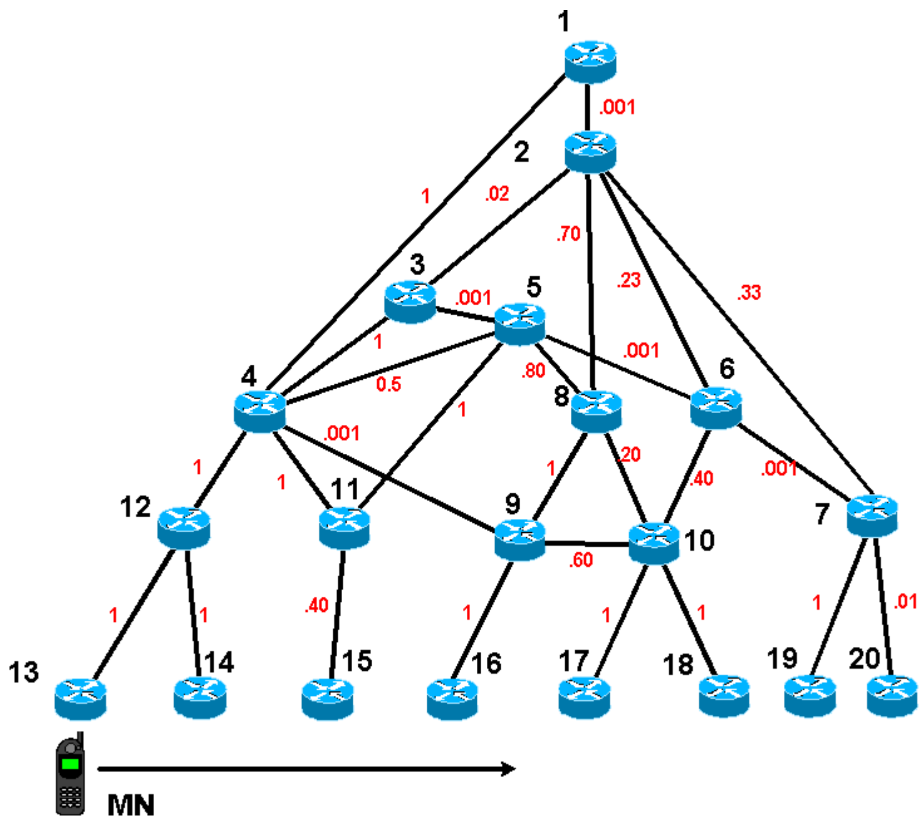

Figure 6.4: Topology for a 20 access router network

network and the settings as described using a Pentium 4 1.7 GHz PC with 512 MB RAM and the results obtained are summarised in Table 6.2 for the 10 node access router network of Fig. 6.3. For each test scenario, a network simulation experiment was setup based on the selection of nodes determined by our prediction algorithm.

10 Node AR Network proposed MMP algorithm				
	No. of selected nodes by prediction	Residual bandwidth	hop count	total cost
Tree 1	1	1.702	5	3.351
Tree 2	2	2.702	6	4.351
Tree 3	3	4.702	8	6.351
10 Node AR Network - K-Minhop algorithm				
	No. of selected nodes by prediction	Residual bandwidth	hop count	total cost
Tree 1	1	2	2	2
Tree 2	2	3	3	3
Tree 3	3	6	6	6

Table 6.2: Table showing the results from proposed algorithm and K-min hop algorithm for the 10 node access router network of Fig. 6.3.

10 Node AR Network - CAR-set algorithm				
	No. of selected nodes	Residual bandwidth	hop count	total cost
Tree 1	7	5.704	11	8.351

Table 6.3: Results from CAR-set algorithm for a 10 node access router network of Fig. 6.3.

For each experiment, we calculated the minimum cost tree for bandwidth and minimum hop count as shown. The table also shows the total cost as per equation 6.4 where $w_1 = w_2 = 0.5$. Each row in the table represents a set of tests performed for a given source and set of destinations. The bandwidth savings are shown in Tables 6.2 and 6.3. It can be seen that the results show very good performance of the algorithm proposed in terms of cost. In addition, we have compared (in Table 6.3) the model with the CAR-set algorithm which selects all the AR's irrespective of the mobile nodes' movement discussed by Helmy et. al. in [28]. It is worth noting that, in all cases, the total cost obtained by our algorithm is always less than the CAR-set algorithm. This suggests that it is unnecessary to reserve as many resources and not to flood the network with multicast packets. However, one disadvantage with this approach is the effect if our prediction algorithm fails. A possible reason for such a failure might be a Şblack spotŤ where there is no received signal strength. The reaction to this situation by the CAR-set algorithm could be better as more resources are available with that method. We believe that our prediction algorithm is accurate to within $\pm 0.02dB$ thus it is able to detect the signal strength as well as any other algorithm and matches any other proposed method to the present time. Table 6.4 and 6.5 shows the difference between the cost solutions obtained by the CAR-set algorithms and our algorithm for a 20 node access router network (Fig. 6.4) and it shows the various scenarios when more nodes are selected by our prediction algorithm.

20 Node AR Network - proposed MMP algorithm				
	No. of selected nodes by prediction	Residual bandwidth	hop count	total cost
Tree 1	1	2.702	6	4.351
Tree 2	2	4.102	8	6.051
Tree 3	3	5.102	9	7.051
20 Node AR Network - K-minhop algorithm				
	No. of selected nodes by prediction	Residual bandwidth	hop count	total cost
Tree 1	1	3	3	3
Tree 2	2	4.4	5	4.7
Tree 3	3	5.4	6	5.7

Table 6.4: Table showing the results from proposed MMP algorithm and K-minhop algorithm of Fig. 6.4.

20 Node AR Network - CAR-set algorithm				
	No. of selected nodes	Residual bandwidth	hop count	total cost
Tree 1	7	8.504	14	11.252

Table 6.5: Results from CAR-set algorithm for a 20 node access router network of Fig. 6.4.

6.10 Summary

In this chapter, we have provided an overview of how mobility prediction and multicasting can be combined together to help improve overall handoff performance. The methods of some current algorithms overload the network during handoff. Now with our approach it is possible to improve handoff in terms of bandwidth savings based on application requirements. Specifically, with this idea of mobility prediction and multicasting we can improve handoff delay. We have formulated a problem that takes into account a weighted cost involving bandwidth constraints and hop count supported by a prediction method that improves handoff performance in a multicast environment. Accordingly, two algorithms were proposed, the MMP algorithm and the K-MinHop algorithm respectively, and our results were tabulated and discussed in some detail.

Wireless multicast is required for a range of advanced wireless applications employing group communications among mobile users. Applying multicast methods to wireless networks is difficult for many reasons, for example, available bandwidth; the user's mobility can lead to the loss of packets, delays and incorrect routing. Multicasting packets to the set of candidate access routers can cause significant overheads by the duplication or replication of packets. Our ongoing and future work will address the above problems. Experiments are also to be conducted to test the performance in terms of handoff delay. In addition, our focus is on the development of good algorithms that enable us to optimise the two techniques jointly to improve handoff in wireless networks.

The design of mobility prediction based multicast algorithms for improving handoff performance has given an insight of how we can save on resources. An extensive and rich set of results has shown the performance critically depends on the mobility detection and handoff scheme that are employed in this research. In comparison with the CAR-set, our proposed algorithms provides a minimum cost route from the source to the destination. One of the major contributions of our study is the introduction of a framework using a mobility prediction scheme which provides an efficient proactive handoff procedure. In addition, we provided a thorough and systematic evaluation of our proposed algorithms via simulations and some numerical results.

In future work, we plan to use mobility prediction to measure the packet losses during handoff using the above algorithms. In particular, we would like to implement our algorithms on the network layer protocols for mobile networks namely MIP and HMIP which are under standardisation by the IETF. Finally, we recognise that the dependence of handoff performance on mobility prediction (mainly the detection scheme) deserves further emphasis on how the nodes join and leave the multicast tree. In the next chapter, we shall consider the development of more an accurate join and leave algorithm based on mobility prediction for multicasting during handoff depending on constraints and specific objectives.

Chapter 7

Join/leave Algorithms

7.1 Introduction

In chapter 6, we presented the framework/architecture for a mobile environment to improve handoff performance using mobility prediction and multicasting techniques. This was followed by the two multicasting algorithms namely, MMP algorithm and K-Minhop algorithm to construct the minimum cost multicast tree to suit the application running on the mobile node. For the problem formulation proposed in chapter 6, we gave weights to provide a bias towards one over the other. We also provided the numerical results for the 10 and 20 node networks to prove that resources are saved by using mobility prediction to predict the target access router.

In this chapter, we provide a detailed description of the algorithm required for the join and prune mechanisms, which will help to build an optimal multicast tree with QoS requirements during handoff and also reduce the number of joins and prunes. An analysis of how mobility prediction helps in the selection of potential Access Routers (AR) with QoS requirements is presented which affects the multicast group size and bandwidth cost of the multicast tree. The proposed technique tries to minimise the number of multicast tree join and prune operations. We have examined the performance of our algorithm using simulation studies under various environments and obtained good performance results. Our results show that the expected multicast group increases linearly in size with an increase in the number of selected destination AR's for multicast during handoff. We observe that the expected number of joins and prunes from the multicast tree increases with group size. Thus, for an increased number of destinations, the estimated cost of the multicast tree in a cellular network also increases. We also hope that the discussion presented here will help researchers and pave the way for future research.

Although many multicast algorithms are proposed for wired networks that consider end-to-end delay, bandwidth and inter-destination delay as constraints; none of the research papers considered these parameters for wireless networks. This chapter is dedicated to use parameters like bandwidth constraints and delay threshold on the link to establish a multicast tree during handoff – thus minimising the number of joins and prunes. Once again, we propose our hybrid mobility prediction algorithm based on the Grey model, in order to reduce the number of unnecessary join and prune operations.

The proposed algorithms also takes cares of the QoS constraints involved during the join operation. Furthermore, the solution proposed in this chapter reduces such overheads by performing accurate mobility prediction which is used to select a potential AR. The advance knowledge about the new target AR can be used to initiate a mechanism for avoiding packet losses and accomplishing a new handover (with low packet losses). To improve accuracy, a set of new potential AR's can be formed. Minimising the cost of a multicast tree is an important issue [24, 136]. When a mobile node wishes to join an existing multicast tree, a route from the existing multicast tree to the node must be computed. The Fig. 7.1 shows the basic diagram of our handover procedure when the mobile node moves from one point of attachment to the other. When the mobile node enters a new cell (under the AR) as predicted by our hybrid Grey prediction model, a route is established from the corresponding node to that AR. This is represented by a thick dark line in Fig. 7.1(a). The packets are delivered from the Corresponding Node(CN) to the AR via the intermediate ARs. As the mobile node moves from the current cell to the next potential cell, our multicasting algorithm tries to minimise the overall cost namely, the bandwidth cost and guarantees the delay thresholds are met. The path is established if and only if the objectives and constraints are met. When the mobile node moves, a route is established from the new AR to the CN (source) as shown in Fig. 7.1(b). As soon as the new connection is established the old route is pruned. Thus, by using mobility prediction algorithm we only select the potential AR which will serve the mobile node utilising a minimum of resources. A detailed explanation of the algorithm, and architecture are presented in the following section. This chapter deals with the problem of optimally connecting a node to an existing multicast tree such that the selected node by the prediction algorithm still satisfies our QoS requirements.

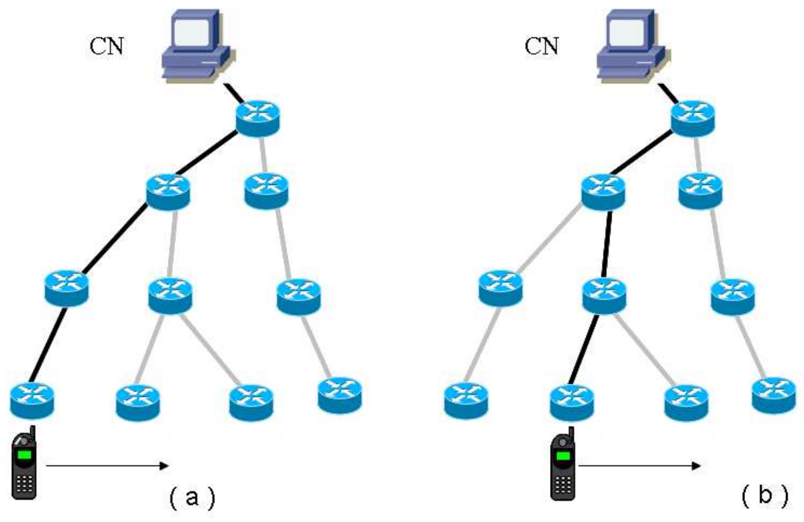

Figure 7.1: The basic join and prune mechanism as the mobile node moves from one AR to another.

We consider the formation of a near optimal multicast tree problem which requires predicting the new AR and proactively setting up a path to it. The main idea is to establish a multicast session from the source to these potential AR's to compute a minimum cost tree with specific constraints. We discuss the details of how mobility prediction can help multicast routing to improve handoff performance in terms of join/prune operations. An algorithm called MBWDC (Multicast BandWidth Delay Constraint) algorithm is proposed to solve this problem. The main motivation is to allow the necessary join operations and to reduce unnecessary joins. As stated, there could be some ARs that may perform unnecessary joins. The overhead of concern is the join and leave operations for the multicast tree. Obviously, the tree management cost will increase when mobility is higher. With this in mind, we provide information about the current state of the art in multicast protocols for wireless networks and compare them against several performance metrics.

The work in this chapter led to the publication [12]. The contribution of this chapter can be summarised as follows:

1. A situation was modelled to use mobility prediction to select the candidate access router for building the multicast tree based on residual bandwidth as an objective and the delay bound as a constraint.
2. Development of the MBWDC algorithm for building an optimal multicast tree based on the selection of candidate access routers, which incorporates the efficient join and leave procedures.
3. Simulation study is conducted to evaluate the accuracy of the proposed algorithm on different topologies which it employs for reducing the size of the multicast tree i.e., to reduce the total cost of the tree. It is also compared to the CAR set (Coverage Access Router) algorithm proposed by Helmy *et. al.*

The rest of the chapter is organised as follows: In section 7.2, we discuss related work on minimum cost multicast trees that were proposed for wired networks and give details of the problem formulation for our research. In section 7.3, we present our problem formulation. Section 7.4 describes the proposed MBWDC algorithm with a detailed description. Simulation parameters and a framework are presented in section 7.5. Finally, we provide some results of our study in section 7.6 and this is followed by conclusions and future work in section 7.7.

7.2 Related work

In cellular networks, communications between two mobile nodes completely rely on the wired backbone and fixed base stations or Access Routers (AR). Cellular networks have limited resources, which need to be conserved. In wireless networks, bandwidth is limited, wireless links are error prone and there is a frequent change in the position of the mobile which may initiate a handoff. For example, there has been a demand for online gaming operations (multiple), where players are located at different locations and use their PDA's or hand-held devices. In order to solve these handoff problems, we can, once again, perform mobility prediction [21, 56, 133]. As we have seen in previous chapters, mobility prediction can be used to highlight the minimum number of access routers required to build the multicast tree. The important methodology that we have used to underpin our methods throughout this work is Grey theory.

It should be also noted that there exists a large amount of literature on multicast communications in wired and wireless networks [140, 141]. Most of the multicast protocols are evaluated in terms of bandwidth consumed by the entire process and the maximum delay encountered in delivering a message to any member of the multicast group [135, 142, 143]. A multicast message can be sent to each member of the group separately, but this wastes bandwidth, as each message has to be sent over the same link several times [75, 78, 144]. There is a need to construct a minimal spanning tree with only the members of the group or the required number of AR's.

There are numerous protocols and algorithms based on multicasting that have been proposed for both wired and wireless networks. Some of them include M&M [72], MOM [70] and MMA [69] and RBMOM [81]. RBMOM (Range Based MObile Multicast) was proposed to support multicast for mobile hosts on the Internet [76, 81]. Here, there is a tradeoff between the shortest delivery path and the frequency of multicast tree re-configurations. It also shows that it can adapt to fluctuations in both host movement and the number of mobile group members. A detailed explanation of the protocols and algorithms such as M&M, MOM, MMA and RBMOM protocols which are proposed for a efficient handover performance in wireless networks were discussed in detail in chapter 2. However, in this section we shall discuss the algorithms which have been proposed for multicasting in wired networks so that we can relate them to our work and our problem formulation. Apart from the fact that we have applied multicasting algorithms that are proposed to minimise resources for a wireless environment, the basic structure which includes intermediate ARs of the network is still a wired network. Therefore, it is necessary to discuss algorithms that are proposed for wired networks and give a similar problem formulation for multicasting in wireless networks. Since we are looking at establishing a minimum cost tree (with objective and constraints) from the CN to the new AR, it can be modelled as a multicast problem. Some of the algorithms that are proposed to establish a minimum cost tree based on suitable objectives and constraints are described below. Based our literature survey, there also exists some multicast routing heuristics for the problem minimising the cost of the multicast tree. Some of the heuristics for this problem include the shortest-path-tree heuristic, which constructs a multicast tree from shortest paths from the source to every destination, and the minimum spanning tree heuristic, which computes a tree that spans the source and all destinations and minimises the total tree cost.

In the algorithm proposed by Kou et. al. [145], the KMB algorithm (Kou, Markowsky and Berman), a graph consisting of edges that represent the shortest paths between a source and a destination is represented as a network. The KMB algorithm computes the minimum spanning tree and the Steiner tree of the original network is obtained by calculating the shortest paths that are represented by edges in the minimum spanning tree.

Rouskas et. al. presented the DVBMT (Delay Variation Bound Multicast Tree) [136] algorithm for multicast routing with end-to-end delay and delay variation constraints. This algorithm studies the problem of constructing multicast trees to meet QoS requirements in real time. Here, the multicast communication depends on 1) bounded delay along the path from the source to each destination and 2) bounded variation among these delays along these paths. They derive heuristics that demonstrate good average case behaviour in terms of maximum inter-destination delay variation of the final tree. The constraints mentioned are the delay (Δ) and the delay variation tolerance (δ). The objective is to determine a multicast tree such that delays along all the source destination paths in the tree are within the two tolerances. The problem is defined as follows: ***Problem: (DVBMT)*** Given a network $G = (V, A)$, a source node $s \in V$, multicast group $M \subseteq V - s$, source to destination delay bound δ, a link delay function $\mathcal{D} : \mathcal{A} \longrightarrow \mathcal{R}^+$, a delay tolerance Δ, a delay variation δ, does there exist a tree $T = V_T, A_T$ spanning s and all the node M, such that:

$$\sum_{l \in P_t(s,v)} \mathcal{D}(\ell) \quad \leq \quad \Delta \qquad \forall \quad v \in M \tag{7.1}$$

$$\sum_{l \in P_t(s,v)} \mathcal{D}(\ell) - \sum_{l \in P_t(s,v)} \mathcal{D}(\ell) \mid \quad \leq \quad \delta \qquad \forall \quad u, v \in M \tag{7.2}$$

They refer to 7.1 as the source destination delay constraint and 7.2 as the inter-destination delay constraint, According to [136], a tree T is a feasible one for a give multicast session with a source s and a destination set M if and only if it satisfies both of these constraints.

Kompella et. al. presented a heuristic for the delay-constrained minimum-cost multicast routing problem [141]. The proposed heuristic depends on 1) bounded end-to-end delay along the path from the source to the destination 2) minimum cost of the multicast tree, where edge cost and edge delay can be independent constraints. In their proposed heuristic, the multicast tree is constructed with the constraint on individual path delay rather than trying to minimise the average path delay to all the destinations. The algorithm provides near-optimal routes that minimise cost and limit path delays.

Parsa et. al. present a delay-bounded minimum-cost multicast routing problem [140]. The proposed heuristic (BSMA) guarantees to find a feasible solution - if one exists. The heuristic first starts with a least-delay tree using any standard shortest path routing algorithm. If the least delay tree violates the end-to-end delay bound, then this bound is too tight. Otherwise, the heuristic manages to iteratively reduce the cost of the current multicast tree by replacing the most costly superedge on the current tree by a cheaper one while the delay bound is still satisfied. A superedge is a simple path containing only Steiner nodes excluding the two end nodes. The replacement of an existing superedge is by removing all inside nodes and edges on that superedge, which results in two disjoint subtrees, then a cheaper and feasible superedge is found to connect the two subtrees. Finding of a replacing superedge is done by a k-shortest path algorithm.

Ramanathan et.al.[142] formulated the problem of multicast tree generation as a directed Steiner tree of minimal cost. They present a polynomial time algorithm that provides a tradeoff solution between the tree cost and runtime efficiency. This work considered a more general model of directed graphs and allowed edge costs to be unequal in each direction. The assumption is that, given the heterogeneity of communication links e.g. radio, satellite etc., they believe that costs would be different in each direction. They also analyse the running time of the algorithm and show it to be $\mathcal{O}(m^2 + e)$ where m is the number of terminals, e the number of edges and k is a "knob value" which is used to set the algorithm to a desired tradeoff between the expected tree cost and running time. As k is increased, the tree cost decreases but the runtime increases. Overall, the contribution is a multicast tree generation procedure that accommodates link asymmetry.

Kodialam et. al. presented [135] a new algorithm for online routing of bandwidth guaranteed multicasts when routing requests arrive one by one without any *a priori* knowledge of future requests. Here, a multicasting routing request consists of a source, a receiver and a bandwidth requirement. This algorithm handles requests arriving one by one and satisfies as many potential future demands as possible. The algorithm develops a multicast tree selection heuristic that is based on the idea of deferred loading on certain links. The links are identified by the algorithm as critical links and uses link-state information and capacity information to build the multicast tree.

Similar to the proposed algorithms in the literature for multicasting in wired networks, we have formulated a problem for the wireless environment. Although, similar constraints are required for our problem, especially in the wired part which includes the ARs and a corresponding node, a lot depends on the wireless part for our architecture. This again depends on our mobility prediction algorithm which predicts the potential AR which is one of the destinations. After the destination is known, the problem becomes simpler, as we have to take care of constructing a minimum cost tree from the CN to the destination AR. In this chapter, we construct a multicast tree similar to that proposed for wired networks which includes specific constraints and objectives. Our idea is compared with the M&M protocol proposed by Helmy et. al. which also has similar techniques to highlight potential ARs for multicasting the packets from the corresponding node to the set of destination ARs. In multicast based mobility (M&M) [28] proposed by Helmy et. al. they define a set of protocol suites to enable multiple access routers to receive traffic for the mobile host. Here, a group called the CAR-set (Coverage Access Router set) of access routers is highlighted or selected to provide the necessary coverage for the mobile node so that there is no loss of packets during handoff. However, this method utilises more network resources than necessary and consequently, more access routers are selected for inclusion in the multicast tree. This, in turn, results in more unnecessary joins and prune operations. However, our proposed algorithm which involves a combined mobility prediction and multicasting algorithm outperforms the CAR-set algorithm when we simulate the algorithms using the same parameters and mobile environment. In the following sections, we describe the problem formulations and detailed descriptions of our algorithms and present some numerical results.

7.3 Problem Formulation

We represent a network as a connected directed graph $G = (V, E)$. V and E are the set of n nodes and m links of the network respectively, with a specific source node $s \in V$, and a set of destinations $D \subseteq V$, the objective is to find the minimum spanning tree rooted at s and spanning all the nodes in D. Fig. 7.2 shows a multicast tree in which CN is the correspondent node, the grey ones are the selected AR's by the prediction algorithm with thick darker lines showing the path taken and the other nodes are Steiner nodes (i.e., non-members in-tree nodes).

A path between a particular source v_s and a particular destination v_d is represented by a sequence of nodes $v_s, v_1, v_2, v_3, ...v_d$ where $v_i \subseteq V$. There could be many such paths based on a given source and destination. However, for multicast routing our focus is to find such paths between a single source and multiple destinations that

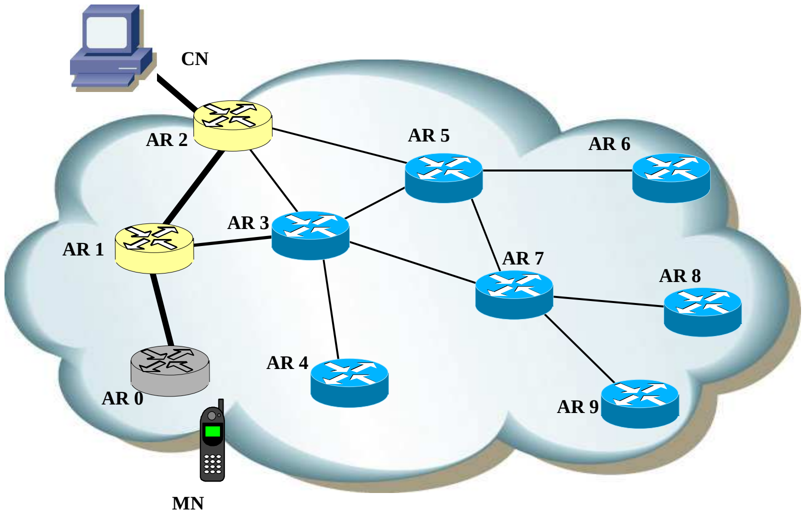

Figure 7.2: Topology of a 10 node network used in the simulation

will simultaneously satisfy our QoS requirements. The collection of multicast paths defines a multicast tree. An efficient allocation of network resources satisfying the QoS requirements is our primary goal. Several algorithms that construct low cost multicast routes are based on heuristics for generating approximate Steiner trees. However, satisfying the individual QoS parameters may be conflicting or may be inter-dependent making it a more challenging task. If we have a single optimisation factor to be satisfied say residual bandwidth constraint then it is easily solved. But, satisfying different QoS parameters simultaneously is an NP-Hard problem [24, 141]. In our formulation, the performance of the multicast tree is determined by two factors:

1. Bounded end-end delay along the individual paths from the source to destination.
2. Minimum cost of the multicast tree, for example, in terms of residual band-

width.

The first QoS optimisation factor chosen is the delay bound δ. In our formulation, a delay bound is specified while constructing the multicast tree. We assume the edge cost and edge delay are different functions. Here, edge cost is the inverse of the residual bandwidth and the edge delay could be the propagation delay or transmission delay or the queueing delay or the combination of all the three. Here, we are also trying to use the constrained minimum cost tree with constraints on the individual path delays. With the requirement of stringent delay constraint, it is often required that the delay from the source to any destination should be within the time bound or a threshold "δ". We can have, $P(v_s, v_d)$, $v_d \in D$, which is the path from source s to destination d in a multicast tree T, then the bounded delay is expressed as

$$\forall v_d \in D : \qquad \sum_{l \in P(v_s, v_d)} d(l) < \delta \tag{7.3}$$

where $d(l)$ is used to indicate the delay of the link.

The second QoS optimisation factor chosen is the residual bandwidth. Generally, the multicast path capable of providing the greatest residual bandwidth is taken as the best choice. The total cost of residual bandwidth in the network is given by $\sum_{l \in E}(c_l(b_l))$, where c_l is the cost of the link $l \in E$ and b_l is the bandwidth allocated for the different hops along the entire multicast tree T. We notice that $b_l = 0$ if $l \notin p$, where $p \in T$.

7.3.1 Problem Specification

The multicast routing problem with bandwidth and delay constraints can be formally stated as follows:

Problem (MBWDC): Given a network $G = (V, E)$,$\{(c_l = 1/b_l, d_l)\}_{l \in E}$, a source node $s \in V$, multicast group $M \subseteq V - s$, source to destination delay bound δ, p is the multicast path, find an in-tree node $t \in T$ and a tree T rooted at s such that it minimises the cost $c(T)$ and $\sum_{l \in p}(d_l) < \delta$. $c(T)$ is defined as $\sum_{l \in p}(c_l)$.

In our formulation, the network is represented by a graph $G = (V, E)$, where V denotes the set of nodes and E, a set of edges that corresponds to the set of communication links connecting the various nodes. In our scenario, packets originating from the source node $s \in V$ have to be delivered to the destination(s). We call this set $M \subseteq V - s$ of destination nodes, the *destination set* or *multicast group* and we shall use $m =\mid M \mid$ to denote its size. The above routing problem involves finding the optimal multicast tree when the source and all destination nodes are specified. The tree is established if and only if the delay constraints are met first after which the bandwidth constraint is applied to get a feasible path from the source to the destination. In the following section, we take a closer look at the problem of determining the multicast tree that guarantees the desired level of performance in terms of reserving the resources with a minimum number of joins and prunes during handoff.

7.4 Join and Leave Algorithm

7.4.1 Host Mobility and Tree Joining Decisions

In our approach, determining the location and speed of the mobile node is very important. The best approach is to use the signal strengths available from the access points. The mobile host requests a tree join process when the received signal strength of the AP is above the required level and each of the constraints is satisfied. If the required QoS requirements are met, then the join operation will start with the information received from the mobile host. The decision rule defined above is very simple; when the predicted value of the signal strength and the various QoS constraints are met, a mobile host requests a join operation. As a part of the connection establishment process, a multicast tree satisfying constraints (factors) 1 and 2 needs to be determined. Our algorithm operates under the assumption that there exists a source node s, and the node to be joined to the tree (in order to execute the handover) should meet the required QoS constraints. The construction of the initial tree (say T) is based on the selected destination AR's using the mobility prediction algorithm. As a first step, the shortest path from the source s to a destination node is noted. If T does not satisfy the path with a delay constraint, no tree may satisfy it, implying that the delay tolerance is too tight. At this point, it may be necessary to repeat the procedure with the next candidate AR (second best AR). So, a negotiation may be necessary to determine a less stringent value of the delay bound or select the path that best satisfies the required QoS. Suppose now that the negotiated value of the delay bound is met for the tree T, it also has to meet bandwidth requirements. If the two requirements are simultaneously satisfied, the tree T is considered to have a feasible path for the particular AR and the join operation is completed. It is also possible that the multicast tree may fail to satisfy condition 2. In our approach, we construct the best possible tree with the selected AR using a suitable search algorithm so that it finds the optimal tree and makes a join operation. After the join operation is successful for the mobile node, the handover is completed. The following subsection explains the step by step process to implement the algorithm as shown in Fig. 7.3.

7.4.2 Proposed MBWDC Algorithm and Description

During the handover session, nodes may join/prune from the initial tree which is constructed by using knowledge of the destination nodes given by the mobility prediction algorithm. It is necessary to dynamically update the multicast tree based

Input : $G = (V, E)$ = graph, s = source node
D = set of destination nodes
N = Number of destination nodes
δ = destination delay bounds threshold
Output: A delay bounded route and a destination node satisfying constraints and objectives

MBWDC-JOIN($\{G, \delta_l, c_l(b_l)\}_{l \in E}$)

```
T ← minimum spanning tree with D ∪ s ;
Initialise all edges of T as unmarked.;
for v ∈ D do
    P_v ← shortest path from s to v ;
    Sort P_v in the increasing order delay and label them as p_1, p_2, ··· p_k ;
    for p ∈ P_v do
        if d(p) > δ then
            delete p from P_v ;
        end
        /* We have a set of paths for a particular destination*/ ;
        return OK ;
        else return FAIL ;
    end
end
/* we have paths from s to multiple destinations */
for p_i = 1 to K do
    Construct a tree T including all the destination nodes D and all links.;
end
p_best = φ /*route from s to a destination*/;
c(p_best) = ∞ ;
for v ∈ D do
    for p ∈ P_v do
        c(p) = Σ_{l∈p} c(b_l) ;
        if c(p) < c(p_best) then
            c(p_best) = c(p) ;
            p_best = p ;
        end
    end
end
/*Prune operation*/
for l ∈ p_best in T do
    Mark link l and corresponding tail and head ;
    Remove all the remaining links and nodes.;
end
```

Figure 7.3: Proposed algorithm for MBWDC-JOIN

on the movement of the mobile host and ensure that the delay constraints and the bandwidth cost are satisfied at all times.

Let D be the destination nodes selected by the mobility prediction algorithm. Fig. 7.3 outlines the proposed MBWDC algorithm as shown. First, all the nodes and edges in the network are initialised and are defined as being *unmarked*. In order to find paths, we construct the l-shortest paths P_i from the node v to s. In lines 3-14, for each of the nodes in D, a minimum delay bounded path P_v is determined for each node. The path from a node $v \in D$ to a source node s is found. It is possible that the most suitable candidate AR with good signal strength may not be the best destination AR node as it may not satisfy the delay bounds. For this reason, all the paths are placed in ascending order (see line 5). The selection process is based on the acceptable signal strength and required delay bounds. The initial prune operation is made to eliminate all the nodes which do not satisfy the delay constraints (lines 7-8). In lines 16-18, we construct the multicast tree T based on the paths obtained. In lines 19-28, for a particular selected node, the bandwidth costs are evaluated for each path from v to the source. The best route and least cost for bandwidth (p_{best}) is taken as the best AR for handover, pruning all the other nodes.

To complete the description of the algorithm, note that if a feasible tree exists, it will contain some path from v to s. Therefore, if the process of path selection does not satisfy the delay bounds initially, there always exists a second path which can be used. Finally, if the algorithm terminates at line 34, the path returned is a feasible one that exists.

7.5 Simulation Modelling and Framework

7.5.1 Simulation Model

For this example, we have selected two base stations A and B, that are separated by D metres. The mobile device moves from one cell to another with a constant velocity and the received signal strength is sampled at a constant distance d_s in metres. The model we are considering includes slow fading [94]. The received signal strengths a_t and b_t $(in\ dB)$ when the mobile is at a given distance are given by

$$a_t = K_1 - K_2 \log kd_s + u_t \tag{7.4}$$

$$b_t = K_1 - K_2 \log (N - k)\, d_s + u_t \tag{7.5}$$

where $N = D/d_s$. The parameters $K_1 = 0$ and $K_2 = 30$ in dB are typical of an urban environment accounting for path loss. The simulation parameters used for the movement detection are the same as shown previously in chapter 6, Table 6.1.

7.5.2 Simulation Parameters and Topology

Experiments were conducted over several sets of topologies, taking into account various factors that can affect the performance of handoff. In the model, we have selected the base stations or access points under the AR to be separated by D metres. The mobile node is allowed to move from one cell to another with a constant velocity and the received signal strengths are sampled at a constant distance d_s in metres. The above parameters were used to select the potential AR with the help of the PSO based hybrid prediction model. However to evaluate our multicasting algorithms we chose different topologies to observe the effects on performance. The simulation setup included topologies with varying depths. In other words, the distance between the source node and the destination node was varied for each topology.

The network types considered are shown in Fig. 7.2, Fig. 7.4 and Fig. 7.5. A number of access points (AP) can be connected to the access routers (AR). When a mobile node moves from one AP to the other without changing the AR, it is called an intra-AR handoff and when it changes from one AR to another it is called an inter-AR handoff. An access point that is connected to the access router serves a mobile node. The access point acts as the radio point of contact to the mobile node. An AR considers that each AP is on a separate subnet [28]. Most studies conducted on mobility use different topologies and scenarios to evaluate their architecture and focus on handover behaviour. In our model, we have defined network topologies with sizes 10, 20, 60 and 100 nodes to test our algorithm. The topologies define the number of nodes and their link connectivity with associated delays and bandwidth costs. A sample of networks with sizes of 60 and 100 are shown in figures 7.4 and 7.5 respectively.

The topologies for the 10 node and 20 node network were similar to the ones that were used in chapter 6. As in the previous chapter, we assume that the packets originate at the corresponding node and are destined to reach the potential AR via intermediate ARs. Our main aim was to minimise the number of joins and prunes and use the resources efficiently - based on our various constraints and objectives. The subsequent sections discuss the results obtained using our MBWDC algorithm and we compare them with the CAR-set algorithm.

7.6 Results

The prediction model used in this chapter is the same that has been discussed in each of the previous chapters 3, 4 and 5. The results of the PSO based Grey prediction model are given in chapter 4, Fig. 4.9 and they show a plot of the actual values of

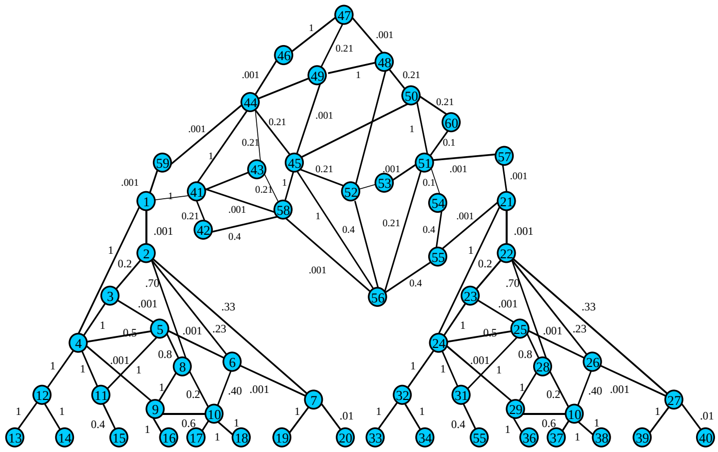

Figure 7.4: Topology of 60 node network used in the simulation

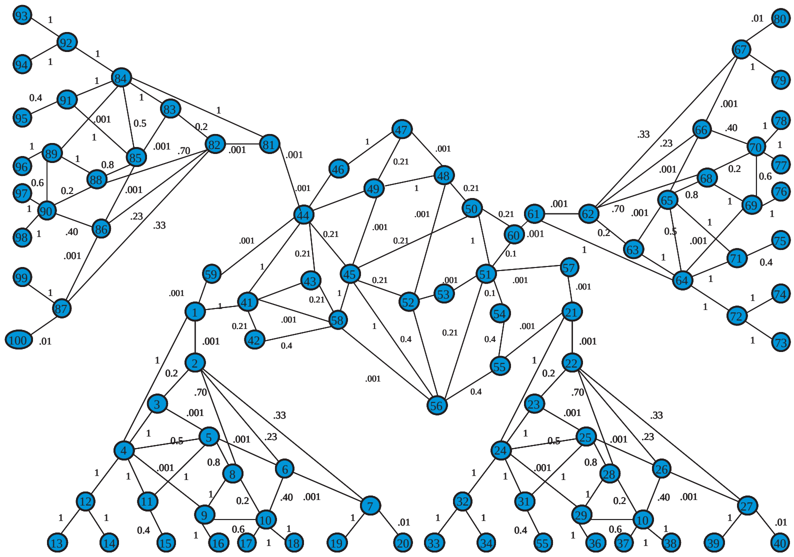

Figure 7.5: Topology of 100 node network used in the simulation

received signal strength and the corresponding predicted values. The PSO based Grey model tracks the curve very accurately. The variations in the prediction values are shown in Fig. 4.10 by plotting the absolute error. Using the above results which provide accurate mobility prediction, tree selection will be minimised and this in turn reduces the use of network resources during handoff. The algorithm

10 Node Access Router Network - proposed MBWDC algorithm					
	Selected by prediction	No of Paths	Residual bandwidth	hop count	Unnecessary joins
MBWDC-1	1	7	1.202	5	2
MBWDC-2	2	17	2.202	6	3
MBWDC-3	3	24	3.202	7	4
10 Node Access Router Network CAR-set Algorithm					
	Selected by prediction	No of Paths	Residual bandwidth	hop count	Unnecessary joins
CAR-set	7	52	5.202	8	6

Table 7.1: Table showing the results from proposed algorithm and CAR-set algorithm for 10 node access router network

20 Node Access Router Network - proposed MBWDC algorithm					
	Selected by prediction	No of Paths	Residual bandwidth	hop count	Unnecessary joins
MBWDC-1	1	5	2.202	6	3
MBWDC-2	2	20	3.202	7	4
MBWDC-3	3	30	4.602	9	5
20 Node Access Router Network CAR-set Algorithm					
	Selected by prediction	No of Paths	Residual bandwidth	hop count	Unnecessary joins
CAR-set	7	60	5.603	11	6

Table 7.2: Table showing the results from proposed algorithm and CAR-set algorithm for a 20 node access router network

presented in the previous section was implemented in C++ according to [134, 138]. We performed our tests on 10, 20, 60 and 100 node networks. We have compared the performance of the CAR-set algorithm against our proposed MBWDC algorithm. For testing purposes, we have considered networks which have a wired network of AR's. The selection of nodes from the prediction algorithm is far better than the CAR-set as it can reduce the number of AR's and thus reduce the total bandwidth required for the multicast. Here, we considered a source node as the corresponding

60 Node Access Router Network - proposed MBWDC algorithm					
	Selected by prediction	No of Paths	Residual bandwidth	hop count	Unnecessary joins
MBWDC-1	1	5	5.743	12	2
MBWDC-2	2	15	6.743	13	3
MBWDC-3	3	25	8.143	13	4
60 Node Access Router Network CAR Set Algorithm					
	Selected by prediction	No of Paths	Residual bandwidth	hop count	Unnecessary joins
CAR-set	7	65	9.144	17	7

Table 7.3: Table showing the results from proposed algorithm and CAR-set algorithm for a 60 node access router network

100 Node Access Router Network - proposed MBWDC algorithm					
	Selected by prediction	No of Paths	Residual bandwidth	hop count	Unnecessary joins
MBWDC-1	1	5	1.254	12	3
MBWDC-2	2	18	3.567	14	5
MBWDC-3	3	28	7.476	15	6
100 Node Access Router Network CAR Set Algorithm					
	Selected by prediction	No of Paths	Residual bandwidth	hop count	Unnecessary joins
CAR-set	7	70	9.076	18	8

Table 7.4: Table showing the results from proposed algorithm and CAR-set algorithm for 100 node access router network

node (e.g. video streaming server) and the remaining nodes could be access routers sending information to the wireless network. We tested our network and the settings as described on a Pentium 4 1.7 GHz PC with 512 MB RAM and the results obtained are summarised in Table 7.1 for the 10 node access router network, Table 7.2 for the 20 node access router network, Table 7.3 for the 60 node access router network, and Table 7.4 for the 100 node access router network. For each test scenario, a network simulation experiment was set up based on the selection of nodes determined by our prediction algorithm. In our simulation, the δ value and bandwidth cost are assigned

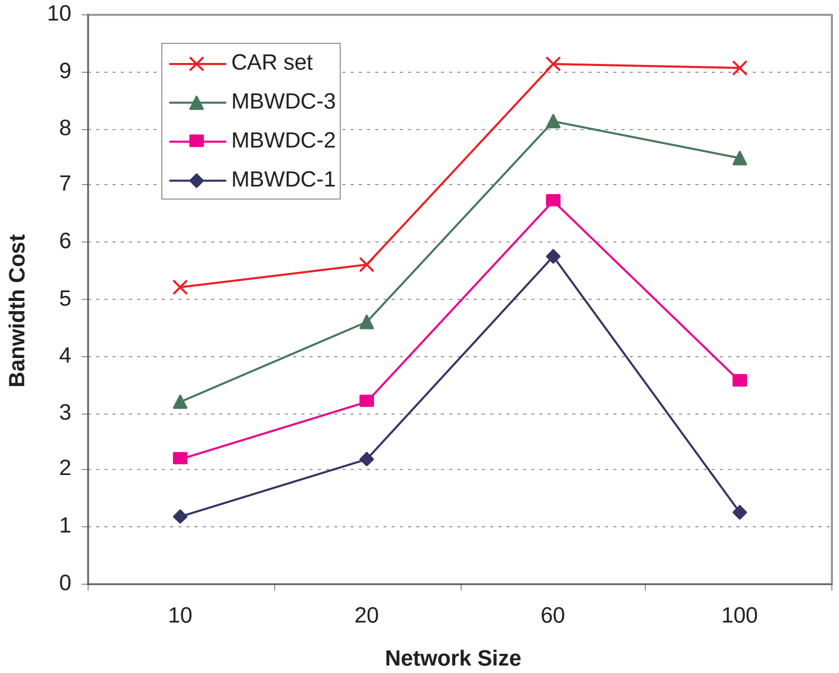

Figure 7.6: Bandwidth cost Vs. Network size.

initially and remain the same for all destinations. The delays on the individual links are generated randomly between 0 and 1. For simplicity, all the links are assumed to be bidirectional and symmetric. Furthermore, all the links are assumed to have enough bandwidth to satisfy the bandwidth constraints.

For each experiment, we performed and calculated the minimum cost tree for bandwidth. In the tables and the plotted graphs, MBWDC-1 represents a single node selection, MBWDC-2 represents a 2-node selection and so on. The tables also show the total cost, hop count, the number of unnecessary joins and the number of paths generated for a given source and destination for different network sizes. Each row in the table represents a set of tests performed for a given source and specified set of destinations. It can be seen that the results show very good performance of our algorithm in terms of cost. In addition, we have compared the model with the CAR-

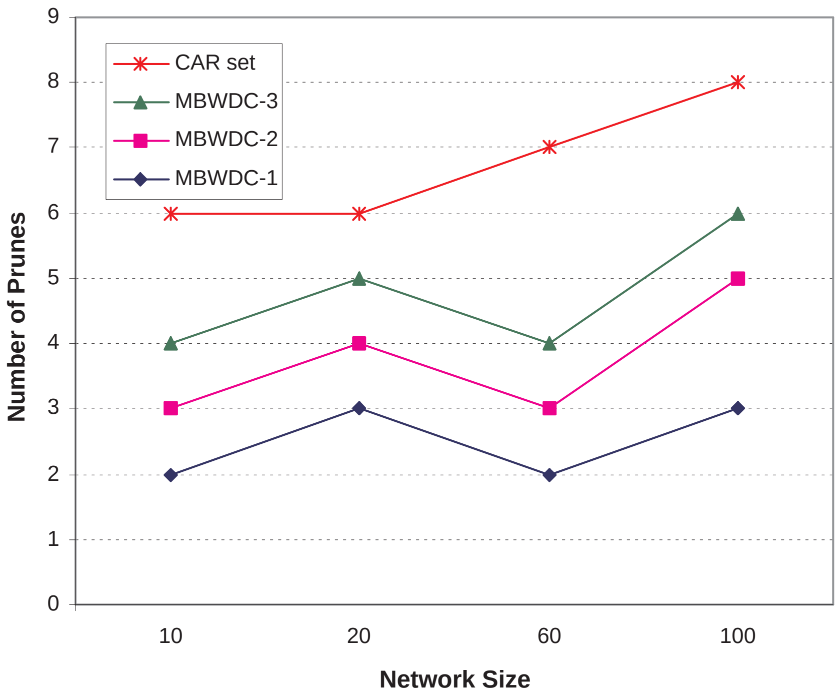

Figure 7.7: Number of prunes Vs. Network size.

set algorithm which selects all the AR's, irrespective of the mobile nodes' movement, as discussed in [28]. It is worth noting that in all cases, the total cost obtained by our algorithm is always less than the CAR-set algorithm. This suggests that it is unnecessary to reserve resources and not to flood the network with multicast packets. However, one disadvantage with this approach is if our prediction algorithm fails. A possible reason for such a failure might be a "black spot" where there is no received signal strength. The reaction to this situation by the CAR-set algorithm may be better, as more resources are available using that method. As previously discussed in chapter 6, we believe that our prediction algorithm is accurate to within $\pm 0.02dB$ and thus it is able to detect the signal strength as well as any other algorithm and matches any other proposed method to the present time. All the tables show the various scenarios when more nodes are selected by our prediction algorithm.

We have also plotted the results in terms of bandwidth cost and the number of prunes against network size. In Fig. 7.6, the bandwidth cost is plotted against network size. With the increase in network size, the cost also increases. This is because, with the increase in network size, the tree becomes more dense resulting in more nodes in the paths. If the number of destinations selected by the mobility prediction algorithm increases there would be more nodes – which is the major factor for our results. We have compared our results in both the graphs against the CAR-set algorithm proposed by [28]. Our results show better performance in terms of bandwidths costs with delay constraints than the CAR-set algorithm. In Fig. 7.7, the number of prunes is plotted against group size. Again, in comparison to the CAR-set algorithm, our algorithm performs better as it selects the most suitable node for prediction to perform the join operation. We argue that the basic reason for this improvement is the selection parameter for the number of nodes highlighted for handover.

Overall, the effect of mobility prediction on the performance of our algorithm was excellent. Depending on how accurately we can predict the next potential AR before the mobile node moves, the exact resources necessary to compute the route had an effect on handoff performance. In our analysis, the advantage of having a mobility prediction algorithm improved the resource consumption 8-fold when compared to the CAR-set algorithm. In other words, the use of mobility prediction resulted in a significant decrease in the required resources as well as reducing the number of joins and prunes and minimising the routing. The results also show the effects of network topology are equally important. This also depends on the distance between the physically adjacent ARs as the resources consumed may be less due to the smaller number of hops. The results also suggest that there is an advantage in placing the adjacent AR's logically close to each other in the wired topology to support mobility. In conclusion, the algorithms proposed effectively reduce the overall consumption of resources in the network.

7.7 Summary

In this chapter, we have provided an overview of how a combination of mobility prediction and multicasting helps to improve handoff performance. A model is formulated by taking into account residual bandwidth as an objective, together with minimum delay requirements as a constraint with the help of mobility prediction that improves handoff performance in a multicast environment. The source-destination delay constraint has been considered previously in the context of designing Steiner trees for real-time multimedia applications but we are not aware of any work that explicitly considers mobility prediction, bandwidth requirements and delay constraints as parameters to select the optimal tree when applied to wireless networks. By providing the values for parameters of cost and δ, we impose a set of constraints on the paths of the multicast tree. Handoff will occur if and only if the tree satisfying these constraints can be found; otherwise the operation will abort. Furthermore, the extra delay incurred through rebuilding a multicast tree can create the possibility of disruption in data delivery.

This chapter proposed an algorithm based on complete topology information for the construction of the delay bounded minimum cost tree. The contribution of the work lies in the novelty of using our mobility prediction algorithm for selection of appropriate AR's and building the multicast tree to improve the performance of handoff. Our algorithm minimises the total link cost of the tree while satisfying delay constraints which could be used for different applications in a wireless environment. The fundamental difference between the CAR-set algorithm and our prediction methodology is a set of access routers are selected to receive the packets destined to the mobile node. This chapter presents an analysis of how mobility prediction helps in the selection of potential AR's with QoS requirements which directly affects the multicast group size and cost of the multicast tree.

By exploiting the idea of mobility prediction, we not only predict the future state of network topology and perform route reconstruction proactively but also reduce the amount of network resources that are used. Moreover, by using the predicted information on the network topology, we can eliminate extra joins and prunes that would be otherwise needed to reconstruct the route and thus reduce overheads. In this chapter, we proposed various schemes to improve routing performance by using mobility prediction. We then evaluated the effectiveness of using mobility prediction via simulation. Effective delivery of data packets while minimising connection disruption is crucial in a wireless environment. In this chapter, we observed that by the use of mobility prediction we can anticipate topology changes and perform rerouting prior to route breaks. Our future idea would be to apply the mobility prediction mechanism to some of the most popular protocols proposed for improving the handoff performance. The results achieved in our research are very encouraging and open the way to further research in several directions.

Our future work mainly lies in the application of the algorithm to existing protocols such as MIP and HMIP and see if we could improve the performance in terms of handoff delay and packet losses. Although many papers discuss packet losses during handoff, we are yet to study the performance of our algorithms in terms of packet losses during handoff. Future work also lies in applying mobility prediction to heterogeneous technologies or 4G next generation networks.

Chapter 8

A Handoff Simulation Tool

8.1 Introduction

In this chapter we shall present the design and implementation of the NeTSim-v3.0 network simulation tool, which incorporates the simulation models considered in previous chapters. We shall describe some on-going work on the homogeneous and heterogeneous handoff. The aim of the simulation tool was not only to develop a model for our research problem but also to provide a platform for future investigations involving both wired and wireless networks. The tool has been specially tailored to deal with issues of mobility management – especially with regard to handoff detection schemes (at the link layer). Our research has mainly focussed on the layer-2 detection schemes which provide the advantage of selection of potential access routers/base stations to improve handoff latency. The tool that has been developed can be used to simulate mobile nodes and access routers together with specific scenarios for studying current protocols and algorithms. We note that the simulation tool incorporates only layer-2 detection mechanisms and future work is focussed on implementing packet losses during handoff for both homogeneous and heterogeneous networks using MIPv6 as the layer-3 protocol. The work described in this chapter details the software design and implementation of the simulation model and tool that has been developed.

There have been many different simulators proposed in the literature that are currently being used by researchers to study the performance of mobile and fixed communication networks. Among the better known examples are: OPNET [146], OMnet++ [147], NS-2 [148, 149], NCTUns1.0 [150], in particular. Although many of these efficient simulation tools exist, the motivation to build our own simulation tool was to enable us to focus on the specific requirements of our research. Moreover, our model incorporates only the necessary assumptions for inclusion of the required parameters/functions for our research. Furthermore, some of the existing simulation tools can be difficult to understand and have more complexity than necessary for our studies. The advantage in developing our own simulation tool is that it provides the freedom to understand, configure protocol modules, draw desired topologies, specify the movement paths of the mobile nodes and to plot the necessary performance graphs.

To date, many different wireless systems exist, which range from indoor to outdoor cellular systems. Features such as higher bandwidth and low cost make wireless LAN (WLAN) systems the competitive choice for high speed wireless Internet access. The integration of WLAN and cellular networks has recently been a subject of great interest. This has lead to what is known as fourth generation networks (4G) that envision the convergence of different access technologies to provide communication “anywhere, anytime". This convergence is aimed at providing seamless connectivity with low-cost when the user is using a hotspot (i.e., WiFi networks with high data rates) and the cost increases as the user moves into a cellular network such as UMTS which provides a lower data rate. For this seamless connectivity to take place, the handover procedure also becomes a bit more complicated. This special type of handover leads to what is referred to as “vertical handoff" when the user

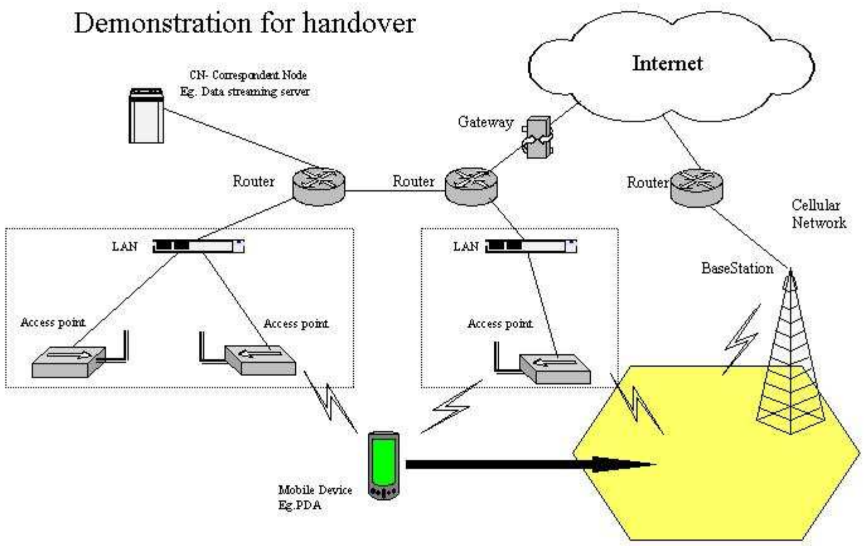

Figure 8.1: A sample network chosen to demonstrate homogeneous and heterogenous network architecture

moves between different networks or heterogeneous networks. So far, researchers have suggested numerous approaches to solving horizontal handoff problems, i.e., the movement between *similar networks*. A typical example is shown in the Fig. 8.1 to demonstrate both a vertical and a horizontal handoff procedure. Problems involving vertical handover are more complex as they are determined by data rates, delays, and the bandwidths of dissimilar networks. It has also been shown in our research that better detection techniques can improve the performance of such networks very significantly. Hence, in this chapter, we shall discuss the simulation tool for such a scenario (a software oriented approach) detailing the graphical user interfaces, simulation engine and data storage developed for our research project.

One of the issues which is of major concern in wireless networks is the data loss that occurs during handoff. We know that the network layer protocol plays an important role in the handoff procedure and that the IP protocol is the universally accepted network layer protocol for wireless technologies as in wired technologies. There has been a lot of research in developing a suitable network layer protocol to help with handoff in mobile networks and some of them include MIP, MIPv6 and HMIPv6 proposed by the IETF working groups some of which are currently being standardised. There is also some effort being directed into using non-prediction techniques as proposed by the IETF using the MIPv6 protocol which have been described in the literature [151]. However, for the purposes described in this chapter we have used MIPv6 with prediction techniques for layer-2 detection to observe the performance of handoff. Although the basics of the MIPv6 protocol were implemented in our tool, we concentrated only on link layer detection mechanisms and mobility prediction for the protocol.

The work in this chapter led to the publications [7] and a research grant [8]. The contribution of this chapter can be summarised as follows:

1. Based on the mobility prediction model, a simulator was developed as a part of the research grant (RPC) achieved from Microsoft Corporation. A model was developed to demonstrate homogeneous and heterogeneous handoff. Vertical handover was demonstrated using the simulator which was build using Microsoft Visual Studio .NET with C# as the programming language.
2. The model consisted of a discrete event simulation for vertical handover from a wireless LAN network to a cellular network and a horizontal handover from one WLAN to another. The protocol specification used the Mobile IPv6 as the standard protocol.
3. The developed tool consisted of a standard Windows GUI and XML to store the simulation results.

The rest of the chapter is organised as follows, section 8.2 discusses the related work on alternative simulators developed by researchers. Section 8.3 discusses the handoff architecture and the scenario considered for our problem. The section 8.4 discusses mainly about the vertical handover in heterogeneous networks. The section 8.5 presents the design and implementation of our simulation tool with respect to our software approach and presents a discussion of the different software blocks in our tool. Section 8.6 presents our simulation modelling and the assumptions made with respect to the homogeneous and heterogeneous handoff scenario and the prediction algorithm for the handoff decision. Finally, we provide some results of using the simulation tool in section 8.7 and this is followed by some conclusions in section 8.9.

8.2 Related Work

This section briefly presents some related work on existing simulation tools that have been proposed in the literature. Details of the simulators that are discussed can be found in references [146, 147, 148, 149, 150] respectively. Future wireless networks should include capabilities such as the three basic areas of connectivity: personal area networking such as Bluetooth, local high speed access points on the network including wireless LAN technologies such as IEEE 802.11, HIPERLAN and cellular connectivity.

Most of the protocols and their variants are supported by the various simulation tools available such as OPNET, OMNET, NS2, NCTUns1.0. NS2 is an open source tool and has been used by many researchers, agencies and organisations. The main drawbacks of NS2 are the lack of a GUI (Graphical User Interface), extensive CPU usage and large memory requirements.

OPNET is a commercial tool by MIL3 Inc., (OPNET = OPtimised Network Engineering Tool) capable of simulating large networks for both wired networks and wireless networks. The features include detailed protocol analysis including different editors such as node editors, process editors, event scheduling and simulation kernel.

The tool also includes the definition of topology and nodes with associated links. The main drawback is that it is an expensive commercial tool with a very steep learning curve.

The other tool is OMNET++ which is a freely available discrete event simulation tool that is designed to simulate computer networks and distributed systems. The main components include a topology description language called the NED language, a command line user interface and a graphical output vector plotting tool. It also includes contributed code from researchers and is similar to NS2. Although it can be easily understood, there are some drawbacks with compilation and installation of the tool in some environments. The above-mentioned simulators work on the Unix/Linux platform and can be quite tedious to work with in some cases. In our research due to the complication and drawbacks of these existing tools, we developed our own tool suitable for our research requirements and assumptions that were made for our specific scenarios to demonstrate both homogeneous and heterogeneous handoff. In the following sections, we discuss the framework/architecture considered for the simulation which is followed by the description of the software architecture.

8.3 Framework/Network Architecture

The network model that we consider is a wireline/wireless network with a number of access routers connected together and is shown in the Fig. 8.1. We consider a fairly large network with many access points, where the Mobile Node [MN] can cross several access points [AP]. As shown in the Fig. 8.1, the Corresponding Node [CN] wishing to send information to the mobile node must send their packets via these access routers. A number of access points [AP] can be connected to the access routers [AR]. Each AP covers a region called a *cell area*. When a mobile node moves from one AP to the other without changing their AR, then this is called a *homogeneous network handoff* and when it changes from one AR to another AR within a different network, it is called a *heterogeneous network handoff*. An access point that is connected to the access router serves a mobile node. A mobile node, throughout its movement joins and leaves these access points. The access point acts as the radio point of contact with the mobile node. An AR considers that each AP is on a separate subnet.

In our framework, we have a number of access points connected to two different subnets connected via the internet to a cellular network which is an overlay network. These are connected via routers and gateways as shown in Figures 8.1 and 8.4. During the simulation, the mobile node is allowed to move in a (prescribed) straight line for simplicity. The aim of such a scenario was to demonstrate both the homogeneous and heterogeneous handoff procedures. The protocol which was used in our scenario was the Mobile IPv6 protocol. However, not every aspect of the protocol details were actually implemented. Our hybrid model was used for the layer-2 detection scheme (using RSSI values) and was shown to predict the next potential access point very accurately.

8.4 Vertical Handover

A vertical handover occurs between base stations that are using different access and network technologies. Proliferation of WLANs has led to low costs and higher data rates for users. The next generation of wireless networks (4G) is aimed at keeping the user connected irrespective of the networks involved in that connection. The future of the next generation wireless internet is expected to consist of different types of wireless networks providing different bandwidths and coverage areas. For this connectivity to be possible, overlay networks provide a good solution with certain limitations such as low data rates. For example, consider a user who is connected to a WLAN/hotspot in a building and wants to move to a nearby building. Once the user moves out of range of the WLAN, the connection should be handed over to the overlay network without loss of connectivity. This handover from the WLAN to the overlay network (such as a cellular network) is called a vertical handover. The WLAN handover is initiated by the mobile node, i.e, the MN decides, according to the signal strength, when it has to perform a handover. A vertical handover may occur between a WLAN and the UMTS network where the UMTS network may be a large overlay network with lower bandwidths [152].

It is also important to note that vertical handover can be performed anywhere within the network, not just at the boundaries of the cell area because different wireless networks do not necessarily overlay with each other only at the boundaries of a cell. Some of the factors that determine handoff in heterogeneous technologies are mobility, bandwidth availability and latency [153, 154]. Although many protocols and algorithms are proposed which address these factors, there is still some room for improvement – mainly in the detection mechanisms at layer-2 to support the upper layer protocols. The following sections discuss the software design of the simulation tool for layer-2 detection mechanisms to provide improvements in performance during handoff.

8.5 Design and Implementation

The architecture of the software was the first consideration for the simulation tool. The application was coded in C#, because of its efficient handling of XML, speed, and the good quality user interface that it produces. The tool has several different module components and features that are discussed in the sections below.

8.5.1 Design Surface

The design surface was developed in VB.NET and used a component from MSDN. There were certain modifications done to this component to meet the customised requirements of the current simulation. The following were requirements of the design surface

- Support multiple graphical elements such as text, pictures, and shapes.
- Allow movement of elements. This was very useful especially to simulate the mobility of the mobile node.
- Support drawing to the printer as well as to the screen.
- Allow finished compositions to be saved to disk and load back into the design surface.
- Visually represent printer page boundaries on the screen to make it easier to position elements relative to the printed page.

Binary Serialisation of the Topology and XML Serialisation of the Results

An important requirement after drawing the topology onto the design surface was to save the state of the scenario so that it can be reused. For this purpose, the serialisation features in .NET were employed. Serialisation can be defined as the process of storing the state of an object to a storage medium. During this process, the public and private fields of the object and the name of the class, including the assembly containing the class, are converted to a stream of bytes, which are then written to a data stream. When the object is subsequently deserialised, an exact clone of the original object is created. Initial XML serialisation was employed but due to speed the design surface, it was modified so that every object could be serialised using binary serialisation. Also the objects can be deserialised and the simulation can be loaded again. We have used XML serialisation because these results can be exported for any application. It is an XML file which can be loaded into memory using the DOM (Document Object Model). Graphics Device Interchange (GDI+) was used to support features such as scrolling and selecting graphic objects. The windows Graphics Device Interchange enables applications to use graphics and formatted text on both the video display and the printer. Windows-based applications do not access the graphics hardware directly; instead, the GDI interacts with device drivers on behalf of applications.

8.5.2 Event Scheduler

The event scheduler was designed to stack events into a simulation queue according to priorities or a rule set that is configurable and the events would then be executed according to the desired schedule. The scheduler can be designed using any data structure. The event scheduler would have to talk to the clock or the timer, which would associate events with a particular time. Since this would be essential information for the simulation, this entire class can be serialised using XML serialisation and the data is collected using XML DOM object. The Document Object Model is a platform and language-neutral interface that allows programs and scripts to dynamically access and update the content, structure and style of documents. The document can be further processed and the results of that processing can be incorporated back into the presented page. The overview of DOM-related materials are available at W3C and around the web [155].

8.5.3 The Object Hierarchy for the Nodes

The nodes that form the object hierarchy are designed so that they optimise the development process. All nodes are derived from a Graphic Object class in the .NET Framework. This is the shared base class which has three child classes, viz: Image

Graphic, Text Graphic and Shape Graphic classes. The Image Graphic class has two children: Linked Image Graphic and Embedded Image Graphic. All the nodes used in the simulation are derived from the Linked Image Graphic child class. The following nodes are derived from the linked Image Graphic: Mobile node, Access Point, Base Station, Router, Gateway and Internet nodes. The individual nodes have features such as number of connections etc. which are features related the specific individual nodes. The shared base class, GraphicObject, contains all of the properties that are common to all of the possible graphic elements, such as position and size. Each individual subclass is built on that base by adding properties such as Text, Font, and FontColor for the TextGraphic subclass and ImagePath for the LinkedImageGraphic subclass. This underlying object library also includes a strongly-typed GraphicObjectCollection class, which will be used to hold a set of GraphicObject instances. Each individual class in this object model overrides the Draw method of the base class so that it knows how to handle its own drawing; this will allow the final design surface code to just call each element's Draw method, without any knowledge of the specific element type. There is a great deal of code in this little object library, even though it only includes a small number of objects, but the most interesting code is found in each class's implementation of the Draw method. In each case, this is the code that will be used to display the object onto the final design surface; it has to be able to handle drawing the element at a specified location and possibly with a certain degree of rotation.

8.5.4 XML Engine and simulation

The XML engine is part of the .NET framework which converts the design surface into XML files and stores them on the disc after the simulation. This engine uses built-in functions, such as getting nodes which contain information related to the

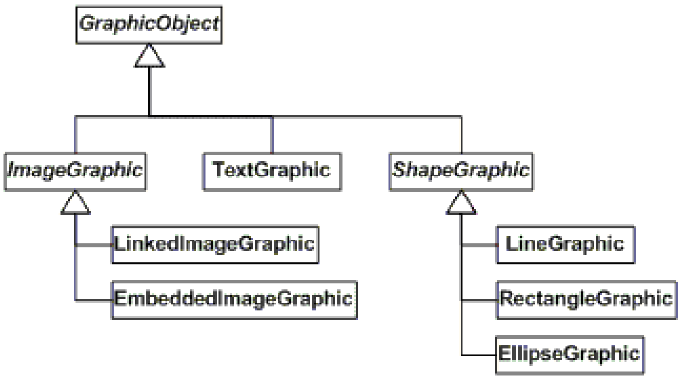

Figure 8.2: Object hierarchy of the graph object

time of arrival of the packet and other simulation related information. The schema in Fig. 8.3 shows a summary of the features of the XML file which stores the simulation details. The engine is also used to store the results of the simulation which will be used later on for further developments.

The simulation engine talks to the event scheduler and the XML engines to run the program. It decides which events should take place and orders them as the simulation proceeds. As previously discussed, the GUI environment allows a user to edit a network topology and configure the modules present in a node. The advantage of the GUI is that the parameters can be set and this generates the simulation job description input files. The simulation engine takes the simulation job description according to the parameters which are set in the input file, runs the simulation and the results are stored in XML format. There is lot of future work related to this, as we can generate events which can be based on scenarios, and they can be ordered based on certain rules sets that can be configured.

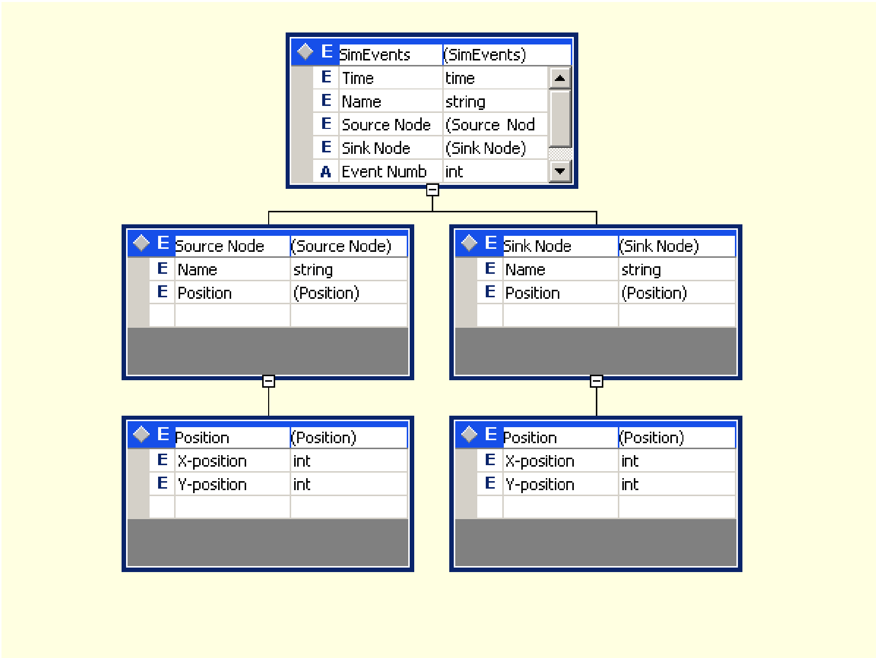

Figure 8.3: The node schema for the features of XML file which stores the simulation results

8.6 Simulation Model

8.6.1 Experimental Setup

In our experimental setup, using the topology editor we set up a suitable scenario to demonstrate a handoff between a WLAN access point and a cellular network. The topology editor provides a convenient way to drag and drop the required components onto the design surface (see Fig. 8.4). It also helps to specify various parameters such as an IP address for a specific router or a node. The links also have specific delays associated with them. The topology editor helps to add and delete a node as per the requirements of the mobility scenario. In our simulation, the mobile node was restricted to move in a straight line and random movements were not implemented.

In this model, about 7 access points were connected to 3 access routers which were separated by a distance of D metres. The mobile device moves from one cell

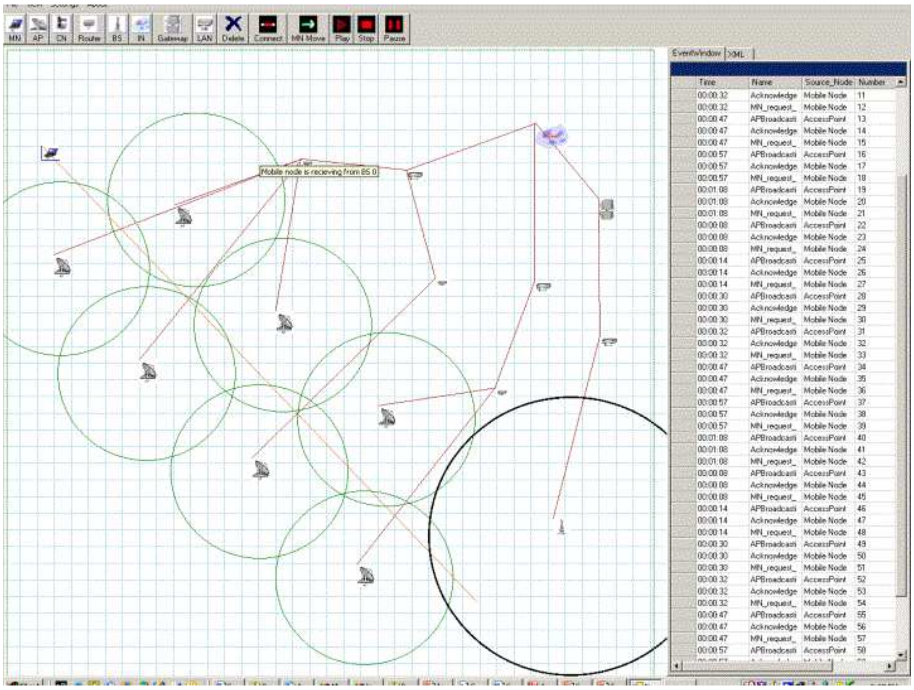

Figure 8.4: The simulation tool user interface developed using Microsoft C#

to another with a constant velocity and the received signal strength is sampled at a constant distance interval (in metres). Here, the mobile node moves from one AP to the other with constant speed. The signals from the AP are affected by two major factors: path loss and log-normal fading. Rayleigh fading is not taken into account as it is assumed that the rapid fluctuations are averaged out. The received signal strengths from the current AP to the target AP are sampled at distances Kd_s. The model also includes slow fading. The received signal strengths a_t and b_t (in dB) when the mobile is at a given distance kd_s are given by:

$$a_t = K_1 - K_2 \log kd_s + u_t \tag{8.1}$$

$$b_t = K_1 - K_2 \log (N - k)\, d_s + u_t \tag{8.2}$$

where $N = D/d_s$. The parameters $K_1 = 0$ and $K_2 = 30$ in dB are typical of an urban environment accounting for path loss. K_1 is the signal strength at distance $d = 1$, and K_2 is the path loss component. The simulation parameters used for the movement detection are as shown in Table 8.1.

Number of Base Stations	1
Number of Access Points	7
Trajectory of mobile node	Straight Path
Sampling distance for Base Stations	10 m
Sampling distance for Access Points	1 m
Distance between Access Points	200 m
Base Stations coverage area	2000 m
Path loss (K)	30 dB
Transmitter power	0 dB
Fading Process	Lognormal fading
Standard Deviation (u_k)	8 dB

Table 8.1: The simulation parameters used for the prediction model

8.7 Results

In this chapter, we have briefly discussed a simulation framework, from a software development perspective, which helps as a platform for future research. The research included only the layer-2 detection mechanisms which was our focus. Although, the packet losses and handoff latency were not measured in the simulation, we implemented the layer-2 detection mechanism which included our hybrid prediction model. The parameters for both the base station and the access points were selected as shown in Table 8.1. It is important to note that the mobile node moved in a straight line over different access points and an overlay network consisting of a cellular network covered all the access points. We made the cell area of the cellular network look smaller just for the visibility of the cell area (Fig. 8.4).

For the detection mechanism at layer-2, we included the mobility prediction algorithm discussed earlier in chapter 4. All the required building blocks of the hybrid prediction methodology were assembled together into a unified algorithm for prediction of the received signal strengths. In this section, we present results based of the hybrid mobility prediction based model for the base station of the cellular network. Similarly, we applied the same algorithms to the access points of the WLAN to predict the signal strengths from the AP. The results of the mobility prediction in Fig. 8.5 show a plot of the actual values of the signal strength and the corresponding predicted values from the basestation of the cellular network. The basic prediction model tracks the signal strengths but with some error. These variations in prediction values are as shown in Fig. 8.6. In the simulation tool, the handoff decision was made based on future received signal strengths and the handoff in the layer-3 is still to be implemented for the MIPv6 protocol which assists the mobility during handoff. In summary, the results show the prediction of the signal strengths of the cellular network and contributions regarding handoff procedures at the network layer is still ongoing work.

8.8 Extensions to the Simulation Tool

Significant amount of work needs to be done to refine the simulation tool so that it can be used to implement and meet the standards proposed by 3GPP and the IETF. The unsolved problems can be broadly classified into two categories, one corresponding to the implementation of the MIPv6 protocol and other in refining the algorithms used by the protocol. Solving the former will ensure that the protocol will work when it is deployed whereas solving the latter will improve the performance

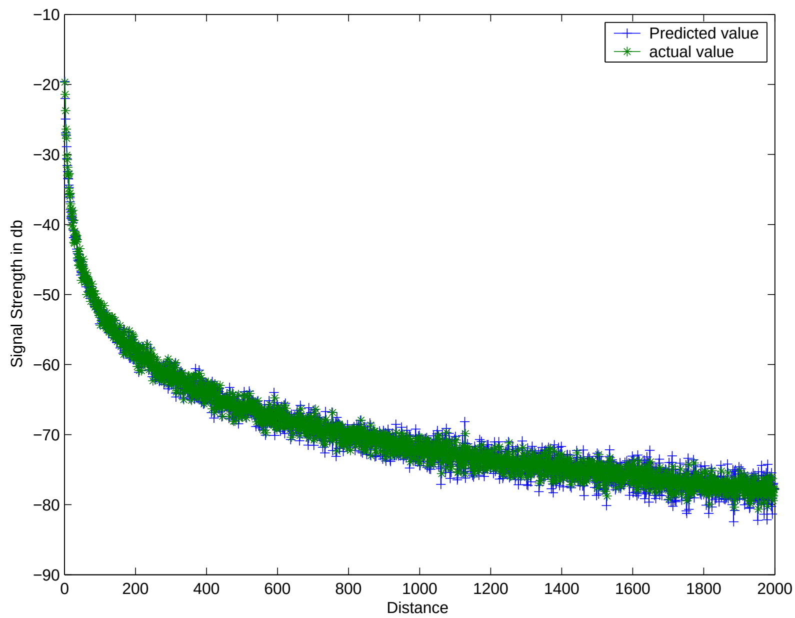

Figure 8.5: The prediction of received signal strength from the base station in the simulation tool

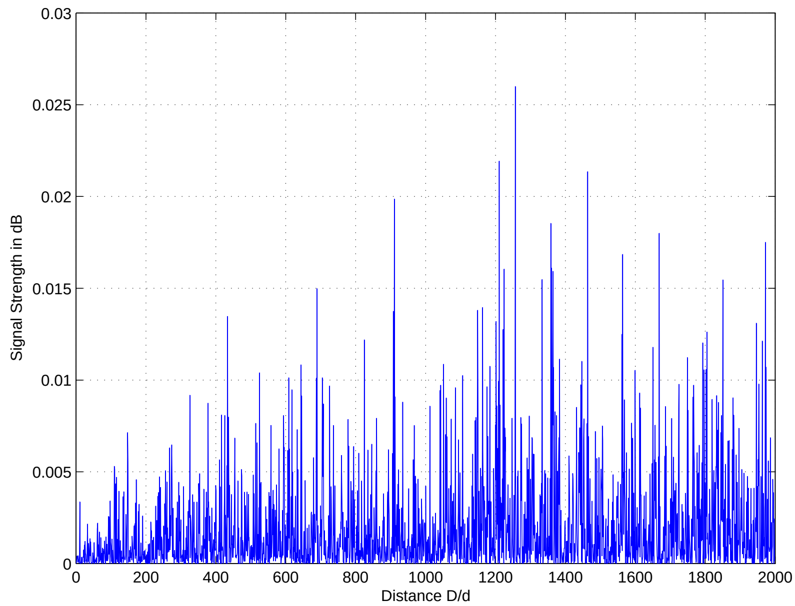

Figure 8.6: The absolute error from the base station in the simulation tool

of the protocol. Some of the areas that need specific attention are:

- UDP streaming from the corresponding node to the mobile node.
- Improved layer-2 performance when the mobility prediction algorithm is employed.
- The binding updates and resource reservation during handoff.
- Any packet losses during handoff

8.9 Summary

In this chapter, we have discussed our work related to the implementation of a mobility prediction algorithm with an emphasis on the layer-2 detection mechanism in our simulation model/tool to improve the performance of handoff in wireless networks. The chapter also discussed the software oriented approach adopted to demonstrate handoff in both homogeneous and heterogeneous technologies; specifically using the mobility protocol viz., Mobile IPv6 protocol (MIPv6). There are many simulation tools that deal with the network layer and transport layer that have been mentioned in the literature. While the research community has addressed the programmability of the various layers and issues of QoS in wireless networks, little work has been done on layer-2 detection mechanisms. In our research, we investigated the layer-2 handoff detection mechanism which is integrated into the simulation tool to show its importance (mobility prediction) when used with the standard existing protocols. Our tool focussed on mobility prediction schemes to identify the next potential access point/base station in order to improve handoff performance. The chapter includes an outline of the software design and implementation of such a tool that is intended to be a platform for our future research activities. Although the specific scenario only implements detection mechanisms and a basic implementation of the MIPv6 protocol, this is still an ongoing work. Our future work involves improving the tool for other standard wireless protocols such as HMIPv6.

Chapter 9

Conclusions and Future Directions

9.1 Research Contributions

In this chapter, we conclude this dissertation with a summary of our research contributions and an examination of areas for future work. In this dissertation, we have examined various methods that allow mobile users to roam without interruption to their network communication. The objective of this work has been to create a networking environment in which mobile users can use continuous media, such as audio and video, as well as more traditional network services. We have developed and used several new techniques to support communications handoff without adversely affecting loss, bandwidth or end-to-end delay. Now, with the need to consider QoS in wireless networks, handoff performance plays an important role and methods to improve this performance should be taken into account. Specifically, it must include techniques which will help in reducing losses and reserve the right amount of resources during handoff. In this thesis, we have examined the issues surrounding the design of an efficient handoff and proposed a framework/architecture with the help of new techniques. The techniques that provide this are the design of mobility prediction and multicasting techniques that significantly improve the performance during handoff. The following are the generic features that will allow the design of suitable mobility environment based on the emerging new technologies:

- Mobility prediction techniques or link layer detection that determines the position and direction to the mobile node.
- Multicasting to efficiently reserve the required resources during handoff.
- Reduce the join and prune operations and select an efficient path from the source to destination over which to deliver the data.

We have incorporated these features into an efficient model which includes decision for handoff and determining the allocation of resources in the network for a possible handoff. To be able to efficiently solve the problem of handoff we proposed a basic mobility prediction model which involves prediction of signal strength based on Grey theory. We extended this basic approach, to improve the performance of the prediction model by proposing a hybrid version of the prediction methodology which includes the use of fuzzy inference rules and investigated three different optimisation techniques to further reduce errors. The three optimisation techniques were a self-tuning algorithm (STA), a genetic algorithm (GA) and a particle swarm optimisation (PSO) algorithm. In the self-tuning algorithm, the prediction errors from the Grey model (the difference between the error from the actual and the predicted values) are fine-tuned using fuzzy rules. The self-tuning algorithm uses a learning method with an iterative approach, minimising the error and thus achieves an optimal value. Similarly to the self-tuning algorithm, the genetic algorithm performs the same function but uses an evolutionary approach by using operations such as selection, mutation and crossover. Finally, we used a particle swarm optimisation technique which is a stochastic approach to reduce the error. All the algorithms were compared with respect to convergence to a minimum value. In this way, we found a suitable optimisation technique to improve the error and achieve superior prediction accuracy.

Although many methods have been used to evaluate the performance of handoff in wireless networks in terms of handoff latency and packet losses, we have used the idea of mobility prediction based on the use of RSSI values to improve on performance during handoff. For our first optimisation technique, namely, the self-tuning algorithm, it was based on the gradient descent method. The main idea here was to tune the membership functions in the antecedent part and the real number in the consequent part of the inference rules. The motivation was to improve the learning speed and achieve an optimal value in a very short time. In the second optimisation technique, the genetic algorithm, we used similar parameters to fine tune the fuzzy inference rules using an evolutionary technique. Although the Genetic algorithm had a shorter convergence time to reach the optimal as compared to the self-tuning algorithm the computation time was greater as it required a fairly large population size. In the third optimisation technique, PSO was used which was the best suited technique for our hybrid model in terms of computation time and the convergence achieved. The PSO was tested with different population sizes and compared to the genetic algorithm approach. In almost every experiment (approximately 90% of the time), PSO faired much better. The PSO based model presented in chapter 5 incorporates results (refer Fig. 5.10) conducted for different PSO population sizes in which all the population sizes faired better than GA in most cases. However, in GA due to the fact that it required a relatively large population size to perform operations such as selection, mutation and crossover the computation time was greater. A simulation study validated our model and it worked best with the PSO model, giving high prediction accuracy for the mobility environment. The compensated models for the GA and PSO were also plotted to show the prediction accuracy which was discussed in chapter 5.

The PSO based hybrid prediction model was proposed along with a suitable multicasting algorithm to save on resources during handoff. For this purpose, we developed a framework/architecture discussed in chapter 6 which incorporates the mobility prediction algorithm to select the potential AR to which the mobile node is moving and construct a minimum multicast tree spanning from the source to the destination(s) based on specified constraints and objectives. The motivation for using the concept of multicasting is that we can reserve the right amount of resources from the source to the destination. Our framework is compared to the CAR-set algorithm proposed by Helmy et. al. and has been to shown to improve handoff in terms of resources being utilised. We also proposed two algorithms namely, the MMP algorithm and K-Minhop algorithm for multicasting during handoff depending on application requirements. These algorithms together with the aid of mobility prediction help to develop a unified framework for improving handoff in mobility environment. The results showed very good performance for this architecture and exposed the significant advantages of this model with respect to selection of potential AR's as well as resource reservation. The various tests with a different number of highlighted potential ARs showed that our model tend to reserve fewer resources than the CAR-set algorithm (< 8 times). Thus, far more resources are saved by using prediction algorithms with multicasting during handoff.

Further work on the architecture was done to observe joins and prunes during handoff. It was important to show that by using mobility prediction and multicasting we not only save on resources but also reduce the number of joins and prunes during handoff. Therefore, we developed another algorithm called the MBWDC algorithm which makes use of our mobility prediction and multicasting techniques discussed in chapter 7 that finds the optimal solution for our problem. The model formulation was based on two factors namely, 1) the bounded end-to-end delay along the individual paths from the source to the destination and 2) the minimum bandwidth cost of the multicast tree. Computational tests on problems of various topology sizes led to the conclusion that the proposed algorithm is quite powerful in comparison with the

CAR-set algorithm. Network sizes of up to 100 nodes were handled with minimal bandwidth utilisation, joins and prune operations.

The work on homogeneous and heterogeneous networks mainly the handoff were demonstrated on the simulation tool which we call "NeTSim-v3.0". The model incorporates our mobility prediction model for homogeneous and heterogeneous handoff in wireless networks. The work includes an outline of the software design and implementation of such a tool that is intended to be a platform for our future research activities.

9.2 Summary of Contributions

We now summarise the list of contributions in this thesis below:

1. Development of a prediction model using the RSSI values for wireless networks with the help of Grey theory which has high prediction accuracy.
2. Introduction of a hybrid prediction model that can be used to minimise the errors in the basic Grey prediction model. The compensation model includes the Grey prediction model, a fuzzy controller and an optimisation technique – namely, particle swarm optimisation – to minimise the errors.
3. The comparison of the model is made with the different optimisation techniques namely the self tuning algorithm(STA), the genetic algorithm (GA) and particle swarm optimisation (PSO) to fine-tune the fuzzy parameters to minimise errors. The convergence to optimal values was tested on the various optimisation techniques with respect to our problem. The genetic algorithm and the PSO were tested for different population sizes by varying the parameters that govern these algorithms.
4. Design and implementation of the hybrid mobility prediction model to select the appropriate access router to support multicasting in wireless networks

during handoff.

5. Development and implementation of a multicasting model for wireless environment to improve the performance of handoff was proposed. Two algorithms namely, MMP algorithm and k-Minhop were proposed based on the objectives and constraints as required by the application.
6. Development and implementation of framework/architecture to support multicasting in wireless environment to reduce the joins and prunes during handoff. The MBWDC algorithm was proposed to minimise the resources utilised during handoff. The comparison was made with CAR-set algorithm to test performance of the proposed algorithms.
7. All the building blocks i.e. the mobility prediction algorithm and multicasting techniques with algorithms were integrated into the software based simulation tool.
8. Finally, the work on heterogeneous networks mainly the handoff between different technologies were demonstrated on the tool. Future work is based on our simulation tools to test the prediction algorithms and multicasting techniques in a heterogeneous environment.

9.2.1 Extensions of the work

Although the required building blocks for consideration of handoff performance with mobility prediction and multicasting are developed in this thesis, the measurement of packet losses has not been implemented in the simulation. Therefore, in this thesis we have provided only an evaluation of the accurate mobility prediction and multicasting algorithms to reserve the right amount of resources for handoff based on proposed objectives and constraints in wired links. In the author's opinion, it is highly desirable to implement these building blocks and the architecture into existing mobility protocols in order to verify our algorithms in practice. The standard approach we have taken is to implement these algorithms separately. As mobility detection or hints from the lower link layer play an important role, it becomes necessary to have a detection mechanism which would significantly vary the performance of handoff. Also with the help of multicasting, we can accurately influence system performance.

The general principle of optimisation techniques used in the hybrid prediction model creates a large area for further research – especially using particle swarm optimisation. Due to the fact that it is a stochastic technique which is governed by parameters such as the velocity of the particle, acceleration coefficients and inertia; these parameters can be further investigated to improve the convergence further. Although the population sizes do not have a significant effect on PSO performance, work has been done on this to observe the speed of convergence to a minimal value. Investigation in the performance of the prediction model based on the variation of different parameters, e.g. higher dimensionality would be very interesting, but it was clearly beyond the scope of this research work.

Further research can also address the developments of the multicasting algorithms proposed for efficient handoff. With respect to the problem formulation presented in chapter 6, which provides a bias of one parameter over the other, we can further investigate for a cost balanced multicast scheme or a trade-off parameter which seeks a balance between the bandwidth cost and the minimum number of hops. The scheme can provide a quantitative trade-off relation between these two routing metrics for efficient multicasting routing.

The problem formulation for the multicasting routing including the bounded delay on the path and the bandwidth objectives could also be investigated for packet losses during handoff. The algorithms could be added to standard protocols such as HMIPv6 and MIPv6 which can provide an advantage over existing standards by the

IETF. Lastly, the development of the simulation tool incorporating all the building blocks of mobility prediction and multicasting can incorporated existing protocols to measure performance during handoff. Currently, extensive research work is underway on the simulation tool, NeTSim-v3.0 for all the proposed algorithms for both homogeneous and heterogeneous networks.

9.2.2 Challenges for Mobile Computing

Next generation networks should be able to access different networks and interoperate with each other to ensure global mobility and access information anywhere, anytime. As new wireless technologies emerge, we have realised that future wireless networks will have unique performance advantages and disadvantages. The variance in performance could be due to various factors, one of these being different network devices. Therefore future wireless networks are likely to be heterogeneous in nature with different devices and technologies being used. Hence the network should be able to support different application services and dynamically adapt to the state of their network connectivity.

For the deployment of different wireless technologies such as 3G cellular networks and WiFi networks there should be efficient protocols for mobile devices and solutions for interoperability problems. For mobile devices that can access different networks, a new type of handoff called “vertical handoff" should be handled efficiently. In this thesis we have mainly discussed horizontal handoff i.e., handoff in the same network type. However, there are research challenges with vertical handoff which incorporate horizontal handoff, which will be the future of wireless networks.

Bibliography

[1] M. Chiu and M. Bassiouni, "Predictive schemes for handoff prioritisation in cellular networks based on mobile positioning," *IEEE Journal on, Selected Areas in Communications*, vol. 18, p. 510 Ű522, March 2000.

[2] J. L. Deng, "Introduction to Grey system theory," *J. Grey Syst.*, vol. 1, no. 1, pp. 1–24, 1989.

[3] G. Pollini, "Trends in handover design," *Communications Magazine, IEEE*, vol. 34, no. 3, pp. 82–90, 1996.

[4] S. Venkatachalaiah, R. J. Harris, and J. Murphy, "Improving handoff in wireless networks using Grey and particle swarm optimisation," in *CCCT*, vol. 5, 2004, pp. 368–373.

[5] S. Venkatachalaiah, R. J. Harris, and J. Murphy, "An approach to improve handoff in wireless networks using mobility prediction," in *The 3rd Australian Telecommunications CRC (ATcrc) Conference and workshop*, vol. 1, 2003.

[6] S. Venkatachalaiah, R. J. Harris, and J. Murphy, "Improving handoff in handoff in wireless networks using mobility prediction and multicasting techniques," in *The 4th Australian Telecommunications CRC (ATcrc) Conference and workshop*, vol. 1, 2004.

[7] S. Venkatachalaiah, R. J. Harris, and D. Jones, "CE.NET embedded IPv6 performance issues - heterogeneous technologies," in *DevCom-04,Microsoft Windows Embedded Developers' Conference 2004*, 2004.

[8] S. Venkatachalaiah, R. J. Harris, and D. Jones, "CE.NET embedded IPv6 performance issues - heterogeneous technologies," in *RPC grant for universities, 2004*, 2004.

[9] S. Venkatachalaiah, R. J. Harris, and R. Suryasaputra, "Improvement of handoff performance in wireless networks using mobility prediction and multicasting techniques," in *in Proc. of 4th WSEAS Int. Conf. on Electronics, Hardware, Wireless & Optical Communications (EHAC 2005).*

[10] S. Venkatachalaiah and R. J. Harris, "Improving handoff in wireless networks using mobility prediction and evolutionary algorithms," in *ICETE*, 2005, pp. 112–118.

[11] S. Venkatachalaiah, R. J. Harris, and R. Suryasaputra, "Improvement of handoff performance in wireless networks using mobility prediction and multicasting techniques," in *WSEAS transactions of Communications,Issue 2, February 2005*, vol. 4, 2005, pp. 104–111.

[12] S. Venkatachalaiah, R. J. Harris, and R. Suryasaputra, "An algorithm using join/prune mechanisms to improve handoff aided by mobility prediction in wireless networks," in *IEEE Malaysia International Conference on Communications and IEEE International Conference on Networks (MICC&ICON 2005*, 2005, pp. 1101–1107.

[13] H. Nomura, I. Hayashi, and N. Wakami, "A learning method of fuzzy inference rules by descent method," *Fuzzy Systems, 1992., IEEE International Conference on*, pp. 203 – 210, 8-12 March 1992.

[14] A. Helal, B. Haskell, J. L. Carter, R. Brice, D. Woelk, and M. Rusinkiewicz, *Anytime, Anywhere Computing, Mobile Computing - Concepts amd Technology.* Kluwer Acedamic publishers, Boston/Dordrecht/London, 1999.

[15] W.-S. Soh and H. Kim, "QoS provisioning in cellular networks based on mobility prediction techniques," *Communications Magazine, IEEE*, vol. 41, no. 1, pp. 86–92, 2003.

[16] S.-L. Su and J.-Y. Chen, "Performance analysis of soft handoff in cellular networks," in *Personal, Indoor and Mobile Radio Communications, 1995. PIMRC'95. 'Wireless: Merging onto the Information Superhighway'., Sixth IEEE International Symposium on*, vol. 2, 1995, pp. 587–591 vol.2.

[17] B. Senadji, "A new handover decision algorithm for cellular and microcellular systems," in *Signal Processing and Its Applications, 1996. ISSPA 96., Fourth International Symposium on*, vol. 1, 1996, pp. 228–231.

[18] B. Senadji, "A new handover decision algorithm for microcellular systems," in *Vehicular Technology Conference, 1997 IEEE 47th*, vol. 1, 1997, pp. 314–318 vol.1.

[19] D. Wong and T. Lim, "Soft handoffs in CDMA mobile systems," *IEEE Personal Communications*, vol. 4, p. 6 Ũ17, Dec. 1997.

[20] A. Jayasuriya and J. Asenstorfer, "Mobility prediction model for cellular networks based on the observed traffic patterns," in *In Proc. IASTED International Conference on Wireless and Optical Communication (WOC 2002), Banff, Canada.*, vol. 1, July 2002, pp. pp. 386–391.

[21] H.-W. Ferng, W.-Y. Kao, D. Shiung, C.-L. Liu, H.-Y. Chen, and H.-Y. Gu, "A dynamic resource reservation scheme with mobility prediction for wireless

multimedia networks," in *Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th*, vol. 5, 2004, pp. 3506–3510 Vol. 5.

[22] W.-S. Soh and H. Kim, "Dynamic bandwidth reservation in cellular networks using road topology based mobility predictions," in *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, vol. 4, 2004, pp. 2766–2777 vol.4.

[23] J. Chan, S. Zhou, and A. Seneviratne, "A QoS adaptive mobility prediction scheme for wireless networks," in *Global Telecommunications Conference, 1998. GLOBECOM 98. The Bridge to Global Integration. IEEE*, vol. 3, 1998, pp. 1414–1419 vol.3.

[24] H. Tran and R. Harris, "Solving QoS multicast routing with genetic algorithms," in *Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on*, vol. 3, 2003, pp. 1944–1948 vol.3.

[25] J. Mysore and V. Bharghavan, "A new multicasting-based architecture for internet host mobility," in *Proceedings of the 3rd annual ACM/IEEE international conference on Mobile computing and networking.* ACM Press, 1997, pp. 161–172.

[26] S. Sanyal, L. Nahar, and S. Bhattacharya, "Real-time multicast in wireless communication," in *Parallel Processing, 1997., Proceedings of the 1997 International Conference on*, pp. 500–507.

[27] D. Waitzman, C. Partridge, and S. Deering, "Distance vector multicast routing protocol RFC 1075," 1988. [Online]. Available: http://www.ietf.org/rfc/rfc1075.txt

[28] A.-G. Helmy, M. Jaseemuddin, and G. Bhaskara, "Multicast-based mobility: a novel architecture for efficient micromobility," *Selected Areas in Communications, IEEE Journal on*, vol. 22, no. 4, pp. 677–690, 2004.

[29] S. Seshan, "Low-latency handoffs for cellular data networks," 1996. [Online]. Available: sherry.ifi.unizh.ch/seshan96lowlatency.html

[30] N. Tripathi, J. Reed, and H. VanLandinoham, "Handoff in cellular systems," *Personal Communications, IEEE [see also IEEE Wireless Communications]*, vol. 5, no. 6, pp. 26–37.

[31] A. Aljadhai and T. Znati, "Predictive mobility support for QoS provisioning in mobile wireless environments," *Selected Areas in Communications, IEEE Journal on*, vol. 19, no. 10, pp. 1915–1930.

[32] M. Asawa and W. Stark, "A framework for optimal scheduling of handoffs in wireless networks," in *Global Telecommunications Conference, 1994. GLOBECOM '94. 'Communications: The Global Bridge'., IEEE*, vol. 3, 1994, pp. 1669–1673 vol.3.

[33] X. Zhou, R. Chen, and C. Gao, "An elliptical shadow algorithm for motion prediction and resource reservation in wireless cellular networks," in *Computer Communications and Networks, 2001. Proceedings. Tenth International Conference on*, 2001, pp. 234–239.

[34] G. Liu and J. Gerald Maguire, "A class of mobile motion prediction algorithms for wireless mobile computing and communication," *Mob. Netw. Appl.*, vol. 1, no. 2, pp. 113–121, 1996.

[35] T. Liu, P. Bahl, and I. Chlamtac, "A hierarchical position-prediction algorithm for efficient management of resources in cellular networks," in *Global*

Telecommunications Conference, 1997. GLOBECOM '97., IEEE, vol. 2, 1997, pp. 982–986 vol.2.

[36] G. Liu and J. Maguire, G.Q., "A predictive mobility management algorithm for wireless mobile computing and communications," in *Universal Personal Communications. 1995. Record., 1995 Fourth IEEE International Conference on*, 1995, pp. 268–272.

[37] I.-R. Chen and N. Verma, "Simulation study of a class of autonomous host-centric mobility prediction algorithms for wireless cellular and ad hoc networks," in *Simulation Symposium, 2003. 36th Annual*, 2003, pp. 65–72.

[38] D. Levine, I. Akyildiz, and M. Naghshineh, "A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept," *Networking, IEEE/ACM Transactions on*, vol. 5, no. 1, pp. 1–12, 1997.

[39] F. Yu and V. Leung, "Mobility-based predictive call admission control and bandwidth reservation in wireless cellular networks," in *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 1, 2001, pp. 518–526 vol.1.

[40] C. Choi and E. Cheung, "Sequential handoff algorithm for cellular mobile communication," in *Global Telecommunications Conference, 1997. GLOBECOM '97., IEEE*, vol. 2, 1997, pp. 716–719 vol.2.

[41] E. Hernandez and A. Helal, "Examining mobile-IP performance in rapidly mobile environments: the case of a commuter train," in *Local Computer Networks, 2001. Proceedings. LCN 2001. 26th Annual IEEE Conference on*, 2001, pp. 365–372.

[42] N. Frangiadakis, M. Kyriakakos, S. Hadjiefthymiades, and L. Merakos, "Realistic mobility pattern generator: design and application in path prediction algorithm evaluation," in *Personal, Indoor and Mobile Radio Communications, 2002. The 13th IEEE International Symposium on*, vol. 2, 2002, pp. 765–769 vol.2.

[43] B. Liang and Z. J. Haas, "Predictive distance-based mobility management for multidimensional PCS networks," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 718–732, 2003.

[44] P. Bilurkar, N. Rao, G. Krishna, and R. Jain, "Application of neural network techniques for location predication in mobile networking," in *Neural Information Processing, 2002. ICONIP '02. Proceedings of the 9th International Conference on*, vol. 5, 2002, pp. 2157–2161 vol.5.

[45] X. Hong, K. Xu, and M. Gerla, "Scalable routing protocols for mobile ad hoc networks," *Network, IEEE*, vol. 16, no. 4, pp. 11–21, 2002.

[46] S.-J. Lee, W. Su, and M. Gerla, "Ad hoc wireless multicast with mobility prediction," in *Computer Communications and Networks, 1999. Proceedings. Eight International Conference on*, 1999, pp. 4–9.

[47] D. Johnson and D. Maltz., "Dynamic source routing in ad hoc wireless networks." 1996.

[48] Y. H. D. B. Johnson, D. A. Maltz and J. G. Jetcheva., "Dynamic source routing protocol for mobile ad hoc networks DSR." 2002. [Online]. Available: http://www.ietf.org/internet-drafts/draft-ietf-manet-dsr-07.txt, Feb 2002

[49] C. Perkins and E. Royer., "Ad hoc on demand distance vector AODV routing," 1998.

[50] Z. J. Haas and M. R. Pearlman, "The performance of query control schemes for the zone routing protocol," *IEEE/ACM Trans. Netw.*, vol. 9, no. 4, pp. 427–438, 2001.

[51] M. Pearlman and Z. Haas, "Determining the optimal configuration for the zone routing protocol," *Selected Areas in Communications, IEEE Journal on*, vol. 17, no. 8, pp. 1395–1414, 1999.

[52] C.-C. Chiang and M. Gerla, "Routing and multicast in multihop, mobile wireless networks," in *Universal Personal Communications Record, 1997. Conference Record., 1997 IEEE 6th International Conference on*, vol. 2, 1997, pp. 546–551 vol.2.

[53] Y.-B. Ko and N. H. Vaidya, "Location-aided routing (LAR) in mobile ad hoc networks," *Wirel. Netw.*, vol. 6, no. 4, pp. 307–321, 2000.

[54] A. Bhattacharya and S. K. Das, "LeZi-update: an information-theoretic framework for personal mobility tracking in PCS networks," *Wirel. Netw.*, vol. 8, no. 2/3, pp. 121–135, 2002.

[55] E. Kaplan, *Understanding the GPS: Principles and Applications.* Artech House, Boston, MA, Feb. 1996, 1996.

[56] W. Su, S. J. Lee, and M. Gerla, "Mobility prediction in wireless networks," in *MILCOM 2000, 21st Century Military Communications Conference Proceedings*, vol. 1, 2000, pp. 491–495.

[57] K. Wang and B. Li, "Group mobility and partition prediction in wireless ad-hoc networks," in *Communications, 2002. ICC 2002. IEEE International Conference on*, vol. 2, 2002, pp. 1017–1021 vol.2.

[58] W. G. Yanlei Shang and S. Cheng, "Clustering algorithm based on wavelet neural network mobility prediction in mobile ad hoc network," in *Lecture Notes in Computer Science, Springer-Verlag Berlin Heidelberg, 2005*, vol. 3498, 2005, pp. 391 – 396.

[59] S. Abhyankar and D. Agrawal, "Distributed mobility-aware route selection for wireless ad hoc networks," in *Performance, Computing, and Communications Conference, 2002. 21st IEEE International*, 2002, pp. 241–247.

[60] A. J. N. S. Chellappa Doss, R., "A comparative study of mobility prediction in fixed wireless and mobile ad hoc networks," *IEEE International Conference on Communications, International Conference on Communications (ICC), USA*, 2003.

[61] R. Chellappa, A. Jennings, and N. Shenoy, "Route discovery and reconstruction in mobile ad hoc networks," in *Networks, 2003. ICON2003. The 11th IEEE International Conference on*, 2003, pp. 549–554.

[62] X. Hong, M. Gerla, G. Pei, and C.-C. Chiang, "A group mobility model for ad hoc wireless networks," in *MSWiM '99: Proceedings of the 2nd ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems.* New York, NY, USA: ACM Press, 1999, pp. 53–60.

[63] N.-K. Ha, B. Lee, K.-J. Kim, and K. J. Han, "A partition prediction algorithm for group mobility in ad-hoc networks." in *ICN (2)*, 2005, pp. 210–217.

[64] A. B. McDonald and T. Znati, "Predicting node proximity in ad-hoc networks: a least overhead adaptive model for selecting stable routes," in *MobiHoc '00: Proceedings of the 1st ACM international symposium on Mobile ad hoc networking & computing.* Piscataway, NJ, USA: IEEE Press, 2000, pp. 29–33.

[65] J. Shengming, H. Dajiang, and R. Jianqiang, "A prediction-based link availability estimation for mobile ad hoc networks," in *INFOCOM*, 2001, pp. 1745–1752. [Online]. Available: citeseer.ist.psu.edu/article/jiang01predictionbased.html

[66] A. McDonald and T. Znati, "A mobility-based framework for adaptive clustering in wireless ad hoc networks," *Selected Areas in Communications, IEEE Journal on*, vol. 17, no. 8, pp. 1466–1487, 1999.

[67] W. Su and M. Gerla, "IPv6 flow handoff in ad hoc wireless networks using mobility prediction," in *Global Telecommunications Conference, 1999. GLOBECOM '99*, vol. 1A, 1999, pp. 271–275 vol.1a.

[68] W. S. S.-J. Lee and M. Gerla, "On-demand multicast routing protocol ODMRP for ad hoc networks."

[69] T. G. Harrison, C. L. Williamson, W. L. Mackrell, and R. B. Bunt, "Mobile multicast (MoM) protocol: multicast support for mobile hosts," in *MobiCom '97: Proceedings of the 3rd annual ACM/IEEE international conference on Mobile computing and networking.* New York, NY, USA: ACM Press, 1997, pp. 151–160.

[70] C. L. Williamson, T. G. Harrison, W. L. Mackrell, and R. B. Bunt, "Performance evaluation of the MoM mobile multicast protocol," *Mob. Netw. Appl.*, vol. 3, no. 2, pp. 189–201, 1998.

[71] F. Rui, J. Yue-hui, and C. Shi-duan, "ScaleMoM: a mobile multicast scheme and its performance analysis," in *Info-tech and Info-net, 2001. Proceedings. ICII 2001 - Beijing. 2001 International Conferences on*, vol. 2, 2001, pp. 378–383 vol.2.

[72] A. Helmy, M. Jaseemuddin, and G. Bhaskara, "Efficient micro-mobility using intra-domain multicast-based mechanisms (M&M)," *SIGCOMM Comput. Commun. Rev.*, vol. 32, no. 5, pp. 61–72, 2002.

[73] F. Bai, G. Bhaskara, and A. Helmy, "Building the blocks of protocol design and analysis: challenges and lessons learned from case studies on mobile ad hoc routing and micro-mobility protocols," *SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 3, pp. 57–70, 2004.

[74] A. Helmy, "State analysis and aggregation study for multicast-based micro mobility," in *Communications, 2002. ICC 2002. IEEE International Conference on*, vol. 5, 2002, pp. 3301–3306 vol.5.

[75] C. L. Tan and S. Pink, "MobiCast: a multicast scheme for wireless networks," *Mob. Netw. Appl.*, vol. 5, no. 4, pp. 259–271, 2000.

[76] C. R. Lin and K.-M. Wang, "Scalable multicast protocol in IP-based mobile networks," *Wirel. Netw.*, vol. 8, no. 1, pp. 27–36, 2002.

[77] C. Perkins, "IP mobility support RFC 2002," 1996. [Online]. Available: http://www.ietf.org/rfc/rfc2002.txt

[78] J. Moy, "Multicast routing extensions for OSPF," *Commun. ACM*, vol. 37, no. 8, pp. 61–ff., 1994.

[79] T. Ballardie, P. Francis, and J. Crowcroft, "Core based trees CBT," in *Conference proceedings on Communications architectures, protocols and applications*. ACM Press, 1993, pp. 85–95.

[80] S. Deering, D. L. Estrin, D. Farinacci, V. Jacobson, C.-G. Liu, and L. Wei, "The PIM architecture for wide-area multicast routing," *IEEE/ACM Trans. Netw.*, vol. 4, no. 2, pp. 153–162, 1996.

[81] S. J. Yang and S. H. Park, "An enhanced multicast routing protocol for host mobility in IPv6 based networks," in *Consumer Communications and Networking Conference, 2005. CCNC. 2005 Second IEEE*, pp. 386–390.

[82] A. Helmy, S. Gupta, and D. Estrin, "The STRESS method for boundary-point performance analysis of end-to-end multicast timer-suppression mechanisms," vol. 12, no. 1. Piscataway, NJ, USA: IEEE Press, 2004, pp. 44–58.

[83] A. Helmy, "A multicast-based protocol for IP mobility support," in *COMM '00: Proceedings of NGC 2000 on Networked group communication.* New York, NY, USA: ACM Press, 2000, pp. 49–58.

[84] S. Wang and C.-H. Wu, "Effective handoff method using mobile location information," in *Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS 53rd*, vol. 4, 2001, pp. 2585–2589 vol.4.

[85] S. Wang, A. Rajendran, and M. Wylie-Green, "Adaptive handoff method using location information," in *Personal, Indoor and Mobile Radio Communications, 2001 12th IEEE International Symposium on*, vol. 1, 2001, pp. D–43–D–47 vol.1.

[86] H.-M. Feng and C.-C. Wong, "A on-line rule tuning Grey prediction fuzzy control system design," in *Neural Networks, 2002. IJCNN '02. Proceedings of the 2002 International Joint Conference on*, vol. 2, 2002, pp. 1316–1321.

[87] J.-R. Wu and M. Ouhyoung, "A 3D tracking experiment on latency and its compensation methods in virtual environments," in *Proceedings of the 8th annual ACM symposium on User interface and software technology.* ACM Press, 1995, pp. 41–49.

[88] C. Holt, *Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages.* ONR Research Memorandum 52, Carnegie Institute of Technology, Pittsburgh, Pennsylvania, 1957.

[89] R. Brown, *Smoothing, Forecasting and Prediction.* Englewood Cliffs: Prentice Hall, 1963.

[90] P. Winters, *Forecasting Sales by Exponentially Weighted Moving Averages.* Management Science 6, 324-42, 1960.

[91] D. Wright, *Forecasting data published at irregular time intervals using an extension of HoltŠs method.* Management Science, 32, 499-510, 1986a.

[92] D. Wright, *Forecasting irregularly spaced data: An extension of double exponential smoothing.* Computers and Industrial Engineering, 10, 135-147, 1986b.

[93] G. Janacek and L. Swift, *Time series Forecasting, Simulation, Applications.* Ellis Horwood, Great Britain, 1993.

[94] T. Rappaport, *Wireless communications Principles and practice,3rd Ed.* New Jersey: Prentice Hall publication, 1996.

[95] Y. Shi and M. Mizumoto, "A learning algorithm for tuning fuzzy inference rules," *Fuzzy Systems Conference Proceedings, 1999. FUZZ-IEEE '99. 1999 IEEE International*, vol. 1, pp. 378 – 382, Aug. 1999.

[96] F. Herrera and M. Lozano, "Fuzzy adaptive genetic algorithms: design, taxonomy, and future directions," *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, vol. 7, no. 8, pp. 545–562, 2003.

[97] M. Maeda, M. Oda, and H. Miyajima, "Reduction methods of fuzzy inference rules with neural network learning algorithm," in *Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on*, vol. 5, 1999, pp. 262–267 vol.5.

[98] Y.-P. Huang and H.-J. Chen, "The extension-based fuzzy modeling method and its applications," in *Electrical and Computer Engineering, 1999 IEEE Canadian Conference on*, vol. 2, 1999, pp. 977–982 vol.2.

[99] S. Fukumoto, H. Miyajima, K. Kishida, and Y. Nagasawa, "A destructive learning method of fuzzy inference rules," in *Fuzzy Systems, 1995. International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium., Proceedings of 1995 IEEE International Conference on*, vol. 2, 1995, pp. 687–694 vol.2.

[100] I. Requena, A. Blanco, and M. Delgado, "A constructive method for building fuzzy rule-based systems," *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, vol. 11, no. 2, pp. 217–233, 2003.

[101] C.-C. Wong, B.-C. Lin, and C.-T. Cheng, "Fuzzy tracking method with a switching Grey prediction for mobile robot," in *Fuzzy Systems, 2001. The 10th IEEE International Conference on*, vol. 1, 2001, pp. 103–106.

[102] L. Zadeh, "Fuzzy sets versus probability," *Proceedings of the IEEE*, vol. 68, no. 3, p. 421, 1980.

[103] L. Zadeh, "Fuzzy logic," *Computer*, vol. 21, no. 4, pp. 83–93, 1988.

[104] T. Perry, "Lotfi a. zadeh [fuzzy logic inventor biography]," *Spectrum, IEEE*, vol. 32, no. 6, pp. 32–35, 1995.

[105] M. Sugeno, "On stability of fuzzy systems expressed by fuzzy rules with singleton consequents," *Fuzzy Systems, IEEE Transactions on*, vol. 7, no. 2, pp. 201–224, 1999.

[106] R. C. Berkan and S. l. Trubatch, *Fuzzy systems design principles.* IEEE Press, New York, 1997.

[107] Y. Jin, J. Jiang, and J. Zhu, "Neural network based fuzzy identification and its application to modeling and control of complex systems," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 25, no. 6, pp. 990–997, 1995.

[108] T. Krink, J. Vesterstrom, and J. Riget, "Particle swarm optimisation with spatial particle extension," in *Evolutionary Computation, 2002. CEC '02. Proceedings of the 2002 Congress on*, vol. 2, 2002, pp. 1474–1479.

[109] T. Krink and R. Thomsen, "Self-organized criticality and mass extinction in evolutionary algorithms," in *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on*, vol. 2, 2001, pp. 1155–1161 vol. 2.

[110] M. Lovbjerg and T. Krink, "Extending particle swarm optimisers with self-organized criticality," in *Evolutionary Computation, 2002. CEC '02. Proceedings of the 2002 Congress on*, vol. 2, 2002, pp. 1588–1593.

[111] R. Eberhart and Y. Shi, "Guest editorial special issue on particle swarm optimization," *Evolutionary Computation, IEEE Transactions on*, vol. 8, no. 3, pp. 201–203, 2004.

[112] J. James Kennedy and E. R.C., *Swarm Intelligence.* san Francisco,CA: Morgan Kaufmann Publishers, 2001.

[113] Y. Shi, R. Eberhart, and Y. Chen, "Implementation of evolutionary fuzzy systems," *Fuzzy Systems, IEEE Transactions on*, vol. 7, no. 2, pp. 109–119, 1999.

[114] Y. Shi and R. Eberhart, "Fuzzy adaptive particle swarm optimization," in *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on*, vol. 1, 2001, pp. 101–106 vol. 1.

[115] R. Eberhart and Y. Shi, "Tracking and optimizing dynamic systems with particle swarms," in *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on*, vol. 1, 2001, pp. 94–100 vol. 1.

[116] S. Sheu and C. Wu, "Using Grey prediction theory to reduce handoff overhead in cellular communication systems," *Personal, Indoor and Mobile Radio Communications, 2000. PIMRC 2000. The 11th IEEE International Symposium on*, vol. 2, no. 6, p. 782 Ű786, 2000.

[117] M. Maeda and H. Miyajima, "Constructive methods of fuzzy rules for function approximation," in *The 2002 International Technical Conference On Circuits/Systems,Computers and Communications, Phuket, Thailand*, 2002.

[118] A. Arslan and M. Kaya, "Determination of fuzzy logic membership functions using genetic algorithms," *Fuzzy Sets and Systems*, vol. 118, no. 2, pp. 297–306, 2001.

[119] C. Karr and E. Gentry, "Fuzzy control of pH using genetic algorithms," *Fuzzy Systems, IEEE Transactions on*, vol. 1, no. 1, p. 46, 1993.

[120] C. Karr and S. Sharma, "An adaptive process control system based on fuzzy logic and genetic algorithms," in *American Control Conference, 1994*, vol. 3, 1994, pp. 2470–2474 vol.3.

[121] K. C. L. Meredith D. L. and K. K., "The use of genetic algorithms in the design of fuzzy logic controllers." in *In Proc. Third Workshop on Neural Networks: Academic/Industrial/ Defence (WNN'92)*, vol. 1721, February 1992, Auburn, pp. 549–555.

[122] M. Lee and H. Takagi, "Integrating design stage of fuzzy systems using genetic algorithms," in *Fuzzy Systems, 1993., Second IEEE International Conference on*, 1993, pp. 612–617 vol.1.

[123] S. J. Go, M. C. Lee, and M. K. Park, "Fuzzy-sliding mode control of automatic polishing robot system with the self tuning fuzzy inference based on genetic algorithm," in *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, vol. 3, 2001, pp. 2962–2967 vol.3.

[124] C.-C. Kung and W.-C. Lai, "GA - based design of a region-wise fuzzy sliding mode controller," in *Electrical and Computer Engineering, 1999 IEEE Canadian Conference on*, vol. 2, 1999, pp. 971–976 vol.2.

[125] Y.-P. Huang, C.-C. Huang, and C.-H. Huang, "Optimizing the fuzzy adaptive learning by the gradient descent approach," in *Fuzzy Systems, 1994. IEEE World Congress on Computational Intelligence., Proceedings of the Third IEEE Conference on*, 1994, pp. 701–705 vol.1.

[126] C.-L. Hwang, "A novel takagi-sugeno-based robust adaptive fuzzy sliding-mode controller," *Fuzzy Systems, IEEE Transactions on*, vol. 12, no. 5, pp. 676–687, 2004.

[127] F. Moreno-Velo, I. Baturone, R. Senhadji, and S. Sanchez-Solano, "Tuning complex fuzzy systems by supervised learning algorithms," in *Fuzzy Systems, 2003. FUZZ '03. The 12th IEEE International Conference on*, vol. 1, 2003, pp. 226–231 vol.1.

[128] F. Guely and P. Siarry, “Gradient descent method for optimizing various fuzzy rule bases,” in *Fuzzy Systems, 1993., Second IEEE International Conference on*, 1993, pp. 1241–1246 vol.2.

[129] K. F. K. F. Man, *Genetic algorithms : concepts and designs.* 1951- Date: London ;New York :Springer,c1999, 1999.

[130] H. Eriksson, “MBONE: the multicast backbone,” *Commun. ACM*, vol. 37, no. 8, pp. 54–60., 1994.

[131] A. Acharya, A. Bakre, and B. Badrinath, “IP multicast extensions for mobile internetworking,” vol. 1, pp. 67–74, 1996.

[132] Y.-J. Suh, H.-S. Shin, and D.-H. Kwon, “An efficient multicast routing protocol in wireless mobile networks,” *Wirel. Netw.*, vol. 7, no. 5, pp. 443–453, 2001.

[133] C. Sankaran and A. Ephremides, “Multicasting with multiuser detection in ad-hoc wireless networks,” in *Broadband Communications, 2000. Proceedings. 2000 International Zurich Seminar on*, pp. 47–54.

[134] R. Sedgewick, *Algorithms in C++: part 5 - Graph Algorithms*, 3rd ed. Indiana Polis: Addison-Wesley, 2001.

[135] M. Kodialam, T. V. Lakshman, and S. Sengupta, “Online multicast routing with bandwidth guarantees: a new approach using multicast network flow,” *IEEE/ACM Trans. Netw.*, vol. 11, no. 4, pp. 676–686, 2003.

[136] G. N. Rouskas and I. Baldine, “Multicast routing with end-to-end delay and delay variation constraints,” Raleigh, NC, USA, 1995.

[137] G. M. and J. D, “Computers and intractabilty: A guide to the theory of NP-completeness,” NY, USA, 1979.

[138] H. Taha, *Operational Research: An Introduction*, 6th ed. New Jersey: Prentice Hall International, 1997.

[139] A. Roy and S. K. Das, "Optimizing QoS-based multicast routing in wireless networks: A multi-objective genetic algorithmic approach," in *Proceedings of the Second International IFIP-TC6 Networking Conference on Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; and Mobile and Wireless Communications.* Springer-Verlag, 2002, pp. 28–48.

[140] M. Parsa, Q. Zhu, and J. J. Garcia-Luna-Aceves, "An iterative algorithm for delay-constrained minimum-cost multicasting," *IEEE/ACM Trans. Netw.*, vol. 6, no. 4, pp. 461–474, 1998.

[141] V. P. Kompella, J. C. Pasquale, and G. C. Polyzos, "Multicast routing for multimedia communication," *IEEE/ACM Trans. Netw.*, vol. 1, no. 3, pp. 286–292, 1993.

[142] S. Ramanathan, "Multicast tree generation in networks with asymmetric links," *IEEE/ACM Trans. Netw.*, vol. 4, no. 4, pp. 558–568, 1996.

[143] S. Kapoor and S. Raghavan, "Improved multicast routing with delay and delay variation constraints," in *Global Telecommunications Conference, 2000. GLOBECOM '00. IEEE*, vol. 1, 2000, pp. 476–480 vol.1.

[144] K. Bharath-Kumar and J. Jaffe, "Routing to multiple destinations in computer networks," *Communications, IEEE Transactions on [legacy, pre - 1988]*, vol. 31, no. 3, pp. 343–351, 1983.

[145] G. M.-L. Kou, L., "A fast algorithm for steiner trees," in *Acta Informatica.* Springer-Verlag, 1981, pp. 141–145.

[146] "OPNET Inc." [Online]. Available: http://www.opnet.com

[147] "OMNET++ discrete event simulation system." [Online]. Available: http://www.omnetpp.org/

[148] S. McCanne and S. Floyd, "ns-lbnl network simulator." [Online]. Available: http://www-nrg.ee.lbl.gov/ns/

[149] "NS-2 simulator." [Online]. Available: http://www.isi.edu/ns/nsnam

[150] "NCTUns 1.0 network simulator." [Online]. Available: http://nsl10.csie.nctu.edu.tw/

[151] N. Moore, "Non-predictive handovers for IPv6 (mobileip WG)." IETF draft, March 2004. [Online]. Available: http://www.ctie.monash.edu.au/ipv6/fastho/no_predict.pdf/

[152] G. Liu and C. Zhou, "HCRAS: A novel hybrid internetworking architecture between WLAN and UMTS cellular networks," in *Consumer Communications and Networking Conference, 2005. CCNC. 2005 Second IEEE*, 2005, pp. 374–379.

[153] M. Bernaschi, F. Cacace, and G. Iannello, "Vertical handoff performance in heterogeneous networks," in *Parallel Processing Workshops, 2004. ICPP 2004 Workshops. Proceedings. 2004 International Conference on*, 2004, pp. 100–107.

[154] Y. Sun, W. Trappe, and K. Liu, "A scalable multicast key management scheme for heterogeneous wireless networks," *Networking, IEEE/ACM Transactions on*, vol. 12, no. 4, pp. 653–666, 2004.

[155] "Document object model." [Online]. Available: http://www.w3.org/DOM/

Abbreviations

1G	First Generation
2G	Second Generation
3G	Third Generation
4G	Fourth Generation
AP	Access Point
AMPS	Advanced Mobile Phone System
AR	Access Router
BSC	Base Station Controller
BSMA	Bounded Shortest Multicast Algorithm
CBT	Core Based Tree Protocol
CAR	Coverage Access Router
CIP	Cellular IP
CN	Corresponding Node
DVMA	Delay Variation Multicast Algorithm
EMAF	Exponential Moving Average Filters
DVMRP	Distance Vector Multicast Routing Protocol
DVBMT	Delay Variation Bound Multicast Tree
FDMA	Frequency Division Multiplexing
FLC	Fuzzy Logic Controller
GA	Genetic Algorithm

GDI	Graphics Device Interchange
GPS	Global Positioning System
GSM	Global System for Mobile Communication
HAWAII	Handoff Aware Wireless Access Internet Infrastructure
HMIP	Hierarchical Mobile IP
IGMP	Internet Group Management Protocol
IP	Internet Protocol
IMTS	Improved Mobile Telephone Service
MAHO	Mobile Assisted Handoff
MCHO	Mobile Controlled Handoff
MBone	Multicast Backbone
MMP	Multicast Mobility Prediction
MBWDC	Mobility BandWidth Delay Constrained Algorithm
MIP	Mobile IP
MIPv6	Mobile IP version 6
MOSPF	Multicast extensions to Open Shortest Path First
MOM	MObile Multicast
MAF	Moving Average Filters
MMA	Mobile Multicast
MN	Mobile Node
MSC	Mobile Switching Centre
NP	Non-Polynomial
NCHO	Network Controlled Handoff
PACS	Personal Access Communication Systems
PIM	Protocol Independent Multicast
PSO	Particle Swarm Optimisation

QoS	Quality of Service
RBMOM	Range Based MObile Multicast
RFC	Request for Comments
RSSI	Received Signal Strength Indicator
SGA	Simple GA
TCP	Transmission Control Protocol
TDMA	Time Division Multiplexing
UMTS	Universal Mobile Telecommunications System
XML	Extensible Markup language
WLAN	Wireless LAN
WWW	World Wide Web

VITA

Suresh Venkatachalaiah received the B. Eng. in Instrumentation and electronics engineering from Bangalore Univeristy, Bangalore, India in 1997. He did his M. Eng. in Information Technology from RMIT university. During his Ph.D. candidature at the School of Electrical and Computer Engineering, RMIT University, he also wrote for various research grants and achieved a grant from Microsoft Corporation Inc., Redmond. His research interests are in layer-2 and layer-3 protocols for wireless networks, unicast and multicast routing protocols for wire-line/wireless IP-networks, wireless mobile ad-hoc networks, sensor networks, mesh networks, network detection and address auto-configuration in IPv6 networks.