Seamless Application Handover Across Radio Access Networks (SAHARA Net)

Taleb Rabah Benouaer

A dissertation submitted in partial fulfilment

of the requirements for the degree of

Master of Philosophy

of the

University of London

Department of Electronic and Electrical Engineering

University College London

Supervisor: Prof. George Pavlou

London, September 21, 2010

Declaration of Authorship and Originality

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Department of Electronic & Electrical Engineering
University College London
Torrington Place
WC1E 7JE London
UK
Email: taleb@ee.ucl.ac.uk

Date: September 21, 2010

Taleb Rabah Benouaer

Abstract

The explosion in mobile and data traffic in the last decade has led to a rapid proliferation in wireless networks. A plethora of wireless access technologies are available today each with a different offering. Some offer high data rates within a restricted coverage area such as 802.11 hotspots. Others, offer lower data rates but with a much wider coverage such as UMTS. This diversity can be harnessed in a way that creates a ubiquitous communications platform for the user. This is the premise of the heterogeneous networks vision/architecture: an environment where disparate technologies cooperate together and complement each other. However, there are various technical challenges in the way of such convergence. The first obstacle is enabling communication between disparate mobility protocols. Once this is achieved, the diversity of networks in itself poses a challenge for the user as to which network he connects to.

This thesis answers the first question by reviewing the low-latency handover literature to identify the most credible solutions. The general consensus amongst researchers in the field has been to bridge the gap between the network and link layers so that IP protocols can react quickly to link changes. To answer the second question, this thesis defines a framework to assess handover decision algorithms based on application performance. The merit of the handover algorithm's decision is measured by how well the application performs after handover. In order to facilitate this process, a simulation module was created within the NS2 network simulator that allows mobile devices to collect network measurements and feed that information into a decision algorithm to decide whether or not handover should be triggered.

Through this evaluation process, a number of issues emerged as possible stumbling blocks. The first such issue is the inconsistency between local network conditions measured at the Access Point or Base Station, and the end to end conditions experienced by the user's application. Another issue is the algorithm's adaptability to user and application preferences. Personal users might be cost aware opting to trade off quality for a lower cost in certain circumstances. The handover algorithm must be able to accommodate such scenarios. Furthermore, algorithms must be able to adapt their decisions according to the application's requirements. Using application profiles with thresholds or utility functions can result in better decisions than using absolute values. If an application is satisfied with the current network conditions, it might not be in its benefit to move to a different network even if it offered better conditions. In fact, it might suffer as a result of possible handover disruptions.

Acknowledgements

First, I would like to thank my supervisor Professor George Pavlou for his valuable guidance and support. I would also like to extend my thanks to Prof. Izzat Darwazeh for his continued assistance and support. Many thanks go to Dr. Wei Koong Chai for his help and effort.

My sincere gratitude goes to all my friends and colleagues, especially Mussie Woldeselassie and Redouane Ali, whose presence eased the burden of this challenging experience. Their contributions, both personal and technical, were invaluable in the realization of this project.

Last but not least, I am ever indebted to my parents, brothers and sisters who have been the burning fuel that drives me in life. Their unconditional love and support, their prayers and continuous encouragement have always given me hope and strength. But above all, I would like to thank God who blessed me with health and gave me the patience to persevere through this endeavour, all Praise is due to him.

- **This thesis was completed under the financial sponsorship of the Algerian Government.**

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The proliferation of mobile devices and the substantial diversification in mobile applications in recent years meant that users are faced with a large choice of technologies for their networking needs. These include, among many others, GSM, UMTS, 802.11 and Bluetooth. If users are to have the flexibility of always having the best connection for their application, vertical handovers across heterogeneous networks become a necessity. However, this poses a number of challenges. The first challenge is network selection: how a user chooses which network to use for a particular application. This is not straightforward and needs to take into account a several factors. For example, a user may be connected to an 802.11 Access Point (AP) for a large file transfer but his device's battery level decreases to a point that does not permit the completion of the transfer. The user may then want to switch to a technology that requires less transmission power but might offer lower data rates. In such a scenario, the choice of the network has to take into consideration the various parameters at play and produce an optimal (or sub-optimal) utility.

There have been many proposals that deal with handover decision in heterogeneous networks, using various approaches such as Multi-Attribute Decision Making (MADM) [67], Markov Decision processes (MDP) [61] and Fuzzy Logic processes [50]. Regardless of what technique is used, the goal is to maximize some utility function, of the network attributes, that tries to best accommodate the requirements of the application. This often leads to relaxing some constraints for a particular gain. In the example mentioned above, the decision process may chose to handover to a lower power network, compromising on the application's throughput requirements to save power. In fact this problem is not unique to vertical handover, but applies to a variety of situations where a decision has to be made on conflicting criteria. Examples may be found in economics [58], [57], social sciences [68] or biology [53].

Handing over between various networks is, in itself, a complex process. Issues related to latency, signalling and computational costs must all be addressed in order to establish whether or not handing over to another network is a sound decision. The complexity is increased even further when aspects related to pricing and subscription rights are considered. In all, it is desirable to have the flexibility to move between networks in order to maximize the quality of service but it is also important to study the effects and the costs associated with such schemes.

1.2 Contributions

Having established the complexity of vertical handover decisions, it is important to evaluate the effectiveness of decision algorithms in selecting the appropriate network for the user's needs . There are numerous approaches proposed in the literature that tackle this issue, however, due to the their diversity, it is difficult to compare the performance of one approach against the other. This work's main contribution lies in proposing a common evaluation framework than can be used to compare and asses decision algorithms. The key elements of this contribution are summarized as follows:

1. A framework for evaluating handover decision algorithms is defined, based on application performance. A number of key metrics can be used to determine the performance of the application after each handover. The algorithm's merit is assessed based on whether or not (and how much) the application is deemed to have benefited from the handover decision.

2. A set of simulation tools that facilitate the evaluation process are built. These consist of a measurement module that gathers the necessary network information and a decision module that evaluates the various attributes to reach a conclusion. The modular structure of the decision modules means new algorithms can be easily added alongside existing ones without the need for modifications to the main structures of the tool.

3. Some of the difficulties in making the right handover decisions are identified. First, the difficulty in judging end-to-end network conditions through measurements made at Access Point or Base Station level, makes it difficult to predict the application performance after handover. Second, handover algorithms must be able to trade off cost and quality in accordance with user preferences.

4. Some enhancements to improve the handover decision process are proposed. These include using application profiles or thresholds and utilities rather than absolute metric values to evaluate network parameters. In addition, dynamically adapting the decision algorithm according to the application's requirements, so that the most suitable network for the current application is selected.

1.3 Publications

1. T.R. Benouaer and J.K. Pollard, ”Seamless Adaptive Handover Across Radio Access Networks”, European Modelling Symposium (EMS), London, 2006.

1.4 Report Structure

The remainder of this document is structured as follows:

The next chapter provides background information about regarding mobility management systems. Chapter 3 provides further details presenting the various research efforts aiming to reduce mobile IP handover latency, as well the various vertical handover decision schemes. This is followed by two chapters on handover decision algorithms. The first one presents an evaluation of a number of decision algorithms, and the other comparing the decision pattern of those algorithms. The last chapter concludes the thesis by summarizing the main findings and contributions.

Chapter 2

Background

2.1 Heterogeneous Wireless Networks

Heterogeneous networks are a set of dissimilar wireless technologies that coexist and cooperate to provide uninterrupted connectivity to the user. Control information is exchanged between adjacent networks to maintain connectivity as the user moves from one access network to the other. Heterogeneous networks provide an inter-working platform where different wireless technologies interoperate to extend their network coverage. Several wireless technologies exist today each with its own service offerings. The heterogeneous networks framework aims to exploit this diversity to create a ubiquitous communication system encompassing all the different technologies.

Wireless access technologies fall into one of two categories: packet-switched Internet technologies or circuit-switched cellular technologies. Circuit-switched technologies were developed to provide voice services to mobile users. Early examples of these systems include the global system for mobile communications (GSM). They have the advantage of wide area coverage but have very limited data rates, up to 9.6 kbps for GSM [69]. Packet-switched technologies such as the IEEE 802.11x standard family were developed to provide wireless access to data networks. They have higher data rates, up to 54Mbps for 802.11 [34], but with very restricted mobility within a relatively small area compared to cellular technologies.

Both circuit-switched and packet-switched systems enjoyed great success in their respective domains. Circuit-switched systems benefited from the worldwide proliferation of mobile voice communications. Packet-switched systems profited from the explosive growth of Internet data services. This enormous success in these two sectors fueled a new drive to offer Internet data services on the move. This posed major challenges for both systems. Circuit-switched systems had very limited data rates to support Internet data services. Packet-switched systems did not have adequate mobility support to manage application mobility. Both systems needed to evolve to meet these new challenges.

A host of new technologies were introduced to cater for the emerging mobile data services. The general packet radio system (GPRS) was the first one to emerge in the cellular domain. It was introduced as an upgrade into the GSM architecture and offered packet data services at rates in the region of 144 kbps[20]. Further enhancements to the GPRS system lead to the advent of the universal mobile telecommunications system (UMTS)[26]. The UMTS technology offered both circuit-switched and packet-switched services at data rates of up to 2Mbps. This was still far less than the data rates available through 802.11 and other packet-switched technologies.

Packet-switched technologies offered high data rates but lacked adequate mobility support. There have been various proposals to incorporate mobility into 802.11i and increase its transmission range [42, 16, 75]. Nevertheless, the predominant deployment strategy has been in the form of scattered service islands or hotspots within public spaces. Only users within range of these hotspots are able to access their data applications. New emerging packet-switched technologies offer better reach and mobility support. These include 802.16 [33](branded WiMax) and 802.20 [8] (branded mobile-Fi) with coverage areas of up to 30 km and 20 km respectively. These two technologies are based on cellular architectures and are expected to rival traditional cellular technologies such as UMTS.

The distinctions between cellular and packet-switched networks are becoming increasingly vague. Packet-switched networks have adopted many cellular techniques and vice versa. This reflects the service convergence trend between voice and data applications. Voice and data are bundled together in a variety of services such as Microsoft's instant messaging service and online ”ring back” services. Offering such services through an integrated voice and data network is more cost-effective than two separate networks. Operational and maintenance costs are consolidated resulting in major savings in expenditure. The integration drive is also an indication of the success of the Internet Protocol (IP) as an inter-working platform between networking technologies. This can be seen in the ubiquity of the Internet, which is available through a wide range of access media.

Most new radio access technologies, including UMTS and WiMax, integrate IP into their protocol stack to facilitate access to Internet data applications. This creates an opportunity for interoperation between IP-enabled networks. The heterogeneous networks architecture exploits this interoperation to create an integrated communications platform using IP as a common network layer (layer 3 of the Open Systems Interconnection- OSI- networking model). This allows communications sessions to be maintained uninterrupted across network domains. A set of IP-based networking protocols are used to enable the transfer of control information across disparate access networks. The purpose of these protocols is to facilitate interoperation between wireless technologies. They implement the functions required to maintain cross-network communication sessions independently of the underlying access network. Such functions include managing application mobility, handover between networks, and maintenance of QoS and security settings.

2.2 Mobility in Heterogeneous Wireless Networks

Mobility management is one of the most vital elements of the heterogeneous networks architecture. The mobility management system encompasses all the network components and protocols required to maintain the user's (application) connectivity to the network on the move. It is mainly concerned with two issues: location tracking and session handover.

2.2.1 Location Tracking

The location-tracking element is responsible for maintaining an up-to-date record of the user's (application) current location. The session-handover element is responsible for transferring live communication sessions between network cells or domains.

Tracking the location of the user requires the network to hold a static and a dynamic record of his current location. The static record is the first point of reference to which any queries about the user's location are directed. It is a permanent record that stores the user's dynamic location and is updated whenever that location changes. The dynamic record holds a temporary log of the user's details at his current point of attachment to the network. It is created when the user first registers his details with a new network cell (domain) and is removed as soon as he moves away from it.

2.2.2 Session Handover

The handover process constitutes a major part of the mobility management system. It extends the reach of the access network and ensures uninterrupted communications for mobile applications. As the user moves out of range of a network cell (domain), he is seamlessly handed-over to the adjacent cell (domain) without interrupting the progress of his application. This is achieved through a continuous signal monitoring and assessment process. A handover is triggered when the quality of the received signal deteriorates below a specific threshold. The handover process examines neighbouring network access points and identifies one with adequate signal quality. A connection is established to the selected access point and the user's communications session is seamlessly handed over to it.

2.2.3 Heterogeneous Networks

Managing mobility across heterogeneous networks introduces new challenges. One of the major challenges is the lack of interoperation between the mobility management systems of different wireless technologies. Mobility support has traditionally been implemented as part of the radio access system itself (in the link layer or layer 2 of the OSI model). As a result, signaling protocols and control information are only functional within access networks implementing the same technology (homogeneous networks). Providing mobility support at the network layer, using IP, averts this problem and allows the transfer of control information across heterogeneous networks. Location tracking and handover functions are implemented using IP-based protocols to enable interoperable and seamless mobility across dissimilar radio access networks [4].

However, mobility support in IP is very limited. Network addresses in IP are hierarchical addresses that are associated with a specific subnet within a domain. They define a specific network point and are not mobile. When a user changes his network attachment point, he is allocated a new IP address. In addition, two IP addresses are required for every user to enable the network to track their location: a static permanent IP address and a dynamic temporary one. Allocating IP addresses dynamically can be achieved through a number of IP-based protocols. However, there must be an association between the dynamic and the static IP addresses to enable packets to be delivered to the user. Furthermore, IP is a layer 3 technology and does not have access to information regarding the link status. Hence, IP cannot promptly determine if the user has moved out of the current cell (domain) [6].

Prompt access to link status information is more critical for session handover. Delays in detecting the user's movements slow down the handover process. This may affect the Quality of Service (QoS) of the mobile application or lead to a temporary loss of connectivity. In addition, handover decisions in heterogeneous networks can be triggered by a variety of parameters. These might include link bandwidth, end-to-end delays, service cost, and load balancing. Furthermore, handover across heterogeneous networks (intersystem handover or vertical handover) requires the user's mobile device to be aware of the different wireless networks available within range. This imposes further constraints on the mobile device's power resources. The handover process must adopt power-saving mechanisms to minimize these constraints.

Mobility management across heterogeneous networks requires interworking protocols that interoperate across different wireless technologies. Implementing mobility using IP-based protocols enables interoperable operation across dissimilar access networks. Several IP-based mobility management schemes have been developed. These fall into two main categories: macro-mobility, such as mobile IP, and micro-mobility management systems.

2.3 Mobile IP

Macro-mobility management systems enable user reachability across several network domains. They maintain an updated record of the users current location to ensure incoming packets can be delivered to it. Micro-mobility is concerned with the movements of mobile nodes within a small area designated by a network cell/domain Mobile IP provides macro-mobility support in IP networks. It was developed to enable mobile nodes to remain reachable regardless of their point of attachment to the network. The mobile IP protocol architecture consists of the following components:

Mobile Node (MN) :

An IP node that is able to maintain its IP address while changing its point of attachment to the network.

Correspondent Node (CN) :

An IP node that is communicating with a mobile node (MN).

Home agent (HA) :

An IP router that is able to provide mobility services to mobile nodes belonging to its network.

Foreign agent/Access Router (AR) :

An IP router that is able to provide mobility services to visiting MNs. In mobile IPv6, ordinary access routers (ARs) are capable of providing foreign agent (FA) functionality.

Mobile IP defines two IP addresses for the MN: a home address and a Care of Address (CoA). The home address is a permanent address that remains unchanged when the MN changes its point of attachment to the network. The leading bits of the MNs home address (network prefix) define its home network. The CoA is a temporary address that defines the MNs current point of attachment to the network. It is assigned to the MN while away from its home network. Mobile IP tracks the location of the MN by maintaining a binding between the two addresses at the MNs home agent. Packets destined to the MN are intercepted by its home agent and routed towards its current location. The home address is used to route the packets from the sender to the MNs home address. The CoA is then used to route those packets from the home network to their destination at the MNs current location. This binding can also be used by correspondent nodes (CNs) to allow them to deliver packets directly to the MN. The packet delivery procedure in mobile IP is illustrated in Figure 2.1 below.

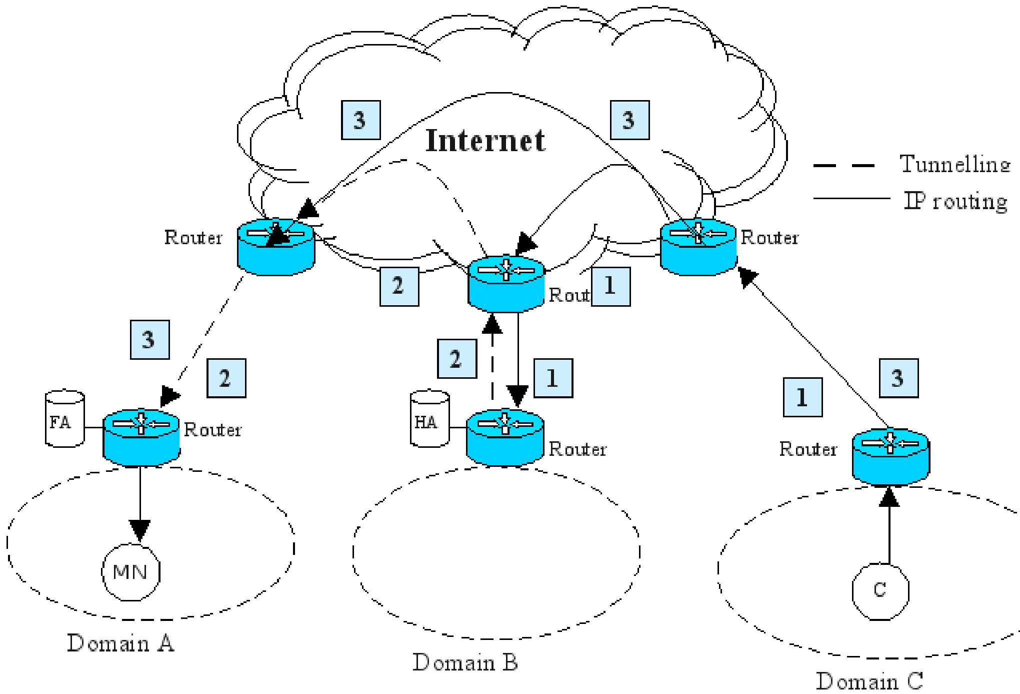

Figure 2.1: Mobile IP architecture

1. packets addressed to the MN are sent using its home address and are routed towards its home network.
2. at the home network (domain B), the home agent intercepts the packets, encapsulates them within a new IP packet destined towards the CoA. The packets are tunneled towards the FA which decapsulates them and routes them to the MN.

3. mobile IPv6 allows CNs to bypass the HA and send packets directly to the MN using the route optimization procedure. CNs implementing route optimization maintain a binding of the MNs home and CoA addresses, and use that information to sends packets directly to the MNs CoA.

Mobile IP Handover

The handover procedure in mobile IP is performed in three stages: Movement detection, FA registration, home registration.

1. **Movement detection** :

   Movement detection in mobile IP relies on the router advertisement messages (router adverts) sent by FAs to announce their presence. Mobile IP proposes two movement detection algorithms based on information contained within router adverts. The first algorithm inspects the lifetime of the router advert. A movement is detected if the lifetime of the router advert has expired. The second algorithm uses the network prefix field in the router advert to detect movements. A movement is detected if a router advert with a different network prefix is received.

2. **FA registration** :

   When the MN detects that it has moved to a new location, it initiates the neighbour discovery procedure to register with a new FA and obtain a new CoA.

   (a) The MN attaches to the link and verifies the uniqueness of its link local address by sending a Duplicate Address Detection (DAD) solicitation to its neighbours. If no response is received within the DAD timeout, the address is considered unique.

   (b) The MN sends a router solicitation to discover neighbouring routers.

   (c) When the MN receives a router advert, it forms a new CoA using the New Access Routers (NAR) network Prefix and its own Interface ID, and perform a DAD check on it

   (d) The completion of the DAD check indicates the end of the FA/NAR registration.

3. **Home registration** :

   The MN performs the home registration procedure to update its home agent binding with its new CoA. The MN sends a binding update message to its home agent, containing the MNs new CoA. The home agent replies with a binding acknowledgment. This concludes the handover procedure and subsequent packets destined to the MN are sent to the new CoA.

The procedures described above introduce lengthy delays that cannot be tolerated by real-time applications ,making mobile IP unsuitable as a seamless mobility solution. Two factors contribute to the handover delays. The movement detection procedure does not detect the handover until the old connection is lost. This leads to a disconnection period where the MN cannot receive or send packets. Connectivity is only restored after the MN completes the home registration. Hence, the length of this disconnection period would depend on the end-to-end delays between the MN and its home agent.

In the micro-domain, macro-mobility solutions such as mobile IP cannot respond to the speed and frequency of movements. The signaling overheads would escalate because of the frequency of the movements. The MN would suffer delays during each movement because of the home registration procedure. Micro-mobility management systems avert these delays by localizing the signaling associated with movements in the micro-domain. They introduce regional nodes, which manage mobility within the local domain and use mobile IP to provide global reachability. Several micro-mobility solutions have been proposed. A brief description of their protocol architecture and operation is given below.

2.3.1 Mobile IP Regional Registration (MIPRR)

The Mobile IP regional registration (MIPRR) protocol [3, 10, 17] is a variant of mobile IP, which introduces a new hierarchical approach (Hierarchical mobile IP (HMIP) is the equivalent protocol within the IPv6 architecture. A new node is introduced called the Gateway Foreign Agent (GFA). Each GFA designates a regional network and holds records of all MNs within that network (the equivalent node within the HMIP architecture is called the mobility anchor point (MAP). A number of FAs exist within a regional network and are all connected to the GFA. The architecture of the MIPRR protocol is illustrated in figure 2.2 below. The hierarchical architecture of the MIPRR protocol allows it to localize handover signaling messages. This reduces registration delays caused by lengthy round trip times to the HA. However, this only applies to movements within a GFA domain. When the MN moves to another domain, it is required to register its new regional address with its HA. In this scenario, the MN does not

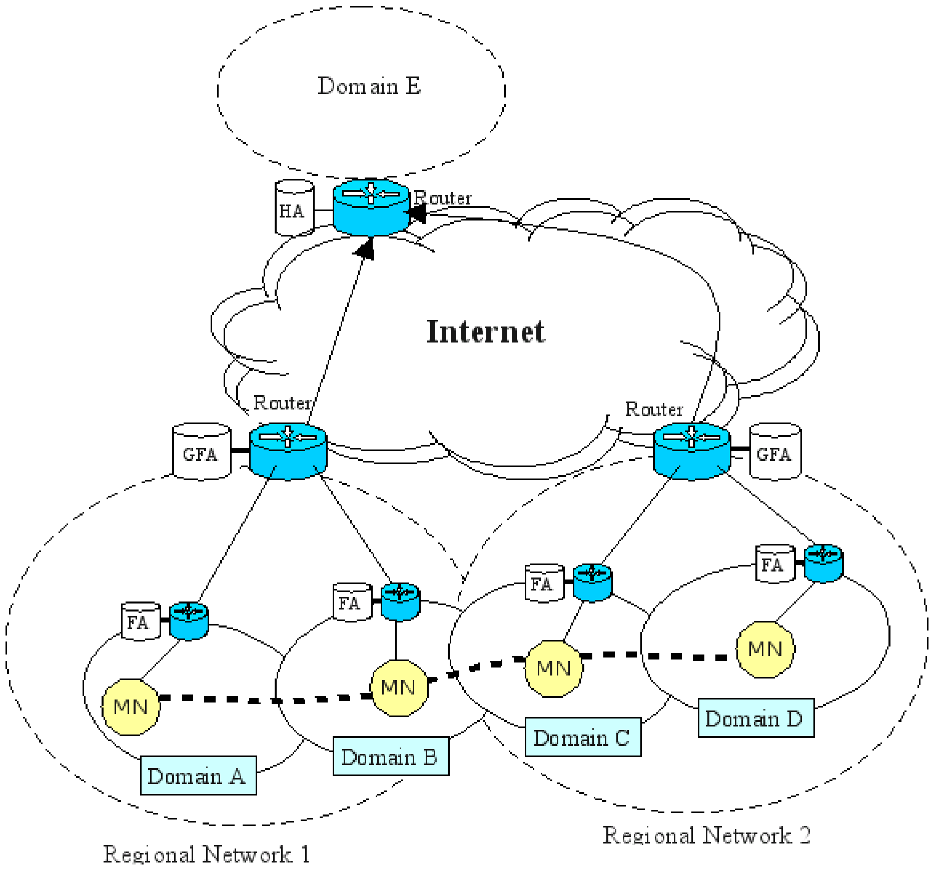

Figure 2.2: Mobile IP Regional Registration Protocol Architecture

benefit from MIPRRs localized approach and will suffer the same home registration delay as it did with mobile IP. Furthermore, it will incur the additional overhead of regional registration with the GFA. As a result, the MIPRR protocol cannot be considered as a viable solution for inter-domain mobility and its benefits are restricted to intra-domain movements.

2.3.2 Intra-Domain Mobility Management Protocol (IDMP)

The Intra-Domain Mobility Management Protocol (IDMP) [3, 10, 45, 18] employs a hierarchical approach to provide mobility support within network domains. It is based on a two-level hierarchy characterized by two classes of agents: the Subnet Agent (SA) and the Mobility Agent (MA). The SA handles mobility inside the subnet, whereas the MA handles mobility across subnets within the domain. Each MN within the IDMP network is allocated two CoAs: a Gateway CoA (GCoA) and a Local CoA (LCoA). The GCoA identifies the MA and the LCoA identifies the SA serving the MN.

Fast handover :

Handover occurs when the MN moves to a new subnet and registers with a new SA.

While the MN is establishing a connection to the new SA, the MA is unaware that it has changed its serving SA and continues forwarding packets to the old SA. Packets arriving at the MA during this period are delivered to the old SA, and are consequently lost. A fast handover procedure is proposed to address this problem. This assumes the MN is capable of anticipating impending handovers, either by monitoring signal power levels or listening to beacon signals from the SA. If the MN identifies an imminent handover, it requests the MA to multicast packets to all neighbouring SAs, until it notifies it of its new LCoA. Neighbouring SAs buffer the received packets until the MN is able to register with its new SA. The new SA then forwards the buffered packets to the corresponding MN. This procedure is called fast handover because it reduces the amount of time during which the MN is incapable of receiving incoming packets. The MN does not have to wait for the completion of the registration process to receive incoming packets. Instead, as soon as the MN registers with its new SA, the SA delivers the buffered packets to it.

Paging support :

The paging procedure is used to enable idle MNs to save power by reducing the level of location registration/update required from them. The IDMP system uses a multicast procedure similar to that used in fast handover to provide paging. Paging areas are identified by unique identifiers, which are communicated to the MN either through agent advertisements or as part of the beacon signal. Idle MNs are then free to move within the paging area, without having to obtain a new LCoA. When the MA receives packets for an idle MN, it broadcasts the received packets to all SAs within the paging area. The SAs buffer the packets until the MN registers with the MA. The SA serving the MN delivers its incoming packets. Upon receiving these packets, the MN obtains an LCoA address and registers with the MA. [3, 10]

The IDMP protocol, through the fast handover procedure, reduces the disruption caused by handover delays and limits packet loss. However, it does not address the handover delays. Although packet buffering might hide the delay for a file transfer application, the disruption will be apparent for real-time applications. The handover anticipation information could be used to reduce the handover delays. Furthermore, the fast handover procedure is only useful for movements within the MA domain. Therefore, it cannot be used to provide seamless inter-domain mobility for applications.

2.3.3 Cellular IP (CIP)

Cellular IP (CIP) [17, 9] is a micro-mobility management scheme that combines IP routing methods, with cellular location management procedures. It employs two handover mechanisms, hard and semi-soft handover, which trade-off very low packet loss for higher signaling costs. The CIP architecture consists of Gateways (GWs) and Base Stations (BSs). The Gateways interwork with Mobile IP to provide macro-mobility management. A GWs IP address is used as the Mobile IP CoA for all MNs within its domain. Within a CIP network, Gateways act as root nodes. All packets originating from the CIP network, regardless of their destination, are routed to the Gateway. The Gateway regularly broadcasts beacon packets to Base Stations (BS) within its domain. The BSs learn the path to the Gateway by keeping a record of the interface through which the beacon packets are received, and use it to route received packets to the Gateway. Distributed Routing and Paging Caches are used for call delivery and location management.

The CIP handover procedure is distributed and avoids the single point of failure. However, the handover delay is dependent on the topology of the network. Longer delays may occur if the MN moves between two BSs which have separate paths to the Gateway. Furthermore, handover across Gateway domains might result in longer delays because the MN needs to setup a path to the Gateway before it can perform the home registration.

2.3.4 Handoff-Aware Wireless Access Internet Infrastructure (HAWAII)

The Handoff Aware Wireless Access Internet Infrastructure (HAWAII) [45, 56] is an intra-domain micro-mobility management protocol. It uses IP routing mechanisms to manage user mobility within the network domain. The HAWAII architecture is similar to that of CIP. The domain root router has the same role as the Gateway in CIP. The main difference between the two protocols is that HAWAII uses specific signaling messages to setup and maintain routes to MNs within the HAWAII network.

The Hawai protocol reduces mobility signaling overhead by using IP routing to forward packets to MNs thereby avoiding tunneling overheads. However, the path setup/refresh procedure required to maintain connectivity to the domain root introduces new signaling overheads. Furthermore, mobile IP is still needed for inter-domain mobility as the Hawaii routing procedure would not be scalable across several domains.

Chapter 3

Literature Review

The previous chapter reviewed macro and micro mobility management techniques and analysed their performance. The analysis showed that micro mobility management techniques are not viable for inter-domain mobility. Most of those techniques use mobile IP for inter-domain mobility. Although mobile IP is capable of providing inter-domain mobility, the delays associated with its handover procedures, especially movement detection and home registration, make it inadequate for real-time applications. This chapter introduces a number of schemes that address mobile IP handover delays and provide enhancements that improve its handover performance.

3.1 Low-Latency Handover Techniques

3.1.1 Mobile IP movement Detection Algorithms

In [21], three mobile IP movement detection schemes are analysed: Lazy Cell Switching (LCS), Prefix Matching (PM) and Eager Cell Switching (ECS). The Lazy cell switching identifies a movement by the expiry of the router advert lifetime. The Prefix matching scheme identifies a movement by a change in the network prefix of the router advert. It has the advantage of avoiding unnecessary same-subnet handovers. Eager cell switching identifies a movement by the reception of an advert from a new mobility agent. They used MNs that cannot connect to multiple networks simultaneously and allocated one mobility agent per subnet. They analysed TCP/ UDP communications. The handover delays observed ranged from 2.77s for ECS, to 5.91s for LCS. These results show that all three scheme are inadequate for seamless application mobility.

In [22], the authors introduce the hinted cell switching (HCS) scheme, which uses link layer information (L2 triggers) to detect the MN's movements. The L2 triggers are generated when the L2 connection is lost. The MN then solicits a router advert to change its mobility agent. This eliminates the need to wait for the absence of router adverts (RA). However, scaling problems might be encountered if a large number of mobile nodes handover at the same time, leading to a huge response from neighbouring MAs.

3.1.2 A Multicast Vertical Handover Scheme

In [59], a multicast vertical handover scheme is implemented that allows users to maintain connectivity for as long as possible with minimum disruption during handover. The scheme uses a multicast CoA address, sending the MN's packets to several neighbouring MAs. The serving MA forwards the packets to the MN while the remaining MAs buffer them in case a handoff happens in the future. The handover decision is made according to signal strength in homogeneous networks. In heterogeneous networks, the MN hands over if it detects a lower-tier network (in terms of coverage area, but with higher bandwidth), or if it detects that it has moved out of range of the lower-tier network. The handover is executed by instructing the new MA to start forwarding packets and the old MA to stop forwarding and start buffering. If the old MA is out of reach, the stop forwarding request is forwarded through the new MA. The system allows for user preference or load balancing by the network to decide which cell the MN should handover to. To save power all interfaces for networks higher in the overlay hierarchy than the current one are turned off, and are only turned on if a handover is anticipated. Interfaces on a lower-tier network are put onto sleep mode where they occasionally check for connections to make sure they know that a network is available, in case a handover is needed.

Experiments with the scheme resulted in an average handover delay of 3s. This is inadequate for seamless application mobility. To improve performance a few enhancements were introduced to the system. These include:

- Faster (more frequent) beacon messages
- Packet double-casting: sending the same packets from two different MAs to the same MN, i.e. setting more than one MA to forward packets to the MN. Missing beacon or packet from one of the MAs signals that it is no longer reachable.
- Header double-casting: setting one MA to forward packets to the MN and another one to forward headers.

The results they obtained show that fast beaconing decreases handover latency but increases overhead (proportionally). Bandwidth is used whether or not data is being sent. Packet double-casting eliminates handover latency and loss but at a huge cost (sending the same data twice:

power and bandwidth cost). Header double-casting eliminates latency with less overhead but it still uses considerable resources.

3.1.3 Link Layer Hints and Notifications

[23] proposes the fast hinted cell switching movement detection scheme for mobile IP. It is based on the HCS scheme but here L2 triggers identify available mobility agents (MA). The MN is then able to connect to the MA immediately without having to solicit or wait for an advert. A single uni-directional GSM-encoded audio stream between two IEEE 802.11 WLAN access points, and used the SSID field in the APs to send the identity of the mobility agent to the MN. The MN was checking periodically whether the SSID and hence the MA has changed to detect movement. The MN was able to connect to the new MA without an agent advertisement and movement was detected almost instantaneously (1.131ms) through link-layer information. However, this does not deal with home registration delay which constitutes a major factor in the handover delay. RFC 4957 [39] reviews the use of L2 triggers, sent from the link layer interface to the IP module, to detect changes in IP configurations. The L2 triggers alert the IP module to the status of the link (link up/down). Some L2 triggers may also contain IP configuration parameters. The nature of information available depends on the underlying link layer technology.

- GPRS provides an L2 trigger to the IP module upon establishing a PDP context. It provides the IPv4 address of the new link, but for IPv6, it gives an interface identifier that can be used by the MN to create a link local address.
- CDMA2000 also provides L2 triggers with IPv4 address, and a link identifier for IPv6.
- For IEEE 802.11, when the MN associates with the AP, an L2 trigger is sent to the IP module along with the BSSID of the AP.

In [44], link layer hints are cataloged into several categories: link type hints, link identifiers, IP address identifiers.

- Link type hints: these describe link characteristics such as MN measured bandwidth, MN measured bit error rate, MN packet error rate, MN link data rate
- Link identifier: which uniquely identifies the link
- IP address identifier: link layer identifiers that can be used to identify the IP address, and detect changes in the IP configuration.

In [74], handover scheme is proposed that uses a likelihood function to trigger handover decisions based on link layer information. The function was implemented on pre-registration mobile IP handover. The function determines whether the link layer handover will result in an IP handover. This ensures pre-registration is only triggered when an IP handover occurs, thereby limiting unnecessary IP handovers and their associated costs. The likelihood function determines the probability of an IP handover based on a number of link layer parameters as well as IP parameters such as the subnet prefix. The probability value is compared to horizontal and vertical handover thresholds to determine if a handover is imminent. If the value is higher than the horizontal threshold, an intra-system IP handover is executed. If the value is higher than both the horizontal and vertical threshold an intersystem handover is executed. An analysis of the signaling cost incurred during handover shows that significant savings can be achieved using the likelihood function to accurately detect IP handovers. The function eliminates unnecessary IP handovers that occur in response to link layer handovers within the same subnet. As a result, the costs associated with the pre-registration and home registration procedures can be avoided.

3.1.4 Neighbour Lists

In [70] an intelligent mobility management system is proposed, that consists of mobile IP extensions and a modified 802.11 handover algorithm. The proposed mobile IP extensions are packet buffering, neighbour list updates and Link layer handover notifications.

- Packets are buffered at the FA when the MN anticipates a handover. When the handover is completed the HA tells the old FA to re-route the buffered packets to the new FA.
- Neighbour lists are held at the FA and contain the IP address and link layer type and quality of its neighbours. The MN may acquire this list from the router advert or send a request to obtain it. The information in the list allows the MN to connect immediately to the new FA in the event of a handover.
- L2 triggers alert the MN immediately when a link layer handover occurs allowing it to connect to the new access point and register with it eliminating the need for the movement detection procedure.

The modified 802.11 handover algorithm monitors the quality/strength of the wireless signal, if this falls below a designated threshold and a new FA with a better signal is detected:

1. The current FA starts buffering incoming packets.
2. The MN initiates a handover to the new FA using the neighbour list information.

3. The new CoA is sent to the HA

4. The HA requests the old FA to forward buffered packets to the new FA.

The combination of L2 triggers and neighbour list extensions reduces the handover delay by eliminating the need to wait for missing router adverts to detect an imminent handover, as well as the need to wait for an advert from the new FA to establish a new connection. Packet loss is also eliminated through packet buffering. However, it does not deal with the home registration delay which constitutes a significant factor in the handover delay.

3.1.5 Advert Cashing and Registration Simulcasting

In [12] a set of enhancements to minimize handover delay and packet loss are proposed. To reduce detection time, more frequent router adverts are suggested to limit the time an MN has to wait before discovering that it had moved away from its serving FA. The current recommendation for router advertisement intervals is 3-10s as specified in the neighbour discovery protocol [49]. However, more recent recommendations have suggested the figure be reduced to 30-70ms. The frequency of router adverts has to be traded off against the increased overhead though, especially over slow links. Fast router adverts can be an effective tool to reduce handover delay, but only at the right frequency, and provided that the signaling overhead is not substantially increased.

Router Advert Caching :

The router advert caching scheme [12] allows the MN to cache received adverts from routers in its vicinity, until it is ready to initiate a handover. When the MN finally decides to handover, it does not have to wait to receive a router advert. Instead, it can use the information in its cache to connect to the new access router. The results obtained using this method show much improved TCP performance, with the detection time close to zero.

Binding Update Simulcasting :

Another technique aimed at reducing registration time is binding update simulcasting [12]. This allows binding updates sent by the MN towards its HA during handover to be delivered across both the old and the new link. This ensures that the registration update reaches the HA through the fastest link.

3.1.6 Semi-Soft mobile IP handover

The term Semi-soft handover refers to handover schemes that maintain the connection to the old FA while establishing a connection to the new one. These schemes anticipate the handover decision and start looking for a new connection before the old one is lost. This ensures minimal interruption to the MN's communication session as delay is minimized and packet loss is close to zero. Mobile IPv4 has two semi-soft handover implementations: pre-registration and post-registration handover. Mobile IPv6 implements semi-soft handover through the fast mobile IP handover scheme.

Mobile IPv4 Pre - registration handover :

The pre-handover registration procedure is executed as follows:

1. the MN receives an L2 trigger indicating an imminent handover to a new FA
2. the MN requests a handoff from its serving FA. The handover request indicates the link layer address of the new FA.
3. the old FA inspects the new FA's address and determines if it belongs to a new access router,
4. the old FA obtains a new CoA from the new FA and forwards it to the MN.
5. The old FA establishes a temporary link to the new FA, to ensure correct delivery of packets during the handover.
6. The MN connects to the new FA and sends a fast binding update to the old FA.
7. If the MN loses its connection to the old FA before the handover is completed, the old FA forwards its packets to the new FA until the handover process is completed.

Mobile IPv4 Post- registration handover :

The post-registration scheme allows the MN's traffic to be directed towards the new FA before mobile IP registration is completed. This procedure can be triggered by either the old or the new FA.

1. When an imminent handover is anticipated, an L2 trigger is sent to either the new or the old FA.
2. When this trigger is received, a bidirectional edge tunnel (BET) is established between the two FAs in preparation for the handover.
3. When the old FA loses it connection to the MN, it starts forwarding traffic destined to it to the new FA.
4. The new FA buffers the received packets, and then forwards them to the MN once a connection is established between them. This is all done prior to mobile IP registration.

In [7], the authors investigated the pre/post-registration schemes [19]. The performance of the two schemes was analysed using the IEEE 802.11 protocol as the link layer. A simple analytical model to investigate the delay characteristics and the buffer requirements of a single node during handover. Both schemes were found to reduce handover delays. The post registration handover is actually faster since the only delay is the time to setup the tunnel. However, individual packets will experience the added delay of going through the tunnel. The pre-registration handover takes longer to complete because it has to wait for the registration, but is still better than normal mobile IP because it starts the registration before it actually connects to the new FA. However, the registration might complete before the MN has moved to the new FA and packets might have to wait for this to happen or they are lost.

Fast Mobile IP handover :

Fast mobile IP [38] is a variant of mobile IPv6 [36] that provides faster handover performance. It anticipates the handover decision and obtains the details of the new access router (NAR) prior to the actual link handover. This allows the MN to attach to it immediately after the link handover. A forwarding tunnel is setup between the previous and the new access routers (PAR/NAR). This is used to forward packets arriving at the PAR to the MN until the handover is complete. A description of the FMIPv6 handover protocol is given below.

1. the MN obtains an access point identifier (AP-ID) either by an L2 trigger or through router discovery.
2. the MN sends a router solicitation for proxy advertisement (RtSolPr) to its AR to resolve the AP-ID
3. the AR responds with a proxy router advert (PrRtAdv) containing the AP information [AP-ID, AR-Info]
4. the MN formulates a prospective NCoA and sends a fast binding update (FBU) message either through the new or old AR.
5. if the FBU has been sent to the PAR:
   - the PAR sends a handover initiation (HI) message to the NAR, in which it sends the NCoA, and to setup the tunnel (association between the MN and NAR so that MN's packets are forwarded to the NAR) between the PAR and the NAR.
   - the NAR responds with a handover acknowledge (HAck) message in which it

confirms the NCoA and it means that the forwarding tunnel between PAR-NAR can be setup.

- the PAR sends a fast binding acknowledgment (FBAck) to the MN to confirm the tunnel has been setup and the NAR accepts its NCoA.
- the MN sends a fast neighbour advertisement (FNA) to the NAR to attach to it.

6. If the MN has not sent an FBU through the PAR or it sent it but left before receiving an FBAck:
   - it sends an FNA containing the FBU to the NAR
   - The NAR processes the FNA and determines if the NCoA is not in use.
   - If it's in use it discards the FBU and sends a neighbour advertisement acknowledge (NAAck) in which it includes an alternate NCoA.
   - If the NCoA is okay, the NAR sends an FBU to PAR to setup the forwarding tunnel.
   - The PAR responds with an FBAck and starts forwarding packets to NAR.
   - The NAR receives the FBAck and starts forwarding packets received from the PAR to the MN.
7. The MN then performs the home registration as described in the mobile IPv6 protocol through the NAR. Packets will continue to arrive at the PAR until registration is completed.

In [5], the authors evaluated the fast handover scheme using real implementations of fast MIP on a network emulator. It was found that fast MIP can meet the requirements of even real-time applications. The results for fast mobile IP handover delays are in the range [3-15ms]. The results show that FMIP is independent of network delays and RA frequency. User perception tests, using an audio-video streaming application, forcing handovers without movements, acknowledged the satisfactory performance of fast MIP.

3.1.7 Adapting Handover Decisions

In [54], the authors attempt to reduce handover delays for real-time applications by limiting the number of unnecessary handovers made. If the application is a real-time application, less handover decisions are made. If it is a non real-time application, more handovers are triggered. The handover algorithm uses the number of beacon signals that are below a certain threshold to determine when the handover procedure should be triggered. This algorithm is adapted depending on the type of the application. For real-time applications less beacon signals are used to trigger the handover. For non-real-time applications, more beacons have to arrive before handover is triggered. For cellular to WLAN handovers, more beacons are required to trigger handover regardless of the application.

When handover is triggered, the Mobility Agent starts multicasting packets to both the old and new SAs. The MN remains connected to the old SA and only disassociates from it when the handover to the new one is finished. This means that no packet loss or delay is experienced by the application even if the handover fails or is delayed. An analysis of the throughput and delay during the transition region, where the beacon is below the threshold, reveals that higher throughput is achieved while moving from WLAN to cellular because more time is spent in the WLAN, and lower delays are experienced overall as less handovers are triggered in the process.

In the algorithm above, when the MN is connected to a cellular network, it has to wait for a long time to switch to WLAN. This would hinder high data rate applications. The application type should have been used here to determine whether or not the application is a real-time application, and trigger the handover accordingly. The delay and throughput improvements shown in the analysis can be achieved by a simple hysteresis handover procedure. However, the differentiation between real-time and non-real-time is still useful. Furthermore, the algorithm does not address handover delays and incurs considerable signaling costs through the soft handover and multicasting procedures.

3.2 Handover Decision Algorithms

The heterogeneous networks architecture presents the user with a diverse array of wireless technologies that are capable of providing network connectivity for his data applications. However, the varying characteristics of the wireless networks on the one hand represent the user with a challenging decision with regards to which network he should connect. Furthermore, networking applications have diverse quality of service requirements. Hence selecting a network that is suitable for the users application varies depending on the type of the application. In addition, a user might simply have a preference towards a certain network due to loyalty, cost or security issues. As a result, network selection becomes a multidimensional problem that has to account for multiple factors including network characteristics, application requirements and user preferences. In the context of a mobile user, the network choice becomes part of the handover decision process. The user has to decide, depending on the choice of available networks, whether or not to handover and to which network he should connect.

Various handover decision algorithms and techniques have been proposed in the literature. These vary in their approach, the criteria they use in their decision making and their adaptability towards user input. Some of these techniques focus on specific parameters such as power consumption and battery life [41, 48, 31, 35] gearing their handover decisions towards maximising these criteria. Others adopt a more holistic approach using multidimensional decision techniques such as fuzzy logic [43, 13, 27, 47]. Overall, most of the algorithms proposed in the literature adopt one of the many multi criteria decision processes to address the problem. These include: classical Multiple Attribute Decision Making (MADM) algorithms [65, 60, 66], Markov Decision Process (MDP) techniques [62, 34, 20] and Fuzzy logic [64, 51, 52, **?**, 71].

3.2.1 Fuzzy Logic

Fuzzy logic lends itself readily to handover the handover decision problem as its flexibility in defining parameters would accommodate for the dynamic nature for network conditions. In a traditional handover scenario, the handover decision is based on whether or not the received signal level (RSS) falls below a certain threshold. When comparing multiple networks, a more flexible approach such as fuzzy logic would provide more granularity in its description. Fuzzy logic uses membership functions which allow a parameter to take two different states at the same time [37]. Taking the RSS example, using a traditional system, a networks RSS is either above or below the threshold. Fuzzy logic can describe how far a certain networks RSS is above or below the threshold by assigning it a membership function for both states. For example, a network with a membership function of 0.4 above and 0.6 below is less far below the threshold than a network with a membership function of 0.2 above and 0.8 below.

In [28, 29], fuzzy logic is used in combination with neural networks to evaluate handover decisions. Network parameters such as signal level, network load and user velocity are evaluated using Multi-level Perceptron (MLP) neural networks [25] which feed the Fuzzy Logic Controller. Algorithm evaluations carried out using the OPNET [46, 14] network simulator considered ftp download response times and TCP throughput. Simulation results show improvement on ftp response time and slight improvement on TCP throughput, as compared to a traditional RSS threshold algorithm.

In [24] the authors use an elman neural network as part of the handover decision process. The neural network is used to predict the number of users, then fuzzy logic is applied to a number of network parameters including bandwidth and velocity to reach a handover decision. Simulations carried out by the authors show that this technique produces better results than conventional RSS algorithms.

In [72], a handover decision algorithm based on grey prediction models[30] and fuzzy logic is proposed. Grey prediction models are used to predict the networks RSS values, which are used in combination with available bandwidth and cost as the decision parameters. These parameters values are fed into the Fuzzy Logic Controller which assess the viability of handover. Simulation results show that this system reduces handover frequency.

3.2.2 Multiple Attribute Decision Making(MADM)

Multiple Attribute Decision Making (MADM) techniques [32] allow the combination of different network network criteria to obtain an optimal solution. An informed handover decision is hence achieved, which takes into account several Quality of Service (QoS) parameters and not only signal level making it ideally suited for an overlay environment. Various efforts have been made to make use of MADM methods for network selection.

In [55] MADM methods are identified as a good technique for handover decisions. They are compared to a few traditional handover methods and identified to produce better results. However, the actual MADM algorithms were not analysed or closely studied to investigate their individual properties. A more thorough evaluation of MADM methods is presented in [73]. It produces results that show the factors that influence handover decisions. Nevertheless, it does not include mechanisms for incorporating user preferences into the handover decision and only refers to it through application profiles.

In [1, 2], a handover architecture is introduced to implement MADM methods on mobile terminals. The architecture presents a flexible method for ranking a scoring network that includes both user preferences and application profiles. However, this scoring method depends purely on predefined scores set by the user, which are modified depending on the current network. It does not incorporate network characteristics directly into the decision process which might lead to inaccurate results. Furthermore, the process was not compared to other MADM methods to evaluate its performance.

3.2.3 Simple Additive Weighting (SAW)

The Simple Additive Weighting (SAW) algorithm [32] is a MADM algorithm that evaluates handover opportunities based on a normalized weighted sum of available networks' QoS parameter values. These parameters are formulated into a decision matrix where the rows represent the networks being evaluated and the columns represent the criteria on which the evaluation is based. The matrix is multiplied it by a weighting vector that defines the relative priorities of each of the criteria considered.

Given a decision concerning four networks (GPRS, WCDMA, 802.11, 802.16) compared on the basis of five attributes: cost, bandwidth, handover delay, battery life (delay, jitter, ...), the decision matrix is illustrated below:

$$M = \begin{bmatrix} network/attribute & Cost & Bandwidth & Delay & Jitter \\ GPRS & x_{11} & x_{12} & x_{13} & x_{14} \\ WCDMA & x_{21} & x_{22} & x_{23} & x_{24} \\ \text{802.11} & x_{31} & x_{32} & x_{33} & x_{34} \\ \text{802.16} & x_{41} & x_{42} & x_{43} & x_{44} \end{bmatrix}$$

To compensate for the varying scales of the different criteria and to ensure that their respective values are comparable, every element in the decision matrix is scaled as follows:

$$r_{ij} = \begin{cases} \dfrac{x_{ij}}{x_j^{max}} & \text{for criteria where a higher value is desired such as bandwidth.} \\ \\ \dfrac{x_j^{min}}{x_{ij}} & \text{for criteria where a lower value is desired such as cost.} \end{cases} \tag{3.1}$$

x_{ij} : denotes the i^{th}network and j^{th}attribute

r_{ij} : denotes the scaled value of the attribute.

x_{ij} : denotes the original value of the attribute before scaling.

x_j^{max} : denotes the highest value of the j^{th}attribute amongst the observed networks

x_j^{min} : denotes the lowest value of the j^{th}attribute amongst the observed networks.

To choose the weights of the different criteria, and given the imprecise nature of the relative importance of one attribute against another, fuzzy values are used initially to distinguish the importance of each attribute on the overall decision. For example, for a voice application, the attributes relative importance can be described as follows:

Cost	High
Bandwidth	Low
Handover delay	High
Battery life	Low

To use these fuzzy values in the decision process, they have to be converted into crisp (numerical) values which can be used in the overall score calculations. Several conversion scales [15] exist that can be used to assign numerical values to the above fuzzy values. The resulting numerical weights are normalized and multiplied by the scaled decision matrix to calculate the overall score of each of the candidate networks (equation 3.2)

$$C_i = \sqrt{\sum_{j=1}^{4} w_j r_{ij}} \tag{3.2}$$

C_i : denotes the i^{th}network's overall score,

w_j : denotes the weight of the j^{th}attribute

r_{ij} : denotes the scaled value of the j^{th}attribute for the i^{th}network.

The SAW ranking depicts the weights assigned to each attribute. The network with the greatest values for the most highly weighted attributes is ranked first. However, the networks' ranking does not always follow the ranking of the high priority attributes [32]. Some networks are penalized for low scores on low priority attributes. The ranking seems to account for all the attributes evenly, taking into account the weighting factors. This will be investigated further in chapter 5.

3.2.4 Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)

TOPSIS is a MADM algorithm that defines the best alternative as the one closest to the ideal solution, which combines the best values for each attribute. The first step is to normalize the attribute values to achieve a more comparable set. This is achieved as follows:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{4} x_{ij}^2}} \tag{3.3}$$

r_{ij} : denotes the normalized value of the j^{th}attribute for the i^{th}network

x_{ij} : denotes the original value of the attribute before normalization

w_j : denotes the weight of the j^{th}attribute

The next step is to scale the decision matrix using a set of weights that determine the importance of each attribute, as was done with the SAW algorithm. Based on this scaled matrix, the ideal and the negative ideal solutions are calculated. These represent the combination of the maximum attribute values and the minimum attribute values respectively. They are calculated as follows:

For a given attribute j,

$$A^{+}{}_{j} = \begin{cases} \max v_{ij} & \textit{if } j \in J \\ \min v_{ij} & \textit{if } j \in \acute{J} \end{cases} \tag{3.4}$$

$$A^{-}{}_{j} = \begin{cases} \min v_{ij} & \textit{if } j \in J) \\ \max v_{ij} & \textit{if } j \in \acute{J} \end{cases} \tag{3.5}$$

$A^{+}{}_{j}$ denotes the positive ideal solution for the j^{th}attribute

$A^{-}{}_{j}$ denotes the negative ideal solution for the j^{th}attribute

J denotes attributes where a higher value is desired.

$\acute{J}$ denotes attributes where a lower value is desired.

v_{ij} denotes the scaled weighted value of the j^{th}attribute for the i^{th}network

The process of determining the relative closeness of each alternative to the ideal solution is done as follows:

1. The distance from the ideal solution is calculated:

$$S_i^{+} = \sqrt{\sum_{j=1}^{5}(v_{ij} - (v_j)^{+})^2} \tag{3.6}$$

2. The distance from the negative ideal solution is calculated:

$$S_i^{-} = \sqrt{\sum_{j=1}^{5}(v_{ij} - v_j^{-})^2} \tag{3.7}$$

3. The relative closeness to the ideal solution is calculated:

$$C_i^+ = \frac{S_i^-}{S_i^+ + S_i^-}, 0 < C_{i+} < 1, i = [1 - 4] \tag{3.8}$$

Using the same decision matrix as the one used for the SAW algorithm and the same weight vector, TOPSIS produces a different network ranking [32]. This can be explained by the TOPSIS algorithm's favouring of higher priority criteria. As the algorithm's evaluation is based on the ideal solution, it is likely to give better scores to high priority attributes. This will be investigated further in chapter 5.

3.2.5 Connectivity Opportunity Selection

In [11], a new concept of Multi-hop connectivity opportunities is introduced. These are identified by their mode of access (single/multi-hop) and the networks through which they're accessed. Handover decisions are triggered by either an application specific event or a generic (link layer) event. Two handover service classes are defined depending on applications' QoS requirements:

- A Quality Guaranteed (QG) decision profile to be used with realtime applications.
- A Quality Flexible (QF) decision profile to be used with non-realtime applications

Both decision profiles take into account, as well as the network characteristics and applications' QoS requirements, the level of user mobility and the cost associated with each network. The profiles are defined by a set of utility functions that are used to assess the decision parameters such as cost and network conditions. The overall formulae combining all the utility functions is given by equation 3.9 below:

$$f(C_i, T, M) = [f_{net}(C_i, M) + f_{mob}(C_i, M)]\mathbf{w_{np}} + f_{cost}(C_i)\mathbf{w_{cost}} \tag{3.9}$$

$f(C_i, T, M)$	denotes the overall network score using the CSA algorithm.
$f_{net}(C_i, M)$	denotes the network's QoS utility function.
$f_{mob}(C_i, M)$	denotes the user's mobility profile.
$f_{cost}(C_i)$	denotes the cost utility function.
$\mathbf{w_{np}}$	denotes the network conditions weight.
$\mathbf{w_{cost}}$	denotes the cost weight.

The two weights used in the equation above allow the trade off between cost and quality. As the sum of the two weights is 1, their relative values represent the respective influence of cost and network conditions on the overall handover decision. Cost-conscious users are able to compromise the quality they receive to reduce costs. Both application profiles use a common utility function for cost. As the objective is to reduce costs, the utility function is inversely proportional to it (equation 3.10).

$$f_{cost}(C_i) = \frac{(c_{max} - c)}{c_{max}} \tag{3.10}$$

c_{max} denotes the maximum price the user is willing to pay for access to the network. As such, if the network's cost exceeds that value, the utility function returns a negative value, and the respective network is penalized.

The other utility functions in equation 3.9 define the network conditions and mobility profiles. They have different definitions in the Quality Guaranteed and the Quality Flexible profiles.

Quality Guaranteed :

The Quality Guaranteed (QG) profile is intended for use with realtime applications. Hence, its utility functions specify performance thresholds for network parameters to make sure prospective networks meet the application's QoS requirements. Network conditions are assessed through equation 3.11 below:

$$f_{net}(C_i, T) = \frac{\lambda - R_{me}}{\lambda} + \frac{D - \delta}{D} + \frac{PLR_{max} - \epsilon}{PLR_{max}} \tag{3.11}$$

The utility function, $f_{net}(C_i, T)$, defines thresholds for the application's mean data rate, R_{me}, maximum delay, D, and maximum packet loss, PLR_{max}. The network's data rate, λ, delay, δ, and packet loss, ϵ, measurements are bound by these thresholds. Any values outside these bounds are penalized by the respective utility functions. Mobility is defined in equation 3.12 below:

$$f_{mob}(C_i, M) = \begin{cases} 0 & \text{if Mobility is high and multi-hop connection} \\ 0.25 & \text{if mobility is low and multi-hop connection} \\ 0.5 & \text{if mobility is high and type is WLAN and single hop connection} \\ 0.75 & \text{if mobility is high, type WMAN or cellular} \\ 1 & \text{otherwise} \end{cases} \tag{3.12}$$

To limit interruptions to connectivity, the utility function, $f_{mob}(C_i, M)$, favours wide coverage networks, such as cellular and Wireless Metropolitan Area Networks (WMAN), over Local Area Networks (WLAN). Preference is also given to single hop over multi-hop connections.

Quality Flexible :

The Quality Flexible (QF) profile relaxes the QoS requirements set by the QG profile. As it is intended for non-realtime applications, which are delay tolerant, it does not specify delay in its utility function (equation 3.13). Throughput and packet loss are the parameters accounted for, as data integrity and speed of overall delivery are more important for non-realtime applications.

$$f_{net}(C_i, T) = \frac{\lambda}{\lambda_{max}} + \frac{\epsilon_{max} - \epsilon}{\epsilon_{max}} \tag{3.13}$$

The maximum data rate, λ_{max}, and packet loss, ϵ_{max}, values specified in the equation above are not QoS parameter thresholds. They represent the highest data rate and packet loss values, respectively, amongst candidate networks, and are used as a normalization factor. The QF mobility profile is defined by equation 3.14 below:

$$f_{mob}(C_i, M) = \begin{cases} 0.25 & \text{if mobility is high and multi-hop connection} \\ 0.5 & \text{if mobility is high, type is WLAN and single hop connection} \\ 0.75 & \text{if mobility is high, type WMAN or cellular} \\ 1 & \text{otherwise} \end{cases} \tag{3.14}$$

As was the case with the QG profile, wide area networks and single hop connections are also favoured here. However, the QF profile shows more tolerance to high mobility as non-realtime applications are better equipped to handle interruptions.

Chapter 4

Evaluating Handover Decision Algorithms

4.1 Evaluation Approach

In the previous chapter, a number of handover decision algorithms were reviewed. These varied in the approach they take in evaluating networks, the network metrics they take into account in their evaluation and the level of user interaction with the scheme. Hence, in order to evaluate and compare a number of these algorithm, a common attribute has to be identified. This common attribute lies in the claim all of the decision algorithms make, and that is to 'select the best network' for the user's applications. Hence, in evaluating the handover decision algorithms, this chapter will focus on how the decisions triggered by the different algorithms affect the performance of network applications.

However, network applications have different properties. Their demands on the network depend on their function and how the user interacts with them. Some are bandwidth intensive and generate high data rates. Other applications are delay-intolerant and require very low end-to-end delays. Any delay variations (jitter) would also harm these applications.

In [63], Internet traffic is classified into elastic and inelastic depending on the time relation between the traffic entities. Inelastic applications can be further classified into tolerant and intolerant depending on their sensitivity to network delays. Elastic applications can also be split into interactive and background applications, characterized by the level of user interaction.

4.1.1 Inelastic applications

Inelastic applications are predominantly real time applications that have stringent delay and throughput requirements. They generate data at a constant or a variable bit rate and require the maintenance of the timing characteristics of their data. They use the UDP transport protocol, usually combined with RTP to ensure the reliable delivery of the data.

Inelastic applications are very sensitive to delay variations in packet delivery. IP networks have varying delay characteristics depending on the load on the network. The network delay is characterised by a constant intrinsic element attributed to packet propagation and transmission by network nodes, as well as a variable element attributed to waiting time on network queues.

Depending on how sensitive they are to jitter, inelastic applications can be further classified into intolerant rigid applications (conversational) and tolerant adaptive applications (streaming). Conversational applications have rigid delay bounds and are unable to adjust their operation in the face of delayed packet delivery. This includes applications such as voice over IP (VoIP) and video conferencing. However, streaming applications such online TV can adapt their delay requirements, and are able to tolerate short interruptions using smoothing buffers. User service requirements can also dictate the level of tolerance to interruptions. A video conference application in a military environment, relaying battlefield data to commanding officers cannot tolerate service interruptions. However, a video conference of an academic lecture can accept some interruption to the service.

4.1.2 Elastic Applications

Elastic applications comprise traditional Internet applications such as email, web browsing and file transfer. They are not greatly affected by network delays but require low-loss reliable transport. Depending on the level of user interaction, elastic applications can be classified into interactive and background class. Interactive applications such as web browsing are characterized by a high level of interaction with the user. Background applications such as file downloads do not typically involve user input during their execution. Higher priority is given to interactive application to ensure responsiveness to user interactions.

Based on this assessment, a number of metrics can be identified to be used in the evaluating the decision algorithms. Two profiles will be defined:

- A realtime profile where delay and throughput are inspected to assess the effect of the handover decision on the performance of the application. Instantaneous throughput is considered here, as temporary drops in throughput may lead to packet loss resulting in a temporary but noticeable degradation in the quality of the application.
- a non-realtime profile where overall throughput is considered to evaluate the soundness of the handover decision

4.2 Simulation Setup

The evaluation is done using simulations as it allows for wider testing of various network types, sizes and conditions. The simulator used in this study in the discrete event network simulator NS2. NS2 provides an excellent platform for network simulations due to the extensive number of network modules already built within it, as well the ability to add new functionality as the specific at hand requires. A number of modules were developed as part of this work in order to provide the necessary tools for the evaluation process. These are described below:

4.2.1 New Developed Software modules for ns2

The software modules created within the project were built around the NIST mobility package. This package provides the 802.21 Media Independent Handover (MIH) functionality which provides the essential bridge between the link layer and the network layer. MIH provides triggers that notify the network layer of events in the link layer such as: Link Detected, Link Up, Link Down... MIH also provides commands that allows handover decisions to be executed from the network layer [40]. To complement the MIH functionality provided through the NIST package, the following modules were created:

Network measurement module :

This module collects network measurements such as delay, jitter, data rate and packet loss in realtime, to be used by the handover decision algorithms.

Handover decision modules :

The handover decision algorithms were created in a modular fashion such that further algorithms can be added without the need to modify the other modules.

4.2.2 Simulation Scenarios

The simulation setup emulates the overlay network scenario present in most urban areas in the developed. The overall structure is of an overlay 802.16 (WiMax) BaseStation (BS), scattered within it are three 802.11 (WiFi) Access Points (APs). One multi-homed Mobile Node (MN) having an 802.11 and an 802.16 interface moves gradually between the hotspots of the 3 802.11 APs while always in range of the 802.16 BS. Two 802.11 and one 802.16 mobile nodes are attached to the two 802.11 APs and the 802.16 BS respectively. The three single interface mobile nodes are stationary. Two routers represent an IP backbone network, attached to the them is a host that starts a communication session with the MN (See Figure 4.1 below). The communication session consists of a CBR application for the realtime profile and an FTP application for the non-realtime profile.

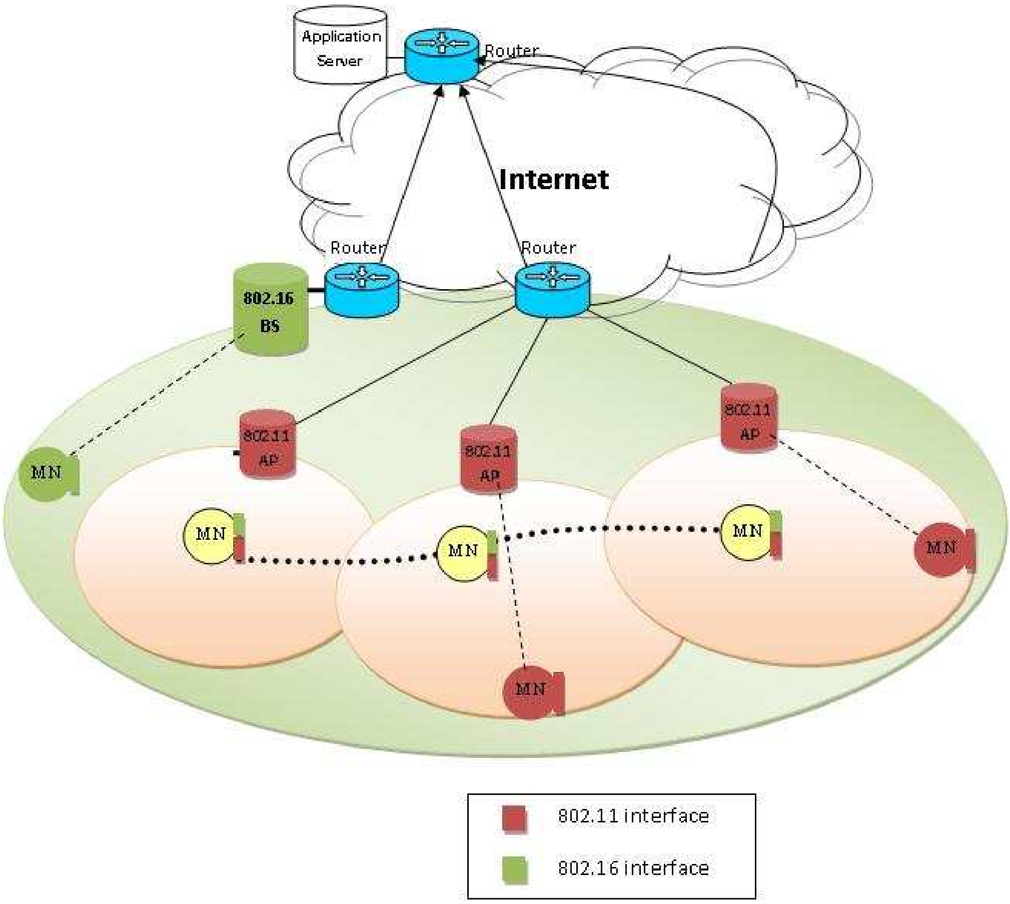

Figure 4.1: Simulation Scenarios

4.3 Decision Algorithms Evaluation Results

Three Handover decision algorithms will be evaluated in this section: TOPSIS, SAW and CSA. TOPSIS and SAW represent two alternatives of MADM solutions, whereas CSA presents a more flexible and user centric algorithm.

4.3.1 Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)

The MN connects to the 802.16 BS at the start of the simulation as it is the only available network. As the simulation develops, the MN starts moving and gradually entering the covering area of the 802.11 hotspots. Table 4.1 below shows how the TOPSIS algorithm scored the different networks as the MN entered the second 802.11 hotspot. The throughput shown is in bits per second and delay is in seconds. The terms MAC 6, MAC 2 and MAC 0 refer to the 802.16 BS, the first and second 802.11 APs respectively. As can been from the table, MAC 0 has the highest score, and the MN is handed over to it.

Figure 4.2 below shows the throughput for the realtime profile where CBR traffic is generated. As the MN is handed over to the new network, the throughput remains unchanged. This is to be expected, as the low data rate required by the application can be easily met by both networks.

Network	Score	Throughput	Delay	Packet loss
MAC 6	0.406362	12460000.00	0.000044	0.00
MAC 2	0.089693	9285714.28	0.010691	0.00
MAC 0	0.651208	27293934.68	0.013355	0.00

Table 4.1: Realtime traffic scenario -TOPSIS Network ranking

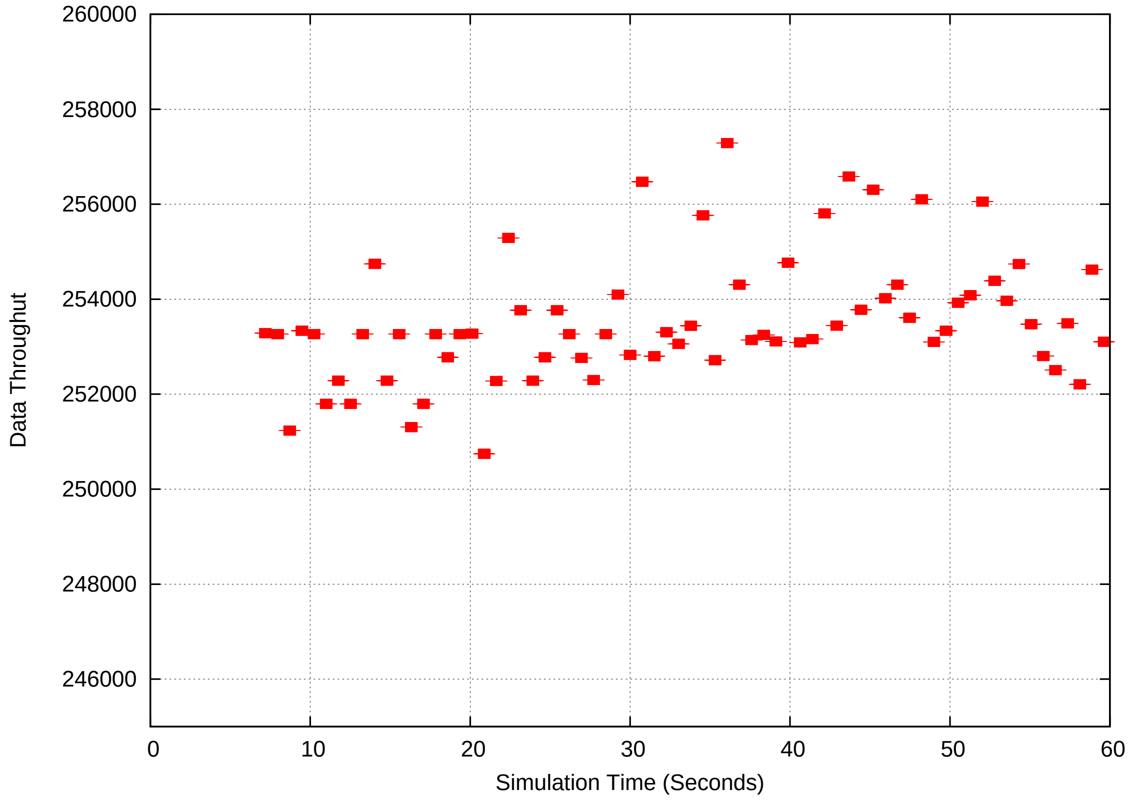

Figure 4.2: Realtime application throughput.

Although the new network has a higher data rate, that is of no significance to the application, as its throughput requirements were already met by the previous network As shown in table 4.1 the queuing delay of the new network is higher than the old network by 2 orders of magnitude, and is at a level that can be detrimental to realtime applications. However, because of the low values being measured here, its effect on the overall score was less than that of the throughput. it is interesting to note however, that the resulting end to end delay experienced by the application is actually lower in the second network

Table 4.2 shows the TOPSIS scores for non-realtime profile. Again handover is triggered to the new network. Here, the main concern is the throughput and how it is affected by the handover decision the handover is executed when the second access point is detected (MAC 0).

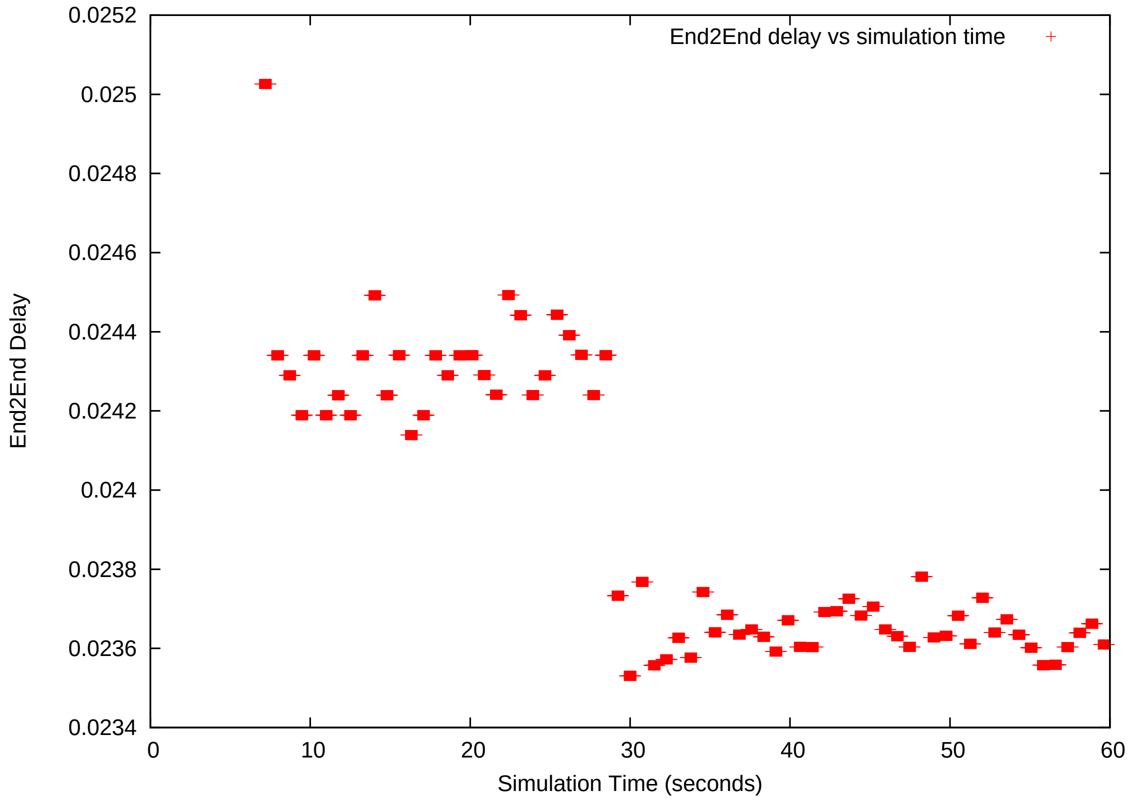

Figure 4.3: Realtime application end to end delay.

Network	Score	Throughput	Delay	Packet loss
MAC 6	0.443923	12460000.00	0.000044	0.00
MAC 2	0.000000	9285714.28	0.003203	0.00
MAC 0	0.818661	27293934.68	0.001232	0.00

Table 4.2: non-Realtime traffic scenario -TOPSIS Network ranking

as in the previous case, this network has higher throughput, and so is given a high score by the topsis algorithm, as shown in Table 4.2 However, inspecting the real throughput experienced by the user application reveals that the data throughput actually drops after the handover, as shown in Figure 4.4 below

Overall, the above two scenarios have illustrated two problems with throughput at the access point does not reflect the throughput to be expected by application data as it does not consider the load on the AP the queuing delay at the AP can give a clear indication of the access delay, but is only a small component of the overall delay, and hence the resulting performance might not reflect the decision higher emphasis should be placed on metrics that are more relevant to the application to ensure that the chosen network provides the best performance for the application. this is illustrated in the first scenario, where handover was triggered despite the

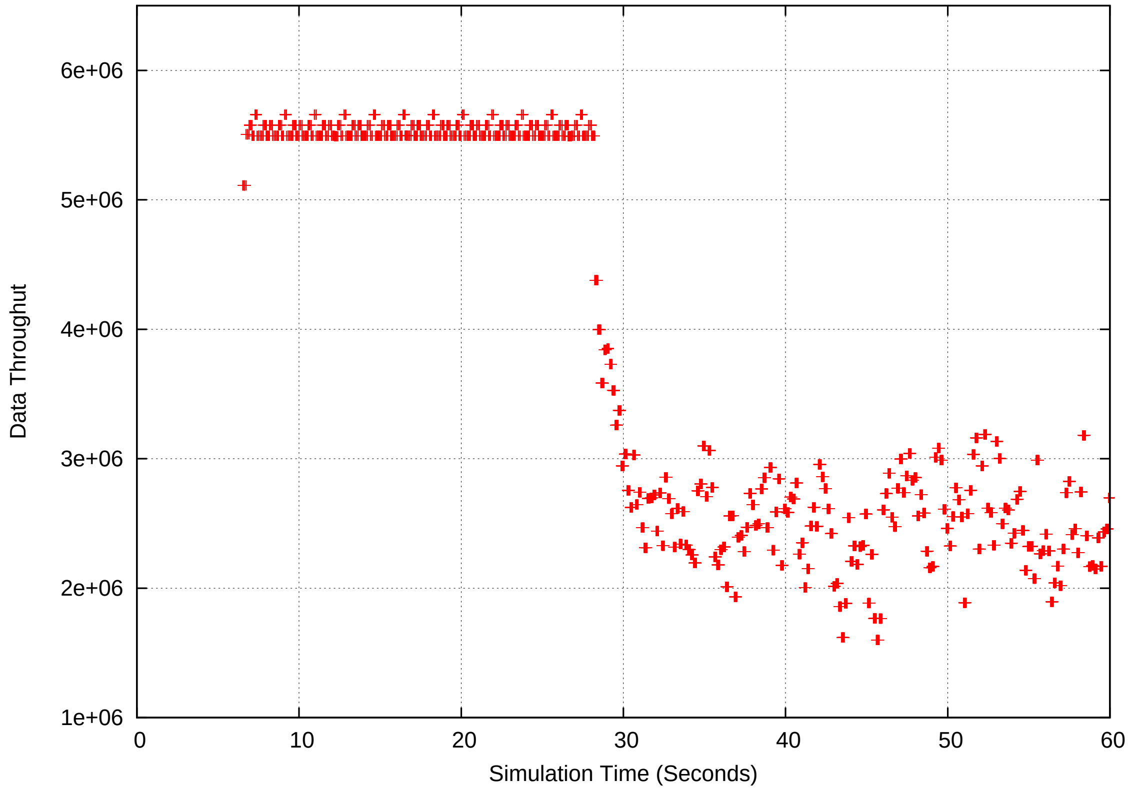

Figure 4.4: non-Realtime application throughput.

new network having longer delays. This decision was based on the higher data rates available at the new network. However, in an instance where the current data rate is satisfactory for the application, the application should not have been handed over to a network that has not longer delays. This is especially relevant as the application is time-sensitive, an attribute that was not considered by the algorithm.

4.3.2 Simple Additive Weighting (SAW)

As the MN moves into the coverage area of a new 802.11 network (MAC 0), the SAW algorithm identifies it as a better network, as indicated by the high score shown in table 4.3. The new network has a higher data rate than the other two networks but also a higher queuing delay.

Network	Score	Throughput	Delay	Packet loss
MAC 6	0.610319	12460000.00	0.000044	0.00
MAC 2	0.245108	9285714.28	0.010691	0.00
MAC 0	0.717942	27293934.68	0.013355	0.00

Table 4.3: Realtime traffic scenario -SAW Network ranking

Despite this, handover is executed to the new network and traffic is diverted towards it. In the first scenario, the simulated traffic emulates realtime application traffic. As such, throughput and delay experienced by application data are assessed. Figure 4.5 shows the data throughput

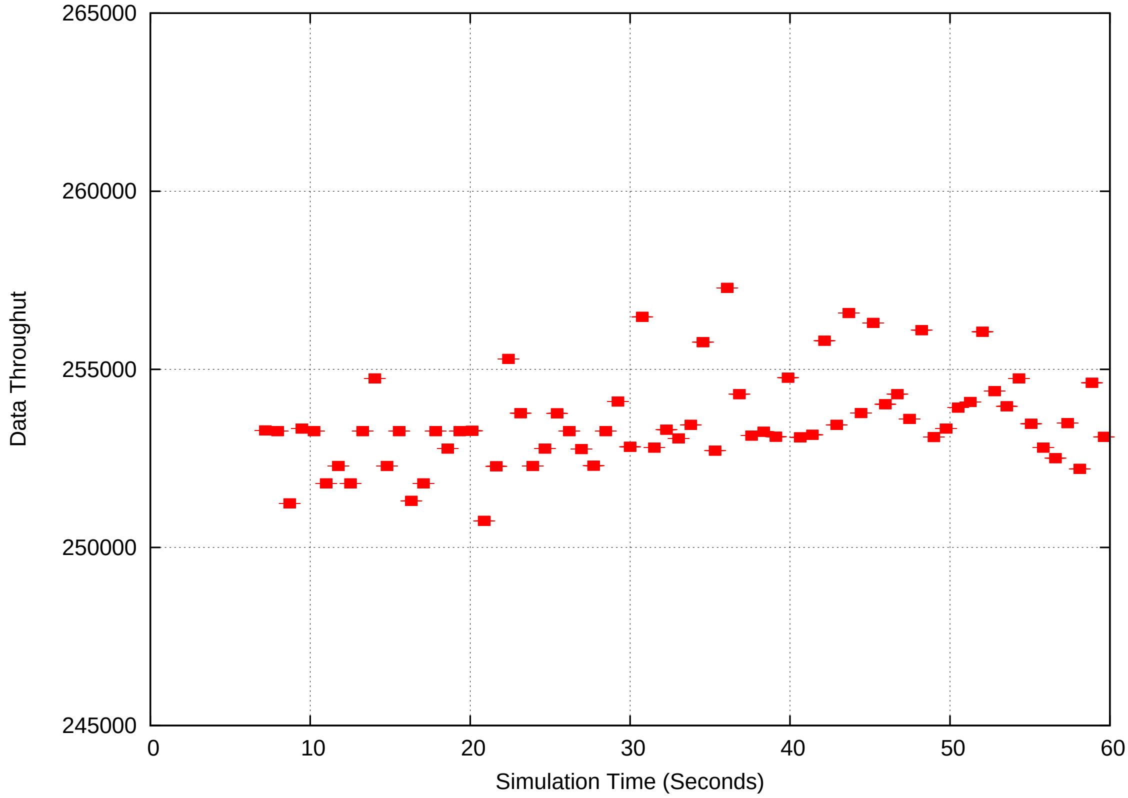

Figure 4.5: Realtime application throughput.

measured during the simulation. As expected, the throughput remains consistent throughout the simulation at around 200kbps, which is the application's data rate. The low data rate of the application meant that it was easily satisfied by both networks. The end to end delay dropped after the handover as shown in figure 4.6 This is inconsistent with the measurements collected during the handover decision process, which asserted that the new network has a higher queuing delay. Nevertheless, the overall end to end delay appears to be lower than the previous network.

In the second scenario, a file transfer application is run and the MN is set on the same movement pattern. Again, handover is triggered once the MN detects the presence of a new 802.11 network. The conditions are similar to the previous scenario with the new network experiencing higher data rates and queuing delays. The score assigned to it by the SAW algorithm deems the benefit of the higher data rate to outweigh the hindrance of the higher delay. This is quite reasonable given the current application is not time-sensitive. However, this information was not part of the decision process of the algorithm. A closer inspection of the application throughput before and after the handover instance reveals that the data throughput actually drops after the

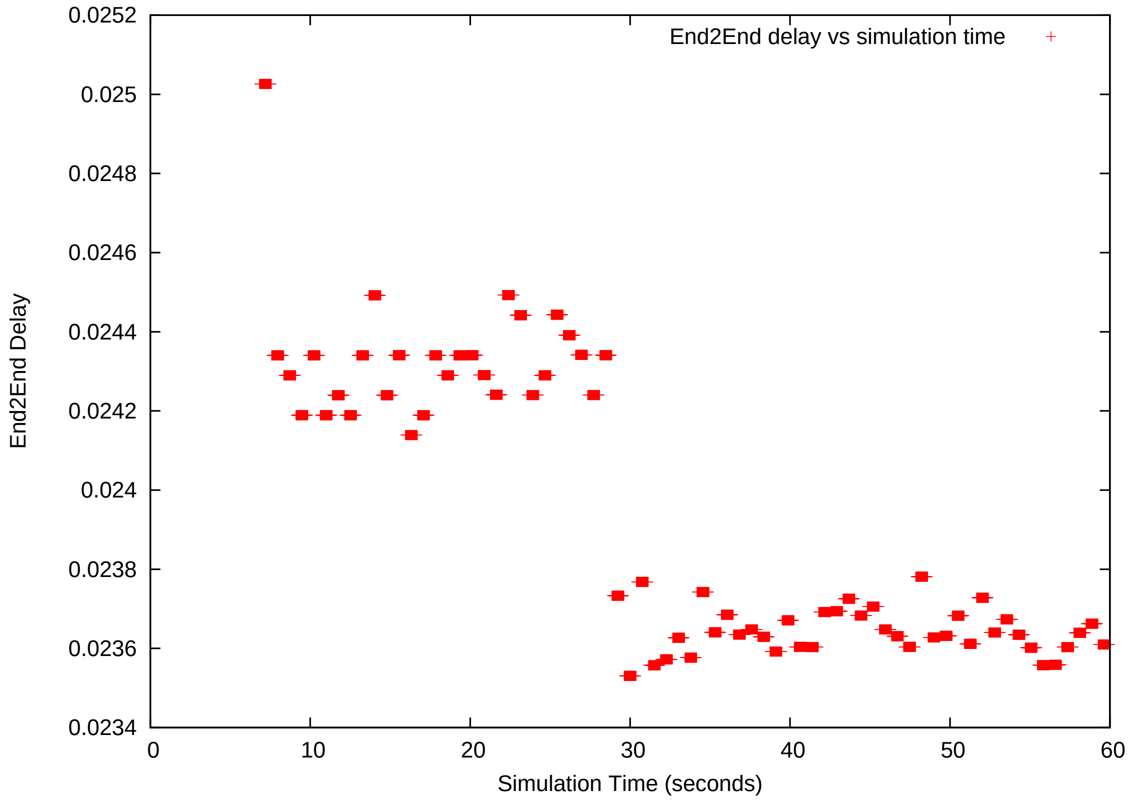

Figure 4.6: Realtime application end to end delay.

Network	Score	Throughput	Delay	Packet loss
MAC 6	0.406362	12460000.00	0.000044	0.00
MAC 2	0.089693	9285714.28	0.010691	0.00
MAC 0	0.651208	27293934.68	0.013355	0.00

Table 4.4: non-Realtime traffic scenario -SAW Network ranking

handover. This can be attributed to a higher load on the new host network, a factor overlooked by the SAW algorithm.

Considering both scenarios, it is clear that more information is needed to ensure handover decisions result in the best available network being chosen for the user's application. As mentioned in the previous section, network conditions at the access point might not reflect the overall end to end picture. Furthermore, the application's requirements have to be considered when trading off the importance of different network metrics. Although SAW does allow for the use of weights to emphasise certain network attributes, these weights are not modified to accommodate different application requirements. Adapting the weights dynamically in line with the current application would result in more informed decisions.

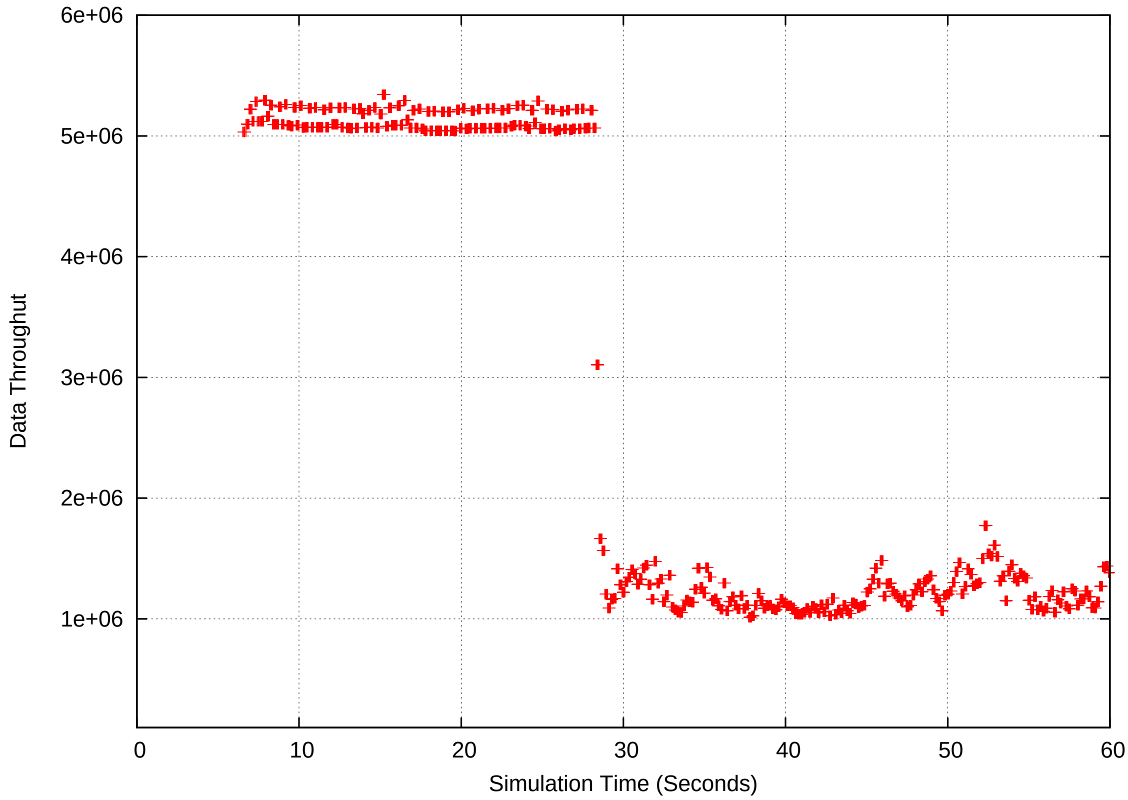

Figure 4.7: non-Realtime application throughput.

4.3.3 Connectivity Opportunity Selection Algorithm (CSA)

In this simulation, handover is triggered earlier than the previous two cases. As shown in table 4.5, handover is triggered to the first 802.11 network encountered by the MN It should be noted from the table, as well, that both data throughput and delay are more favorable in the original network. However, the CSA algorithm considers other factors that are not shown in the table.

Network	Score	Throughput	Delay	Packet loss
MAC 6	2.894226	12460000.00	0.000044	0.00
MAC 2	2.945901	9285714.28	0.013303	0.00

Table 4.5: Realtime traffic scenario -CSA Network ranking

First, its assessment of network attributes is based on a predefined application profile. In this case, as the application being simulated is a realtime application, the Guaranteed Quality (QG) profile is used. The QG profile defines thresholds for network metrics, such as the maximum delay or packet loss. The CSA algorithm's assessment of the network depends on whether or not these limits have been crossed. In this scenario, the parameters that influenced the CSA score are the mean data rate and maximum delay. In this simulation, these two metrics were set to: Mean data rate = 100kbps, Maximum delay = 0.2s. As the values for both networks fall within both thresholds, neither network is considered favorable. However, CSA also takes into account network cost, which is set to 20 for 802.16 networks (MAC 6) and 10 for 802.11 networks (MAC 2). As a result, MAC 2 is chosen as it is cheaper.

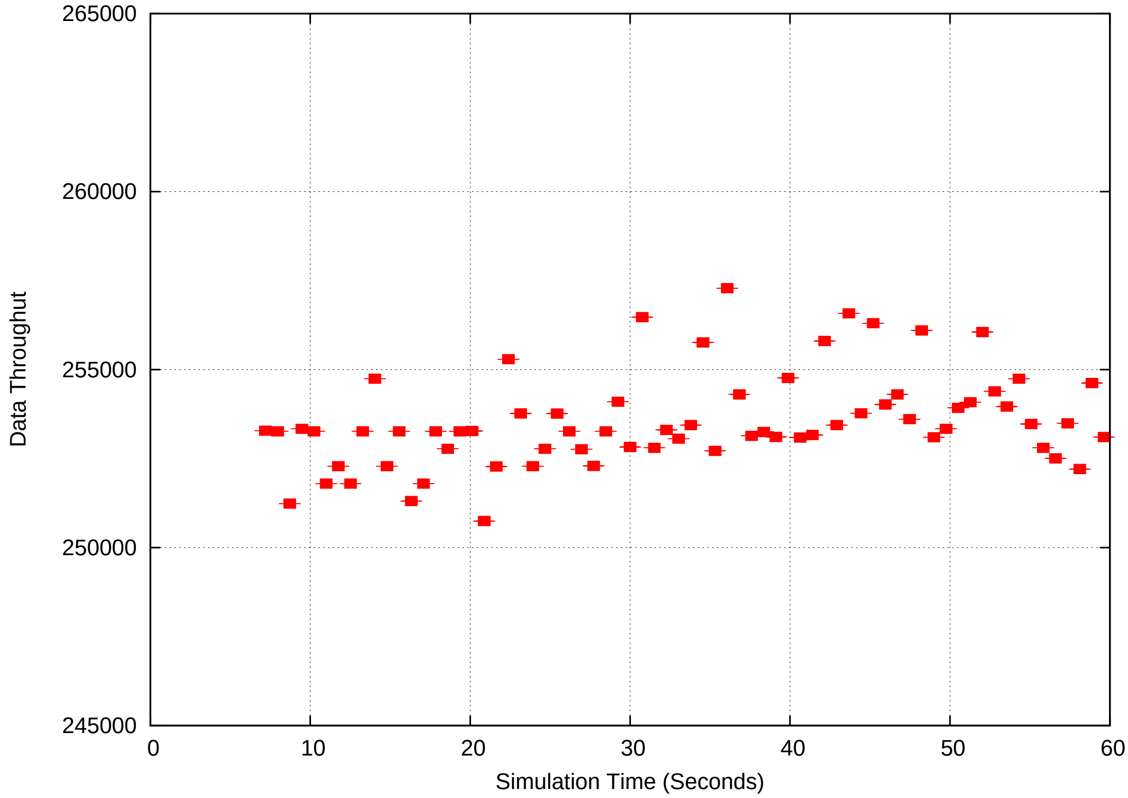

Figure 4.8: Realtime application throughput.

Figure 4.8 shows the application data throughput before and after the handover. Although the throughput observed is similar to that seen with the previous two algorithms, it is only in this case that the algorithm's assessment is consistent with the observed behavior. Using the predefined application profile, CSA concluded that a handover will not affect the application's QoS requirements. This is illustrated by the consistency of the application's data throughput before and after the handover. Considering the end to end delay observed in figure 4.9, although it has dropped after the handover, its value before the handover was already within the acceptable bounds. Hence, it does not constitute a significant improvement to the application's performance.

In the non-realtime application scenario, the CSA algorithm's behavior is different as it relaxes its QoS requirements. This can be noted from table 4.6 where similar network conditions with

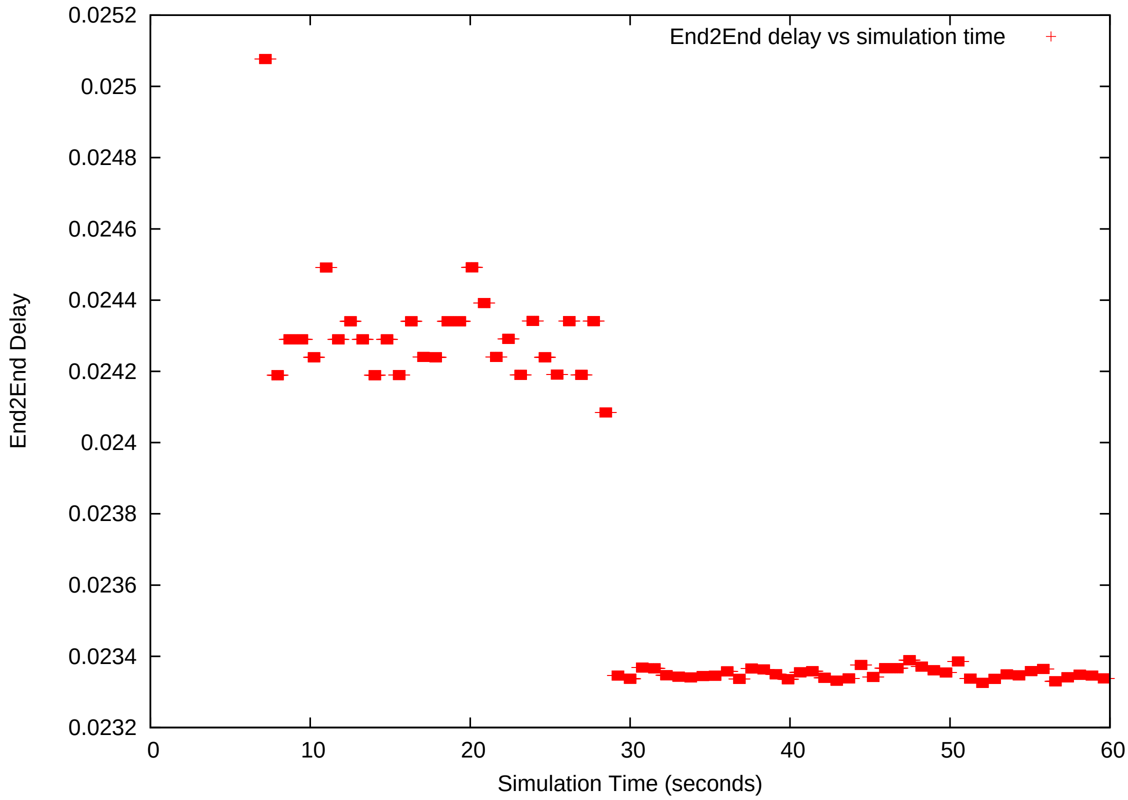

Figure 4.9: Realtime application end to end delay.

realtime applications lead to a handover. In this case, no handover is triggered as the original network (MAC 6) is seen to be better because of its higher data rates and despite it higher cost (20 compared to 10 for MAC 2).

Network	Score	Throughput	Delay	Packet loss
MAC 6	1.294807	12460000.00	0.000044	0.00
MAC 2	0.998097	9285714.28	0.003423	0.00

Table 4.6: non-Realtime traffic scenario -CSA Network ranking

Network	Score	Throughput	Delay	Packet loss
MAC 6	0.900000	12460000.00	0.003074	0.00
MAC 2	0.997312	9285714.28	0.003063	0.00

Table 4.7: non-Realtime traffic scenario -CSA Network ranking

Handover is only triggered when delay in network MAC 6 overtakes the delay in network MAC 2, as can be seen from table 4.7. However, since the present scenario involves non-realtime applications, this decision does not yields the best application performance. This can be clearly seen in figure 4.10 where data throughput drops after handover. The only justification for this action is the low cost of the new network. Hence, the user has to manage the trade-off between cost and quality to obtain the best value service.

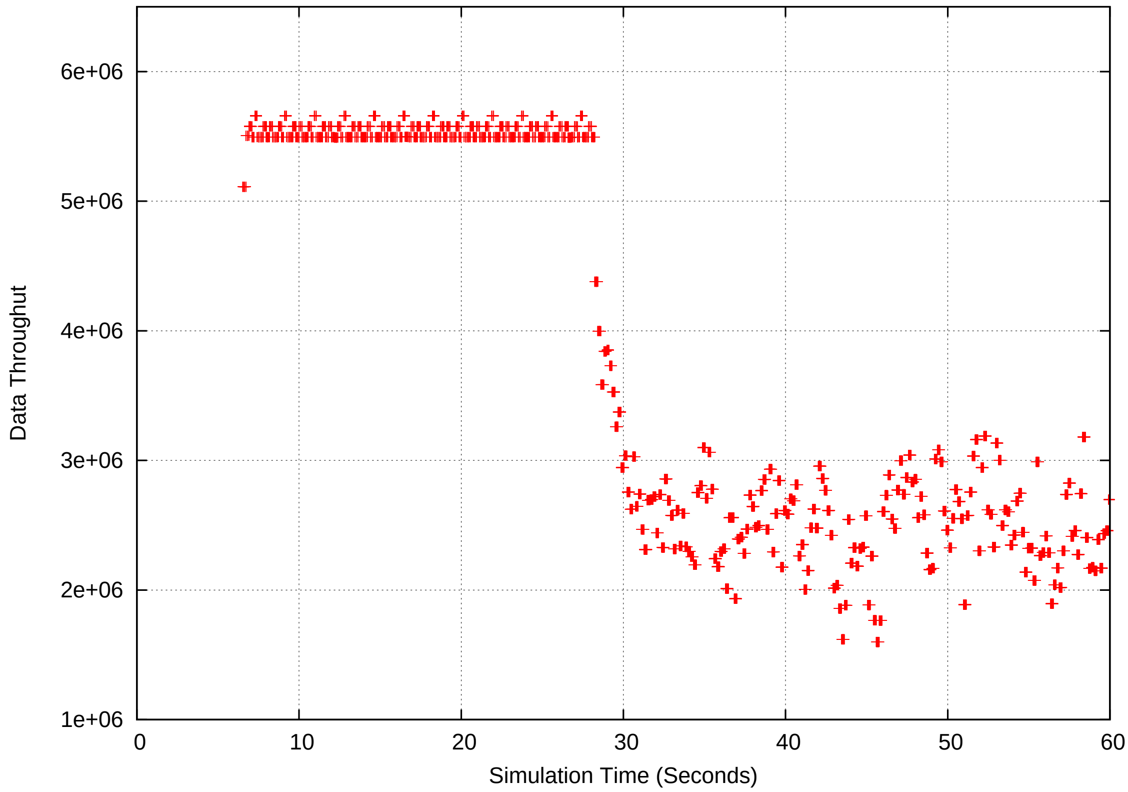

Figure 4.10: non-Realtime application throughput.

Chapter 5

Comparison of Handover Decision Algorithms

5.1 Performance Comparison of the Handover Algorithms

To analyze and further investigate SAW and TOPSIS algorithms, their relative ranking of a set of networks with different attributes will be compared. Four networks will be compared, two WCDMA networks (A1 and A3) and two WLAN networks (A2 and A4). The network attributes used to compare these networks are: cost, bandwidth, signal level, handover delay and battery life. These attributes were assigned priorities based on their impact on the Quality of Service of a voice application. Using fuzzy logic, these priorities were mapped onto the numerical weights shown in the table below.

$$\begin{bmatrix} Cost & High & 0.717 \\ Bandwidth & Low & 0.283 \\ Signallevel & Medium & 0.5 \\ Handoverdelay & High & 0.717 \\ Batterylife & Low & 0.283 \end{bmatrix} \tag{5.1}$$

The decision matrix **M** below shows the respective values of the 5 attributes for each of the four networks. These generic values will be used to compare how the two algorithms rank the different networks based on these attributes values.

$$M = \begin{bmatrix} \mathbf{A1} & 10.0000 & 30.0000 & 80.0000 & 0.9090 & 0.5000 \\ \mathbf{A2} & 7.0000 & 40.0000 & 80.0000 & 0.9090 & 0.5000 \\ \mathbf{A3} & 1.0000 & 80.0000 & 20.0000 & 0.2830 & 1.0000 \\ \mathbf{A4} & 2.0000 & 40.0000 & 40.0000 & 0.2830 & 1.0000 \end{bmatrix} \tag{5.2}$$

Using the same decision matrix and the same attribute priorities, as shown above, the two algorithms generate two different sets of ranking results.

$$SAW \begin{bmatrix} \mathbf{A3} & 0.61427 \\ \mathbf{A2} & 0.50864 \\ \mathbf{A1} & 0.46853 \\ \mathbf{A4} & 0.39877 \end{bmatrix} \tag{5.3}$$

$$TOPSIS \begin{bmatrix} \mathbf{A3} & 0.66630 \\ \mathbf{A4} & 0.53304 \\ \mathbf{A2} & 0.45054 \\ \mathbf{A1} & 0.33370 \end{bmatrix} \tag{5.4}$$

To investigate the factors contributing to the variation of scores between the two algorithms, the values of the attributes will be varied between the maximum and the minimum figures used to observe the effect of different criteria on the overall ranking. In particular, the values of the attributes with the lowest weights are varied, since their values are the ones that differ between networks WLAN2 and WCDMA1/2 whose ranking is swapped around between SAW and TOPSIS. Figure 5.1 and Figure 5.2 below show the ranking of the different alternatives as the values of attributes 4 and 5 are changed.

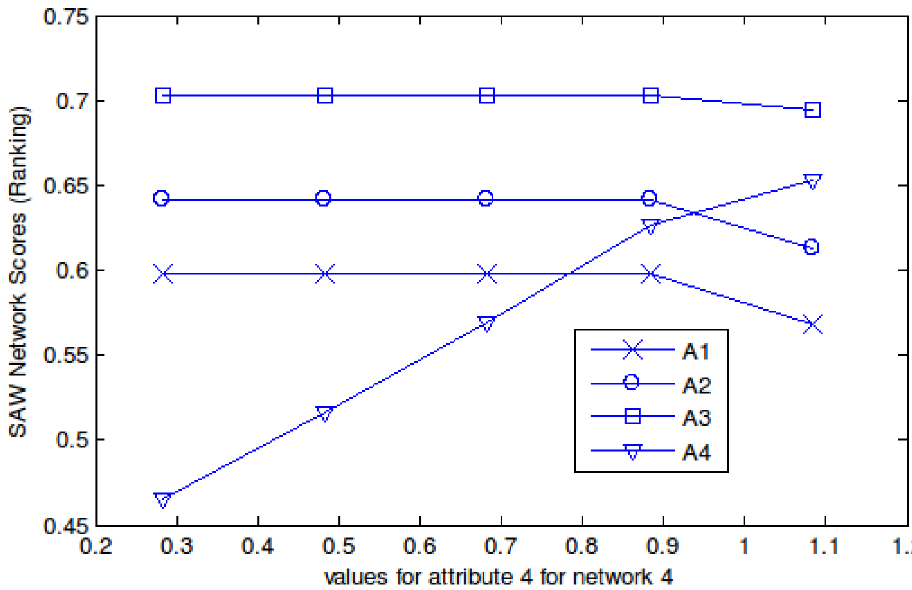

Figure 5.1: SAW ranking as the values of attributes 4 and 5 are changed for network A4.

It can be seen from the graphs that as the value of the attribute is changed with SAW, the

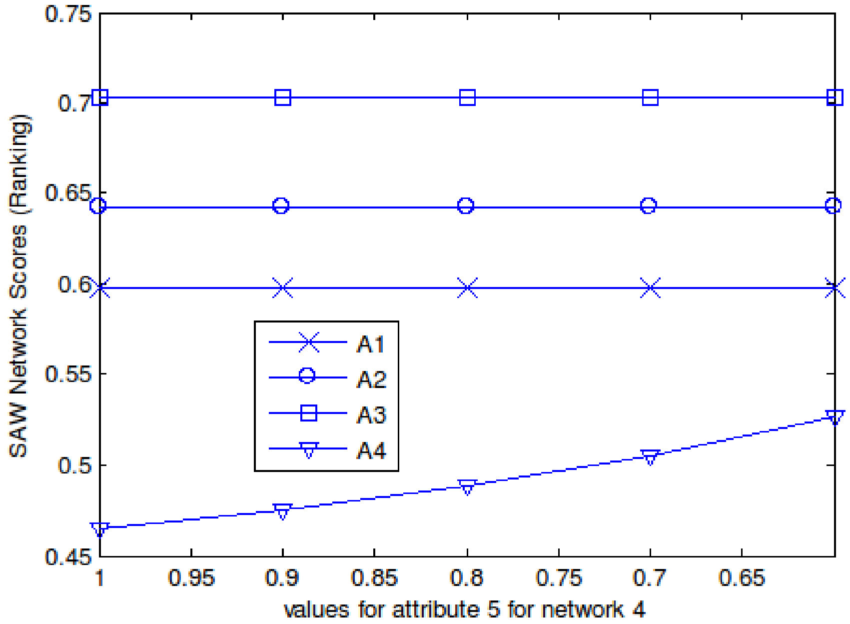

Figure 5.2: SAW ranking as the values of attribute 5 are changed for network A4.

network ranking changes as the value of those attributes for A4 approaches the values for A1/2. Attempting to change only the values of attribute 5 did not result in any changes in the network ranking. This emphasizes the supposition that the SAW reflects a complete picture of all the elements combined and no one attribute dominates the decision process. To verify this property, the highest priority attribute, attribute 1, of a low scoring network, network 1, was increased to check if that alone would improve its ranking. As can be seen from Figure 5.3, no changes can be observed on the ranking of the networks.

This is in contrast with TOPSIS, where as the value for attribute 1 is changed for network A1, its ranking changed and it moved above A2 (see Figure 5.5 TOPSIS ranking as the values of attributes 1 are changed for network A1.). This behaviour can have a major influence on how handovers are handled. As the user moves to a network with a very highly favourable attribute such as bandwidth which could be of use for a file download application. If the decision algorithm prevents handover to such a network because other attributes are not as favourable, the application loses the opportunity to make use of that network. In such a scenario, TOPSIS is more favourable to SAW.

To investigate this further, let us look at the decision matrix below, which was used to calculate

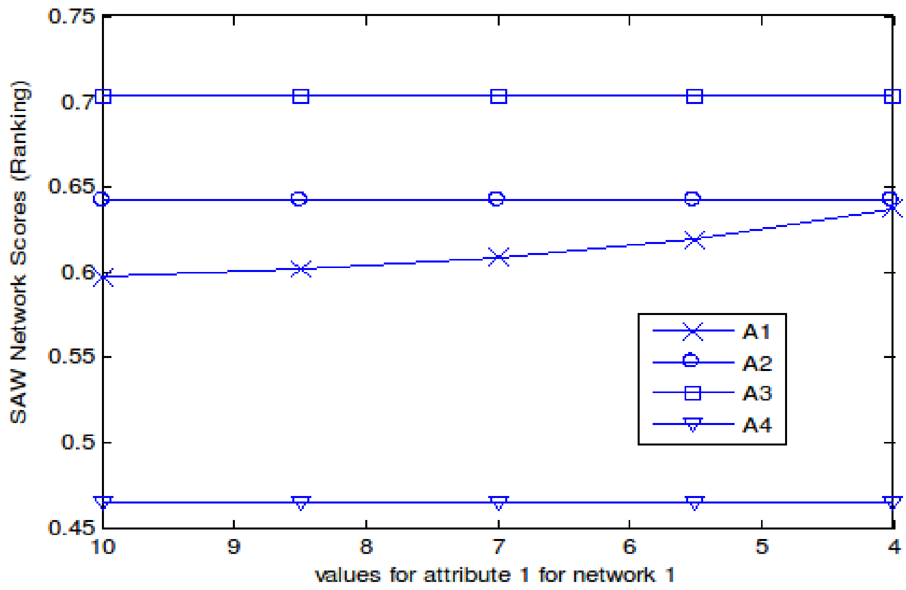

Figure 5.3: SAW ranking as the values of attributes 1 are changed for network A1.

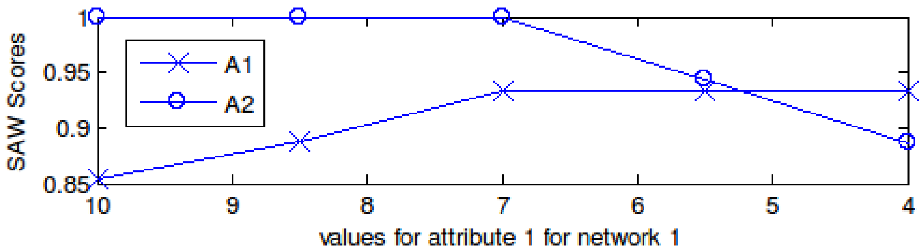

Figure 5.4: SAW ranking as the values of attributes 1 are changed for network A1.

the last score point on the graph, and taking networks A1 and A2 in isolation, it can be seen that:

$$M = \begin{bmatrix} 4.0000 & 30.0000 & 80.0000 & 0.9090 & 0.5000 \\ 7.0000 & 40.0000 & 80.0000 & 0.9090 & 0.5000 \\ 1.0000 & 80.0000 & 20.0000 & 0.2830 & 1.0000 \\ 2.0000 & 40.0000 & 40.0000 & 0.2830 & 1.0000 \end{bmatrix} \quad (5.5)$$

- For attribute 1, which is a cost parameter, the difference between the two networks is 3/7 in favour of network A1.

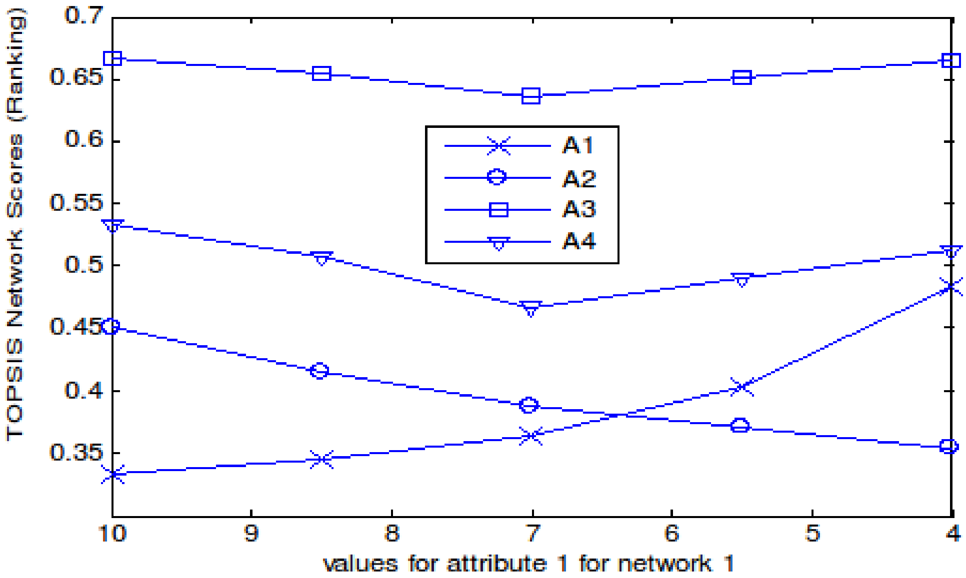

Figure 5.5: TOPSIS ranking as the values of attributes 1 are changed for network A1.

- For attribute 2, the difference between the two networks is in favour of network A2.

Given that these are the only differences between the two networks, and that attributes 1 and 2 have the same weighting, network 1 should be ranked higher than network 2. This is not the case here as can be seen from Figure 5.3. However, if these two networks are evaluated in isolation of the other networks, using a decision matrix as shown below, the results are different (see Figure 5.4 above).

$$M = \begin{bmatrix} 4.0000 & 30.0000 & 80.0000 & 0.9090 & 0.5000 \\ 7.0000 & 40.0000 & 80.0000 & 0.9090 & 0.5000 \end{bmatrix} \tag{5.6}$$

The scores obtained are: [A1 0.93403] [A2 0.8869]

This shows how SAW ranking is distorted by the inclusion of other networks as each network's score is relative to other networks it is compared to, especially the network with the highest score for a specific attribute. The lack of a pairwise comparison means that each network's score is only valid when taken within the group and not independently.

To check if the same problem can exists with TOPSIS, the following decision matrix is used to calculate the network scores:

$$M = \begin{bmatrix} 10 & 30 & 80 & 0.909 & 0.5 \\ 7 & 40 & 80 & 0.909 & 0.5 \end{bmatrix} \tag{5.7}$$

The resulting scores are shown in Figure 5.6 below:

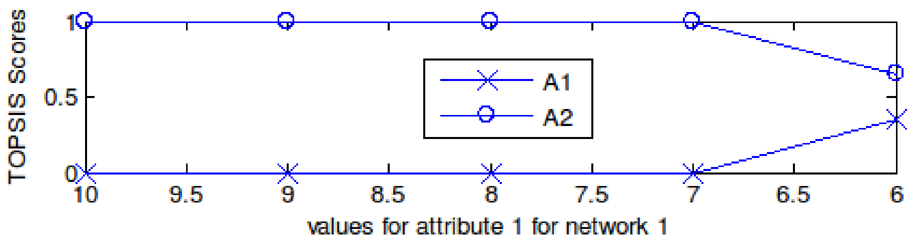

Figure 5.6: TOPSIS ranking as the values of attributes 1 are changed for network A1.

The effect is more apparent with TOPSIS as it relies on the existence of an ideal solution. As there are only two alternatives, one of them is taken as the positive ideal and the other as the negative ideal. Changing the attribute values as in Figure 6 did not allow network A1 to overtake A2's ranking. Overall, network rankings are affected for both SAW and TOPSIS if the number of networks is changed. TOPSIS is more sensitive to higher priority attributes and does not penalize networks on low scores on low priority attributes. SAW gives a more evenly distributed score accounting for all the criteria.

5.2 Pairwise comparison

As was mentioned in the previous section, the relative ranking of networks using the SAW algorithm is affected by the other networks being compared during the evaluation. This distortion can be attributed to how each network's criteria are scaled before being their evaluation using SAW. This scaling results in a value that is proportional to the minimum/maximum value amongst the different networks for that criterion. Hence, each network's score is dependent on how close its criteria values are to these minima/maxima.

The example considered above with reference to network A1 and A2's ranking within the four network group and in isolation demonstrates this issue. In that example, the following matrix was used:

$$M = \begin{bmatrix} 4.0000 & 30.0000 & 80.0000 & 0.9090 & 0.5000 \\ 7.0000 & 40.0000 & 80.0000 & 0.9090 & 0.5000 \\ 1.0000 & 80.0000 & 20.0000 & 0.2830 & 1.0000 \\ 2.0000 & 40.0000 & 40.0000 & 0.2830 & 1.0000 \end{bmatrix} \tag{5.8}$$

As explained earlier, network A1's score for parameter 1 is higher than that of network A2 and vice versa for parameter 2. However, the difference between the two networks' scores is higher for parameter 1 which should result in a higher overall score for network A1. On the contrary, the result shown in Figure 5.3 shows a higher score for network A2. This can be explained by the proportionality of the two networks' scores to the minimum and maximum parameter values. For parameter 1 which is a cost parameter, a minimum value is desired and both networks' scores are very high compared to that value. For parameter 2, a maximum value is desired, and the two networks' scores are closer to that value than for parameter 1. The proportionality of the difference in scores between the two networks to the minimum/maximum value is higher for parameter 2. Hence, the change in this parameter has a higher contribution to the overall score.

To verify this behaviour, the minimum value for parameter 1 is changed from 1 to 2 as shown in the matrix below.

$$M = \begin{bmatrix} 4.0000 & 30.0000 & 80.0000 & 0.9090 & 0.5000 \\ 7.0000 & 40.0000 & 80.0000 & 0.9090 & 0.5000 \\ 2.0000 & 80.0000 & 20.0000 & 0.2830 & 1.0000 \\ 2.0000 & 40.0000 & 40.0000 & 0.2830 & 1.0000 \end{bmatrix} \tag{5.9}$$

This should make the difference between the two networks proportionally higher for parameter 1 relative to the minimum value. The ranking, as shown in Figure 5.7 below, of networks A1 and A2 is swapped for the last point on the graph (which corresponds to the matrix shown above. This confirms the effect of SAW scaling, and particularly its proportionality to the minima/maxima, on the relative ranking of networks.

To address this problem a different scaling and normalization approach has to be used to produce an overall ranking that reflects the relative ranking of each network amongst the other networks regardless of how many networks are evaluated.

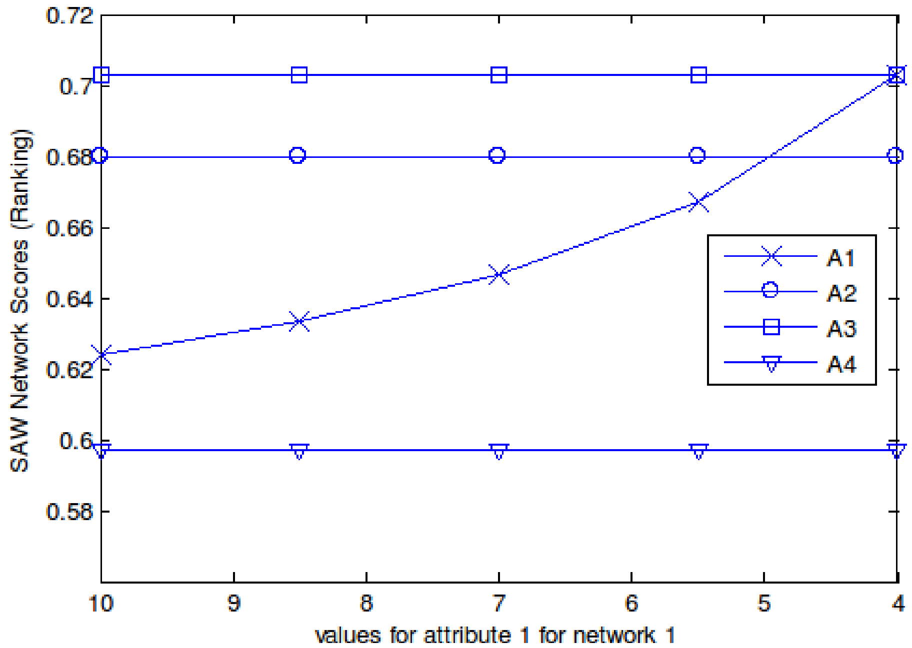

Figure 5.7: SAW ranking as the values of attributes 1 are changed for network A1.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

This work addressed the issue of Seamless handover across heterogeneous networks. The first step was to review the issue of handover latency and how that can be resolved. The consensus among researchers was that link layer information is needed at the network layer to improve the efficiency of layer three handover. The main body of the work centered around handover decision algorithms A framework is defined for evaluating handover decision algorithms. The aim of this framework is to create a common assessment mechanism for the highly varied handover decision algorithms.

This framework assesses the respective algorithms by considering the effect of the handover decisions on applications' performance. The metrics used to evaluate the application vary depending on the type of the application To facilitate the evaluation process, a set of software modules were developed as part of the network simulator NS2. These modules allow mobile devices to gather the necessary information to assess network conditions.

6.2 Open Issues

A number of the difficulties were identified in making the right handover decisions. One of the main issues is the inconsistency between the network conditions as measured at the access point or the base station, and the end to end conditions experience by the application. This leads to unfavourable handover decisions as the handover algorithm is only aware of local conditions.

Another issue is the algorithm's adaptability to user and application preferences. An algorithm might be able to select the most optimal network in terms of QoS metrics, but the user might prefer a cheaper network. Algorithms should be able to accommodate different application requirements and should adapt their decisions to the type of application. Using application profiles with thresholds or utility functions can result in better decisions than using absolute values. If an application is in an operating range that satisfies its requirements, it gains no benefit from moving to another network with better conditions. In fact, it might be penalized due to the disruption that might be caused by the handover.

Bibliography

[1] T. Ahmed, K. Kyamakya, and M. Ludwig. Architecture of a Context-Aware Vertical Handover Decision Model and Its Performance Analysis for GPRS-WiFi Handover. In *Proceedings of the 11th IEEE Symposium on Computers and Communications*, pages 795–801. IEEE Computer Society, 2006.

[2] T. Ahmed, K. Kyamakya, M. Ludwig, K. Anne, J. Schroeder, S. Galler, K. Kyamakya, K. Jobmann, S. Rass, J. Eder, et al. A context-aware vertical handover decision algorithm for multimode mobile terminals and its performance. The IEEE/ACM Euro American Conference on Telematics and Information Systems (EATIS 2006), February 7-10, 2006, Santa Marta, Colombia, 2006.

[3] I. Akyildiz, J. Xie, and S. Mohanty. A survey of mobility management in next-generation all-ip-based wireless systems. *Wireless Communications, IEEE*, 11(4):16–28, Aug. 2004.

[4] S. Balasubramaniam and J. Indulska. Vertical handover supporting pervasive computing in future wireless networks. *Computer Communications*, 27(8):708–719, 2004.

[5] C. Bernardos, I. Soto, J. Moreno, T. Melia, M. Liebsch, and R. Schmitz. Experimental evaluation of a handover optimization solution for multimedia applications in a mobile IPv6 network. *European Transactions on Telecommunications*, 16(4):317, 2005.

[6] M. Bernaschi, F. Cacace, and G. Iannello. Vertical handoff performance in heterogeneous networks. *Parallel Processing Workshops, International Conference on*, 0:100–107, 2004.

[7] C. Blondia, O. Casals, L. Cerdà, N. Van den Wijngaert, and G. Willems. Performance evaluation of layer 3 low latency handoff mechanisms. *Mobile Networks and Applications*, 9(6):633–645, 2004.

[8] W. Bolton, Y. Xiao, and M. Guizani. IEEE 802.20: mobile broadband wireless access. *IEEE Wireless Communications*, 14(1):84–95, 2007.

[9] A. Campbell, J. Gomez, S. Kim, A. Valkó, C. Wan, and Z. Turányi. Design, implementation, and evaluation of cellular IP. *IEEE Personal Communications*, 7(4):42–49, 2000.

[10] A. Campbell, J. Gomez, S. Kim, C. Wan, Z. Turanyi, and A. Valko. Comparison of IP micromobility protocols. *IEEE Wireless Communications Magazine*, 9(1):72–82, 2002.

[11] D. Cavalcanti, N. Nandiraju, D. Nandiraju, D. Agrawal, and A. Kumar. Connectivity opportunity selection in heterogeneous wireless multi-hop networks. *Pervasive and Mobile Computing*, 4(3):390–420, 2008.

[12] R. Chakravorty, P. Vidales, K. Subramanian, I. Pratt, and J. Crowcroft. Performance issues with vertical handovers-experiences from GPRS cellular and WLAN hot-spots integration. In *Proceedings of the Second IEEE International Conference on Pervasive Computing and Communications (PerCom'04)*, page 155. Citeseer, 2004.

[13] P. Chan, Y. Hu, and R. Sheriff. Implementation of fuzzy multiple objective decision making algorithm in a heterogeneous mobile environment. In *Wireless Communications and Networking Conference, WCNC2002*, pages 332–336.

[14] X. Chang. Network simulations with OPNET. In *Proceedings of the 31st conference on Winter simulation: Simulation—a bridge to the future-Volume 1*, pages 307–314. ACM New York, NY, USA, 1999.

[15] S. Chen, C. Hwang, M. Beckmann, and W. Krelle. *Fuzzy multiple attribute decision making: methods and applications*. Springer-Verlag New York, Inc. Secaucus, NJ, USA, 1992.

[16] S. Cherry. Broadband a go-go. *IEEE Spectrum*, 40(6):20–25, 2003.

[17] F. Chiussi, D. Khotimsky, S. Krishnan, et al. Mobility management in third-generation all-IP networks. *IEEE Communications Magazine*, 40(9):124–135, 2002.

[18] S. Das, A. McAuley, A. Dutta, A. Misra, K. Chakraborty, and S. Das. IDMP: an intradomain mobility management protocol for next-generation wireless networks. *IEEE Wireless Communications*, 9(3):38, 2002.

[19] K. El Malki, P. Calhoun, T. Hiller, J. Kempf, P. McCann, A. Singh, H. Soliman, and S. Thalanany. Low latency handoffs in mobile IPv4. *Work in Progress*.

[20] T. ETSI. 101 350. *General Packet Radio Service (GPRS); Overall Description of the GPRS Radio Interface, Stage 2 (GSM 03.64 Ver. 6.4. 0 Rel. 1997)*, 1999.

[21] N. Fikouras, K. El Malki, S. Cvetkovic, and C. Smythe. Performance of TCP and UDP during mobile IP handoffs in single-agent subnetworks. In *1999 IEEE Wireless Communications and Networking Conference, 1999. WCNC*, pages 1258–1262, 1999.

[22] N. Fikouras and C. Görg. Performance comparison of hinted-and advertisement-based movement detection methods for mobile IP hand-offs. *Computer Networks*, 37(1):55–62, 2001.

[23] N. Fikouras, A. Könsgen, and C. Görg. Accelerating Mobile IP hand-offs through Link-layer Information, an Experimental Investigation with 802.11 b and Internet Audio. *International Multiconference on Measurement, Modelling, and Evaluation of Computer-Communication Systems (MMB)*.

[24] Q. Guo, J. Zhu, and X. Xu. An adaptive multi-criteria vertical handoff decision algorithm for radio heterogeneous network. In *Proc. IEEE ICC05*, volume 4, pages 2769–2773, 2005.

[25] S. Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall, 2008.

[26] H. Holma and A. Toskala. *WCDMA for UMTS: Radio access for third generation mobile communications*. John Wiley & Sons, Inc. New York, NY, USA, 2000.

[27] B. Hongyan, H. Chen, and J. Lingge. Intelligent signal processing of mobility management for heterogeneous networks. In *Neural Networks and Signal Processing, 2003. Proceedings of the 2003 International Conference on*, volume 2, 2003.

[28] S. Horrich and S. Jamaa. Neural Networks for Adaptive Vertical Handover Decision. In *5th International Symposium on Modeling Optimization in Mobile, Ad Hoc, and Wireless Networks*, 2008.

[29] S. Horrich, S. Jamaa, and P. Godlewski. Adaptive vertical mobility decision in heterogeneous networks. In *Proceedings of the Third International Conference on Wireless and Mobile Communications*, page 44. IEEE Computer Society, 2007.

[30] Y. Huang and C. Huang. The integration and application of fuzzy and grey modeling methods. *Fuzzy Sets and Systems*, 78(1):107–119, 1996.

[31] J. Hui-ling, Z. Zhao-yang, and L. Shi-ju. A power threshold based policy for vertical handoff in heterogeneous networks. In *Wireless Communications, Networking and Mobile Computing, 2005. Proceedings. 2005 International Conference on*, volume 2, 2005.

[32] C. Hwang and K. Yoon. *Multiple attribute decision making: methods and applications: a state-of-the-art survey*. Springer Verlag, 1981.

[33] IEEE. *IEEE 802.16 Task Group. Air Interface for Fixed Broadband Wireless Access Systems.*, 2003.

[34] Institute of Electrical and Electronic Engineers. *LAN/MAN Standards Committee of the IEEE Computer Society.*, 2003. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz Band. IEEE Std 802.11b-1999 (R2003).

[35] I. Joe, W. Kim, and S. Hong. A network selection algorithm considering power consumption in hybrid wireless networks. *IEICE Transactions on Communications*, 91(1):314, 2008.

[36] D. Johnson, C. Perkins, J. Arkko, et al. Mobility support in IPv6, 2004.

[37] G. Klir and B. Yuan. *Fuzzy sets and fuzzy logic: theory and applications*. Prentice Hall Upper Saddle River, NJ, 1995.

[38] R. Koodli. Mobile IPv6 Fast Handovers. Technical report, RFC 5268, June 2008.

[39] S. Krishnan, N. Montavont, E. Njedjou, S. Veerepalli, and A. Yegin. RFC 4957, Link-layer event notifications for Detecting Network Attachments (August 2007).

[40] G. Lampropoulos, A. Salkintzis, and N. Passas. Media-independent handover for seamless service provision in heterogeneous networks. *IEEE Communications Magazine*, 46(1):64, 2008.

[41] S. Lee, K. Sriram, K. Kim, J. Lee, Y. Kim, and N. Golmie. Vertical handoff decision algorithms for providing optimized performance in heterogeneous wireless networks. In *IEEE GLOBECOM*. Citeseer, 2007.

[42] K. Leung, B. McNair, L. Cimini, and J. Winters. Outdoor IEEE 802.11 cellular networks: MAC protocol design and performance. In *IEEE International Conference on Communications*, volume 1, pages 595–599. Citeseer, 2002.

[43] H. Maturino-Lozoya, D. Munoz-Rodriguez, F. Jaimes-Romera, H. Tawfik, and R. Nortel. Handoff algorithms based on fuzzy classifiers. *IEEE Transactions on Vehicular Technology*, 49(6):2286–2294, 2000.

[44] J. McNair, T. Tugcu, W. Wang, and J. Xie. A survey of cross-layer performance enhancements for Mobile IP networks. *Computer Networks*, 49(2):119–146, 2005.

[45] A. Misra, S. Das, A. Dutta, A. McAuley, and S. Das. IDMP-based fast handoffs and paging in IP-based 4 G mobile networks. *IEEE Communications Magazine*, 40(3):138–145, 2002.

[46] O. Modeler. OPNET Technologies Inc. *Bethesda, MD. URL: http://www. opnet. com*, 2001.

[47] K. Murray and D. Pesch. Intelligent network access and inter-system handover control in heterogeneous wireless networks for smart space environments. *Proc. IEEE ISWCS04*, pages 66–70, 2004.

[48] M. Nam, N. Choi, Y. Seok, and Y. Choi. WISE: energy-efficient interface selection on vertical handoff between 3G networks and WLANs. In *Proceedings of the 15th IEEE International Symposium on PIMRC04*, volume 1, pages 692–698.

[49] T. Narten and E. Nordmark. W. Simpson, Soliman, H.,“ Neighbor Discovery for IP Version 6 (IPv6). Technical report, RFC 4861, September 2007.

[50] Y. Nkansah-Gyekye and J. Agbinya. Vertical Handoff between WWAN and WLAN. In *Proceedings of the International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies*, page 132. IEEE Computer Society, 2006.

[51] Y. Nkansah-Gyekye and J. Agbinya. Vertical Handoff between WWAN and WLAN. In *Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies, 2006. ICN/ICONS/MCL 2006. International Conference on*, pages 132–132, 2006.

[52] Y. Nkansah-Gyekye and J. Agbinya. Vertical handoff decision algorithm for UMTS-WLAN. In *Wireless Broadband and Ultra Wideband Communications, 2007. AusWireless 2007. The 2nd International Conference on*, pages 37–37, 2007.

[53] M. Oren. Decision making by p53: life, death and cancer. *Cell Death & Differentiation*, 10(4):431–442, 2003.

[54] H. Park, S. Yoon, T. Kim, J. Park, M. Do, and J. Lee. Vertical handoff procedure and algorithm between IEEE802. 11 WLAN and CDMA cellular network. *Lecture Notes in Computer Science*, pages 103–112, 2003.

[55] J. Puttonen and G. Fekete. Interface Selection for Multihomed Mobile Hosts. In *2006 IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1–6, 2006.

[56] R. Ramjee, K. Varadhan, L. Salgarelli, S. Thuel, S. Wang, and T. La Porta. HAWAII: a domain-based approach for supporting mobility in wide-area wireless networks. *IEEE/ACM Transactions on Networking (TON)*, 10(3):396–410, 2002.

[57] H. Simon. Theories of decision-making in economics and behavioral science. *The American Economic Review*, pages 253–283, 1959.

[58] H. Simon. Rational decision making in business organizations. *The American Economic Review*, pages 493–513, 1979.

[59] M. Stemm and R. Katz. Vertical handoffs in wireless overlay networks. *Mobile Networks and applications*, 3(4):335–350, 1998.

[60] M. Stemm and R. Katz. Vertical handoffs in wireless overlay networks. *Mobile Networks and applications*, 3(4):335–350, 1998.

[61] E. Stevens-Navarro, Y. Lin, and V. Wong. An MDP-based vertical handoff decision algorithm for heterogeneous wireless networks. *IEEE Transactions on Vehicular Technology*, 57(2):1243–1254, 2008.

[62] E. Stevens-Navarro, Y. Lin, and V. Wong. An MDP-based vertical handoff decision algorithm for heterogeneous wireless networks. *IEEE Transactions on Vehicular Technology*, 57(2):1243–1254, 2008.

[63] Z. Sun, D. He, L. Liang, and H. Cruickshank. Internet QoS and traffic modelling. *IEE Proceedings-Software*, 151:248, 2004.

[64] F. Tansu and M. Salamah. On the vertical handoff decision for wireless overlay networks. In *Computer Networks, 2006 International Symposium on*, pages 111–115, 2006.

[65] R. Tawil, J. Demerjian, G. Pujolle, and O. Salazar. Processing-delay reduction during the vertical handoff decision in heterogeneous wireless systems. In *Proceedings of the 2008*

IEEE/ACS International Conference on Computer Systems and Applications-Volume 00, pages 381–385. IEEE Computer Society Washington, DC, USA, 2008.

[66] R. Tawil, G. Pujolle, and O. Salazar. A vertical handoff decision scheme in heterogeneous wireless systems. In *Proceedings of the IEEE Vehicular Technology Conference*, pages 2626–2630, 2008.

[67] R. Tawil, O. Salazar, and G. Pujolle. Vertical Handoff Decision Scheme Using MADM for Wireless Networks. *Proc. of IEEE WCNC'08*.

[68] L. Trevino. Ethical decision making in organizations: A person-situation interactionist model. *The Academy of Management Review*, 11(3):601–617, 1986.

[69] P. Vidales, R. Chakravorty, and C. Policroniades. PROTON: A Policy-based Solution for Future 4G devices. In *Proceedings of 5th International Workshop on Policies for Distributed Systems and Networks (IEEE POLICY 2004)*. Citeseer.

[70] J. Wu, C. Cheng, N. Huang, and G. Ma. Intelligent handoff for mobile wireless internet. *Mobile Networks and Applications*, 6(1):67–79, 2001.

[71] L. Xia, L. Jiang, and C. He. A Novel Fuzzy Logic Vertical Handoff Algorithm with Aid of Differential Prediction and Pre-Decision Method. In *IEEE International Conference on Communications, 2007. ICC'07*, pages 5665–5670, 2007.

[72] L. Xia, J. Ling-ge, H. Chen, and L. Hong-wei. An intelligent vertical handoff algorithm in heterogeneous wireless networks. In *Neural Networks and Signal Processing, 2008 International Conference on*, pages 550–555, June 2008.

[73] W. Zhang. Handover decision using fuzzy MADM in heterogeneous networks. In *IEEE Wireless Communications and Networking Conference*, volume 2, pages 653–658, 2004.

[74] F. Zhu and J. McNair. Cross layer design for mobile ip handoff. In *Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st*, volume 4, pages 2255–2259 Vol. 4, May-1 June 2005.

[75] H. Zhu, M. Li, I. Chlamtac, and B. Prabhakaran. A Survey OF QUALITY OF SERVICE IN IEEE 802.11 NETWORKS. *IEEE Wireless Communications*, page 7, 2004.