

Doctor Thesis

Shibaura Institute of Technology



ACO-Based Velocity-Aware Handover Scheme
for Seamless Mobile Communications

March 2016

MUHAMMAD ARIFF BIN BAHARUDIN

Acknowledgments

I feel tremendously lucky to have the opportunity to work with Prof Eiji Kamioka on the ideas in this dissertation. Prof. Kamioka have taught me the love for research and agreed to take me as a graduate student, and encourage me to think creatively and out of the box. He is always beside me, understands my difficulties and works side by side with me in finding the right direction. He incubated me in research method, scientific writing skill, etc. He even sacrifices a lot of his time to help me in solving some of the problems that I have faced. I have never met a professor more generous with his time and experience.

I am grateful to Prof. Hiroaki Morino at Shibaura Institute of Technology (SIT), Prof. Katsunori Yamaoka at Tokyo Institute of Technology, Prof. Takumi Miyoshi and Prof. Koichi Gyoda at SIT for spending some of their time and dedication for serving on my dissertation and oral committee.

I wish to thank all staff and faculties at SIT for their kindly taking care of me for not only on study but also on daily life related issues while studying in Japan. I still remember Mrs. Yukiko Sonoi who helped me to settle the apartment arrangement when I first came to Japan. Also to Mrs. Nishishita Yuko who helped me a lot in dealing with many things in Japan. I am grateful to Mrs. Yukari Moriuchi, Ms Ayumi Maemoto, Ms Ueki Emi and others at the Graduate School Section who always kindly support and advise the procedures related to my research duty at SIT. I would like to thank Prof. Ayao Tsuge, Prof Masato Murakami and Prof Chiaki Nakayama for their great Human Resource Fostering Program which positively influences the way I do this research. I also wish to thank SIT's Hybrid Twinning Program (HBT), Japanese Government (Monbukagakusho: MEXT) scholarship, and Electrical Engineering Faculty (Universiti Teknologi Malaysia) for their supports when I study in Japan.

I would like to thank my friends in Japan, especially Mohd Sabri bin Sinal, whom have helped me a lot in work and life in Japan.

Finally, I would like to thank my family, my parents, my wife, and my little son and daughters for their love and support. They were the pillars that have held me up through my ups and downs in my studies and in life. This dissertation is dedicated to them.

ABSTRACT

In mobile communication environments, handover is a very important process to maintain mobile host's connections to the network. Although a lot of researchers have tried to improve the handover performance, little deployment to satisfy the users can be seen in the real network due to modification costs. In the current age, the existence of network with different Radio Access Technologies has created a heterogeneous environment. Moreover, the mobile hosts available to users are becoming more sophisticated, which opens the opportunity for end-point approaches to alleviate the deployment issues. Furthermore, real time applications such as VoIP and Video Conference are very delay and quality sensitive, but not much research has considered qualitative requirement. Additionally, velocity is also one of the issues for the mobile users, because it can affect the success or failure of a handover process. Hence, a suitable technique is needed to alleviate the deployment issues, cope with stringent requirements of real time applications as well as adapt to the velocity of the mobile users.

Among achieved results, most important ones include: (1) The AntNet-based Handover Algorithm (ANHA) which utilizes the Mean Opinion Score (MOS), calculated using the ITU-T E-Model, and the Received Signal Strength (RSS) information to trigger the handover; ANHA adapts the features in an existing ACO-based approach known as AntNet that selects the best network for the MH via pheromone (updated based on MOS changes) and the heuristic component (RSS); (2) A feasibility study to discern the effectiveness of endpoint centric-multi-home approach and network centric-single home approach, in terms of packet loss and handover latency; (3) A feasibility study to differentiate the signaling efficiency between endpoint centric approach and network centric approach; (4) Discussion on the issue of velocity and proposing a simple velocity estimation model and adaptive threshold method based on the velocity to avoid handover failure and unnecessary handover.

Each of the proposed solution and studies mentioned above was evaluated using simulation. The feasibility studies justifies the choice of components used in the proposed method, ANHA. Evaluation results reveal the effectiveness of the proposed method compared to existing methods. For example, ANHA is capable of maintaining a higher average MOS level of the VoIP call compared to the existing method in most of the case studies. ANHA especially excels when the difference in the MOS value of the existing networks are large. Consequently, this thesis opens a new direction in the studies of handover process improvements.

TABLE OF CONTENTS

CH.	TITLE	PAGE
	ACKNOWLEDGEMENTS	i
	ABSTRACT	ii
	TABLE OF CONTENTS	iii
1	INTRODUCTION	1
1.1	Most referable works	4
1.2	Overview of the Ant Colony-based Handover and Velocity-Awareness Approach	6
1.2.1	Considered Topology	7
1.2.2	Ant Colony-based Handover Approach	8
1.2.3	ANN-based Velocity Awareness	11
1.3	Contributions and Thesis Organization	12
2	LITERATURE REVIEW	15
2.1	Network Layer Solutions	15
2.1.1	Mobile IP (MIP)	15
2.1.2	Mobile IPv6 (MIPv6)	18
2.1.3	Mobile IP Regional Registration	19
2.1.4	Hierarchical Mobile IPv6 (HMIPv6)	21
2.1.5	Fast Handover for MIPv6 (FHMIP)	23
2.1.6	Cellular IP (CIP)	25
2.1.7	Handoff-Aware Wireless Access Internet Infrastructure (HAWAII)	27
2.1.8	Terminal Independent Mobility for IP (TIMIP)	29
2.2	End-to-end Solutions	31
2.2.1	TCP Migrate	32
2.2.2	TCP Multi-Home (TCP-MH)	33

2.2.3	Session IP diversity based Generalized Mobility Architecture (SIGMA)	34
2.2.4	MSCTP	36
2.2.5	Cellular SCTP	37
2.2.6	Endpoint Centric Handover (ECHO)	38
2.2.7	Host Identity Protocol (HIP)	40
2.3	Bio-Inspired Algorithms	42
2.3.1	Basic Ant Colony concept based on stigmergy	42
2.3.2	Ant Colony Optimization	44
2.3.3	AntNet	46
2.4	Other Important Information related to this work	49
2.4.1	VoIP Codecs	49
2.4.2	E-Model Mean Opinion Score	51
3	EFFECTIVENESS OF USING ENDPOINT CENTRIC, MULTI-HOME ENVIRONMENT	54
3.1	Introduction	54
3.2	Comparison between Multi-Home (HIP) and non-Multi-Home (MIPv6, HMIPv6, HAWAII, CIP, and TIMIP)	55
3.3	Packet loss rate and handover latency comparison results	56
3.4	Justification on selecting HIP-based approach instead of implementing SCTP-based approach (comparison between endpoint centric multi-home environment)	59
3.5	Concluding Remarks	60
4	SIGNALIN COST COMPARISON BETWEEN ENDPOINT CENTRIC AND NETWORK CENTRIC APPROACHES	61
4.1	Network structure considered in the signaling cost comparison	61
4.2	Notations	63
4.2.1	Notations that apply to both HIP and HMIPv6 signaling cost modeling	63
4.2.2	Notations that apply to only HIP signaling cost modeling	63
4.2.3	Notations that apply to only HMIPv6 signaling cost modeling	64
4.3	Signaling Cost Analysis of HIP	64
4.3.1	HIP Location Update cost	65
4.3.2	HIP Binding Update cost	65

4.3.3	HIP Packet Delivery cost	66
4.3.4	Total Signaling cost of HIP	67
4.4	Signaling Cost Analysis of HMIPv6	68
4.4.1	HMIPv6 location update cost	68
4.4.2	HMIPv6 packet delivery cost	69
4.4.3	HMIPv6 total signaling cost	70
4.5	Result and signaling cost comparison of HIP and HMIPv6	71
4.5.1	Impact of number of MHs for different network (subnet) residence times	73
4.5.2	Impact of average number of communicating CH and location update transmission cost	74
4.5.3	Session to Mobility Ratio	75
4.6	Concluding Remarks	75
5	ANT COLONY-BASED HANDOVER APPROACH	77
5.1	AntNet-based Handover Algorithm (ANHA)	78
5.1.1	Ant deployment and Data collection by Ants	80
5.1.2	Decision Parameter Calculation	81
5.1.3	The proposed ANHA	83
5.2	How ANHA operates	88
5.2.1	ANHA initialization	88
5.2.2	Handover decision using ANHA	88
5.2.3	Packet Duplication process	89
5.2.4	Overview of handover decision	89
5.3	Simulation	91
5.3.1	Sequential Topology, AP2 delay = 10ms, AP1 AP3 delay: 10ms-500ms	93
5.3.2	Sequential Topology, AP2 delay = 450ms, AP1 AP3 delay: 10ms-500ms	99
5.3.3	Overlapped Topology, AP1 delay = 10ms, AP2 delay: 10ms-500ms	104
5.3.4	Overlapped Topology, AP1 delay = 450ms, AP2 delay: 10ms-500ms	109
5.3.5	Handover triggering timing comparison	114
5.4	Discussion on Different Possible Scenario	116
5.5	Concluding Remarks	118

6	MOBILE HOST VELOCITY AWARENESS	119
6.1	Introduction	119
6.2	Main Concepts and Notations	122
6.2.1	Main Concepts related to and used in the proposed methods	122
6.2.2	Travel Distance Estimation Method	123
6.2.3	Overview of the ANN technique	126
6.3	Velocity Estimation Method (VEM)	127
6.3.1	Utilization of RSS and Δ RSS in estimating MH's velocity	128
6.3.2	The proposed ANN for velocity estimation improvement	129
6.3.3	Evaluation of the proposed ANN-based VEM	131
6.4	Proposed Warm-up Time	134
6.4.1	Obtaining the warm-up time	134
6.4.2	The effectiveness of using the proposed warm-up time approach	140
6.5	Adaptive RSS threshold	142
6.6	Concluding Remarks	145
7	DISCUSSIONS, CONCLUSIONS AND FUTURE WORKS	146
7.1	Discussion and summary	146
7.2	A use case scenario	149
7.3	Conclusions	151
7.4	Future Works	152
	REFERENCES	153-161

CHAPTER 1

INTRODUCTION

When considering mobile networks, the handover is a very important process in maintaining Mobile Host (MH) connection(s) to the network. Especially in the current age, a lot of users own at least one or more mobile devices such as smart phones and mobile tablets which are on a par computational capabilities with laptop or desktop computers. This is because of the usefulness, ubiquity and comparably affordable price; consequentially inciting the development of wireless access technologies. Hence, the rapid evolution of RAT (Radio Access Technology) has been done including some new systems such as WiMAX Worldwide Interoperability for Microwave Access) [1] and LTE (Long Term Evolution) [2] which have been deployed nearly worldwide, adding more spice to the pot of the currently available wireless access networks such as the legacy cellular networks and the IEEE802.11 based WLANs (Wireless Local Area Networks) (see figure 1.1) The integration of these wireless access networks has structured a heterogeneous environment consisting of these RATs with different coverage sizes overlapping with each other as shown in figure 1.2. Since most of the mobile devices are capable of connecting to at least two different RATs (i.e. WLAN, cellular, WiMAX, LTE and so forth), there is a growing demand for ubiquitous access to the Internet via various different RATs

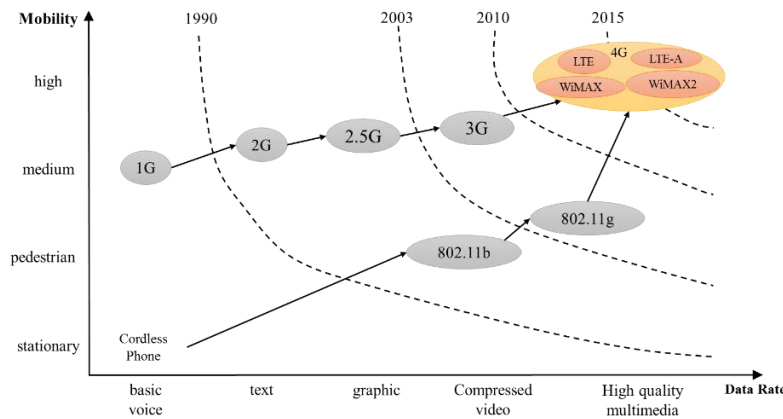


Figure 1.1 Wireless communications evolution from 1G to 4G. The data rate and mobility have increased from generation to generation. Introduction of new Radio Access Technologies in 4G adds more network selection to the mobile hosts.

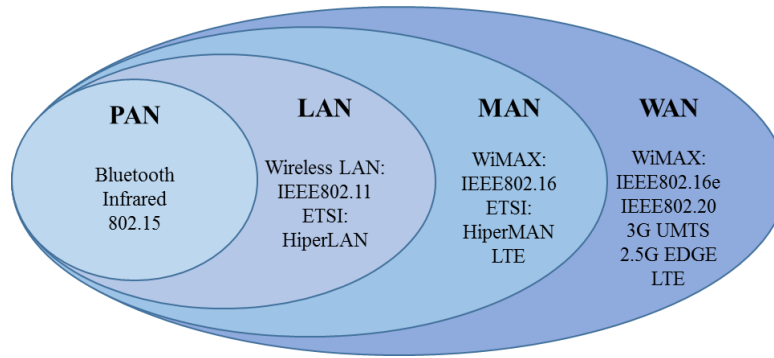


Figure 1.2 Heterogeneous wireless access networks which include a variety of wireless RATs overlapping with each other. These networks can be classified into four different groups based on the size of their coverage area.

Furthermore, with the high capabilities of the existing mobile equipment, the users gain access to a slew of applications that require the Internet connection. The connection requirements for these applications vary; simple applications such as messaging, e-mail and browsing are usually delay tolerant and does not require very high computational needs. On the contrary, real time applications, in particular, Voice over IP (VoIP) [3] and Video Conference, have very stringent requirements due to its delay sensitiveness and its qualitative requirements, not only for the Quality of Service (QoS) of the network, but also for the Quality of Experience (QoE) of the users. Hence, to efficiently utilize the existing heterogeneous networks while maintaining the QoS and QoE, a suitable handover approach is required; in this case, a vertical handover (VHO) approach, which is necessary for multiple different RAT must be well-considered.

The vertical handover process is basically divided into three phases; the Handover Information Gathering Phase (HIGP), the Handover Decision Phase (HDP), and finally the Handover Execution Phase (HEP) as illustrated in figure 1.3. A lot of studies on the HDP have been reported which decide the where and when to trigger the handover. The surveys reported in [4] and [5] have shown, many parameters such as received signal strength (RSS) and other parameters including network contexts, user preferences, mobile terminal's mobility and other VHO contexts. The most important information for real-time applications like VoIP is the expected level of users' QoE, but not many researchers have considered this information in the HDP.

Additionally, from the literature, the HEP can be divided into two parts which are the Handover Management (HM) and the Mobility Management (MM). There are several types of HMs such as Mobile-controlled handoff (MCHO) where the MH is the one that decides and triggers the handover, Network-controlled handoff (NCHO) where the handover is totally controlled by the network equipment such as the Base Station and Mobile-assisted handoff (MAHO) where the handover is initiated and decided by the network entity, but with help and information from the MH [6]. Meanwhile, the MM is mostly depending on mobility management protocols. The popular protocols are Mobile IPv4 (MIPv4) and its variants such as Mobile IPv6 (MIPv6), Hierarchical MIP (HMIP), Fast MIP (FMIP), and Fast Hierarchical MIP (FHMIP). These protocols have been researched extensively, and have shown good performance in research. Nevertheless, even though MIP enabled equipment have been available for some time, not much

deployment of these technologies can be seen in the real network currently [7]. This is because huge modifications have to be done to the existing networks that will incur a very massive cost [7].

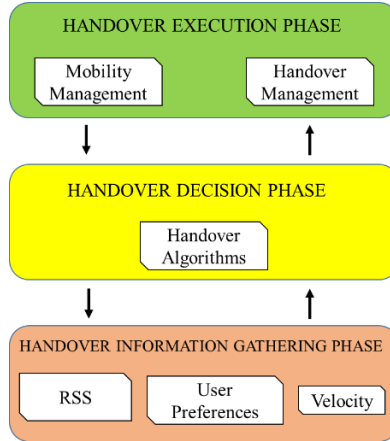


Figure 1.3 Handover management divided into three phases: (i) the Handover Information Gathering Phase (HIGP), (ii) the Handover Decision Phase (HDP) and (iii) the Handover Execution Phase.

Consequently, there is a growing interest in the research communities in studying endpoint-centric approaches to avoid or minimize the alterations in the existing network entities [8]. This approach is possible because of the rapid advancement of mobile hosts; with the capability of the existing and future mobile devices, the complexity of NCHO can be pushed to the endpoints (mobile devices in this case) in order to minimize or totally avoid network modifications. Several existing methods currently available are the the **H**ost base **a**utonomous **M**obile **A**ddress **T**ranslation (HaMAT) [9], the **S**eamless **I**P diversity based **G**eneralized **M**obility **A**rchitecture (SIGMA) [10-12] and the **E**ndpoint-Centric **H**andover (ECHO) [13-15]. These methods are very similar in terms of the implementation albeit using different protocols. HaMAT uses Mobile Address Translation based on the network layer, whilst both SIGMA and ECHO uses SCTP (these are the MM components). The similarity is that all of them use multi-homing (where the mobile nodes are able to connect to two or more networks at the same time via multiple antennas of the same RAT or different RATs), the availability of one or more network connections (the term of “path” will also be used in this manuscript) between the MH and its Corresponding Host (CH) before any handover is needed and all of them are obviously endpoint centric. However, in [104], similar to [110], the handover trigger is based on a critical distance which is calculated based on the network coverage, which is implicitly connected to the RSS value. Meanwhile, SIGMA uses only RSS value to trigger the handover, whilst ECHO, which is based on SIGMA, enhanced the triggering with an additional QoE factor, but implemented only static thresholds for the QoE and RSS value which restrict the method.

Finally, the handover process is only useful if it can be completed before the MH loses its connection to the current serving AP. Failure in doing so will cause a connection breakdown. This is especially important when dealing with multi-home environment, because the advantage of this

environment is in maintaining connection to the network even during the handover process. If it loses connection before completing the handover process, packet losses will occur, hence, the handover signal will be lost, and it will take time to recover or to reestablish the end-to-end connection between the MH and its CH causing a lot of packet losses and high handover latency. This issue comes about due to the MH's lack of ability to detect the right time to trigger the handover process due to the use of static threshold. Most existing methods, including SIGMA and ECHO lack the ability to dynamically adapt to the user mobility and the network parameters in determining the right triggering time to initiate the handover process. Thus, an appropriate algorithm is also needed to overcome this issue.

In a nutshell: the handover algorithm proposed in this study addresses the following issues: (i) dependency on mobility management protocols that requires great modifications to network entities which causes slow deployment due to high network node modification cost, (ii) lack of QoE context awareness in the handover process, (iii) lack of suitable handover algorithm to fully utilize the networks available to the MH and (iv) static handover decision that cannot adapt to the MH's mobility. In the next section, the most referable works in the literature will be discussed in detail.

1.1 Most referable works

The work proposed in this manuscript is closely related to two of the currently existing systems which are SIGMA and ECHO. In this subsection, a brief introduction on these two systems will be presented, to get a gist of how they work and their features. This information will be useful during the discussion on the proposed method.

SIGMA (Seamless IP diversity based Generalized Mobility Architecture) is an architecture proposed by Fu et al. [10-12] to enable multi-homed end-to-end communication environment for mobile users by combining several important components that corresponds to the general handover architecture shown in figure 1.3. When the MH is within the overlapping coverage area of two different Access Points (AP), for example AP1 and AP2, the MH is able to connect to both networks simultaneously, creating a redundant path between the MH and the CH. As discussed in the previous section, to avoid the deployment problems, SIGMA uses the MCHO approach by using Stream Control Transmission Protocol (SCTP) as the mobility management component. However, SCTP itself does not support mobility management, thus they did some modifications by introducing a Location Manager (LM) entity which can also be the DNS server. SIGMA also gains multi-home capabilities from SCTP, using the ASCONF chunks [17] for the MH to update its current location to the LM and its Corresponding Host (CH). Furthermore, SIGMA utilizes the Received Signal Strength (RSS) information to trigger the handover process using classical RSS comparison approach. Let the topology in figure 1.4 be a reference for this discussion. Classical RSS comparison approach compares the RSS level of AP1 to the one of AP2, and when the MH moves beyond the critical point (the point where the RSS levels of AP 1 and AP 2 are the same)

the MH will trigger a handover to AP2. This approach is effective in maintaining MH's connection to the network. However, for real time applications such as VoIP, the important context (e.g. delay, jitter and packet loss requirements) that needs to be maintained is the QoE factor which determines the users' level of satisfaction.

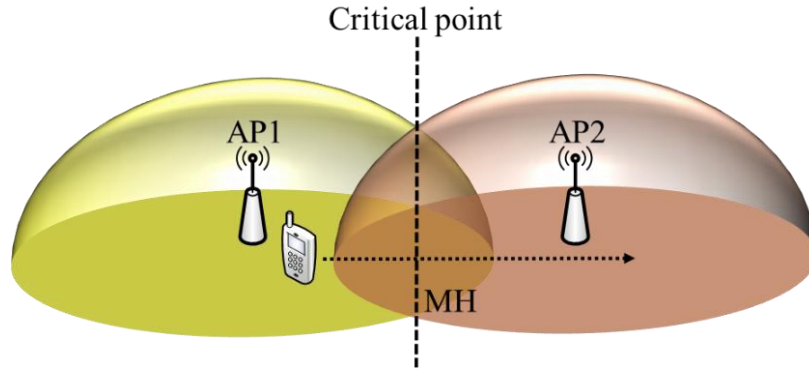


Figure 1.4 Reference scenario for discussion on SIGMA and ECHO. The MH is currently attached to AP1, and moving towards AP2. The handover trigger time depends on the handover strategy of each approach.

ECHO (Endpoint-Centric Handover) [13-15] is developed based on SIGMA. ECHO enhances the performance of SIGMA by considering the user experience level via Mean Opinion Score (MOS), introduced in the E-Model [16] by the International Telecommunication Union – Telecommunication Standardization Sector (ITU-T). In E-Model, the value of R-factor (which has a value between 0 and 100; higher value means better performance or experience) is calculated based on several parameters, in particular, the delay, jitter and the packet loss rate. This R-factor is then mapped to the MOS which is a scale with values between 0 and 4.5 (higher means better). In ECHO, when the MH moves past the critical point (still referring to figure1.4), the MH will send a packet duplication request to the CH. When CH receives this message, it will start sending the same packets to the MH via both available paths (via the connection on AP1 as well as the connection on AP2). The handover is initiated if the alternative network has a MOS level above the configured threshold. If the MOS level is lower than the threshold, MH data flow remains at AP1. Then, when the RSS of the MH becomes lower than the configured threshold, MH will still handover to AP2, regardless of the low MOS level in order to completely avoid losing connection completely.

It can be summarized that SIGMA is the base system, which provides multi-home environment and uses only RSS information to trigger the handover. Enhancing SIGMA, ECHO included a way to monitor the QoE level of a network in order to maintain the quality level experienced by the users, and used static QoE and RSS threshold to trigger the handover. The effectiveness of these two systems have been proven in the literature [10-15]. However, there are limitations to the existing system, due to the limited usage of context (only RSS in SIGMA) and limited usage of handover triggering mechanism (only static threshold values to trigger the handover in ECHO) that can still be improved. The study in this dissertation aims to alleviate these

limitations by incorporating a new handover decision and triggering algorithm to improve the performance during the handover process.

Since in the mobile network, the network parameters, such as RSS, delay and jitter, will be changing constantly (e.g., due to the movement of MH or due to the traffic changes in the network) the MH needs to be able to adapt to these changes in deciding the handover timing. Thus, the work proposed in this manuscript utilizes the RSS value and the QoE information which is obtained using the MOS information, similar to ECHO but with different algorithm for the HDP. The proposed method uses the concept of Ant Colony Optimization (ACO) [18,19]} as the HDP algorithm. Furthermore, an Artificial Neural Network (ANN) based velocity estimation method is proposed to dynamically adapt the MH's handover triggering to its velocity. A summary of the features of these two systems and the proposed approach are as shown in table 1. The overview of the proposed methods will be discussed in the next section and will be described in full in Chapters 5 and 6.

Table 1 Feature comparison between the proposed method and the most related works.

	SIGMA [10-12]	ECHO [13-15]	This research
Mobility Management	SCTP	SCTP	HIP
Handover Decision Contexts	RSS	RSS, MOS	RSS, MOS, Velocity
Handover Decision trigger	RSS comparison	RSS threshold, MOS threshold	RSS+MOS → Ant Colony Optimization-based algorithm, Velocity Awareness → dynamic RSS threshold, unnecessary handover and handover failure avoidance.

1.2 Overview of the Ant Colony-based Handover and Velocity-awareness approach

This section will discuss the proposed Ant Colony-based Handover approach and the ANN-based Velocity-awareness approach that will adapt the handover trigger to the MH's velocity. Before jumping into the discussion on the overview of the proposed methods, a short discussion on the topologies considered in this work and the reasons for consideration will be stated in the following sub-section.

1.2.1 Considered Topology

Throughout this manuscript, there are two main network topologies considered for the discussion and the simulation which are: (i) Sequential Topology (ST) and (ii) Overlapped Topology (OT). These topologies are explained here for ease of discussion in the following sections and chapters of this manuscript.

Sequential Topology: Trains, especially for long distance travel (e.g. shinkansen or high speed train), are widely used. These trains will traverse, not only urban areas, but also rural areas when migrating from one city to another. In some of these rural areas, not many variety of networks will be available, and the APs are deployed in a well-organized sequential way to fully utilize the network coverage. Usually, a sequential topology as shown in figure 1.5(a) is opted. In this research, this topology is used to imitate this scenario. These networks can be of the same or different RAT. In this work, nevertheless, vertical handover is assumed, regardless of the type of RAT.

Overlapped Topology: Another topology considered in this research is the Overlapped Topology as shown in figure 1.5(b), which is a topology normally found in urban areas. This topology came into existence due to the deployment of heterogeneous RAT, with little or no coordination, by service providers, by companies for localized usage and by individual persons in their home or personal workstations. Usually, the accessibility of these networks depends on the owners of the APs. In this work, however, it is assumed that MH are able to utilize these networks as target APs during the handover process.

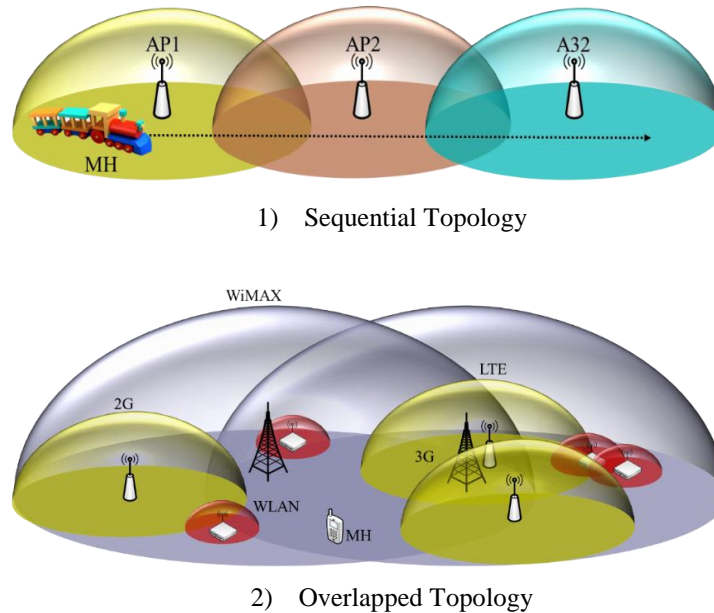


Figure 1.5 The two topologies considered in this research: (a) Sequential Topology, this topology imitates the network topology for trains when moving in rural areas, where the APs are deployed along the railroad; (b) Overlapped topology, this topology imitates the network when the users are in urban areas, where the different RAT overlaps with each other.

1.2.2 Ant Colony-based Handover Approach

As discussed in section 1.1, the handover process can be divided into 3 parts which are the information gathering phase (HIGP), the handover decision phase (HDP) and the handover execution phase (HEP) [4]. In the previous section, SIGMA and ECHO have been discussed, and it can be clarified that the information used (which is RSS only) in SIGMA is not sufficient to maintain the QoE of VoIP applications, thus the improvement has been done by ECHO. However, there is a significant issue in ECHO, that is to say, a simple threshold approach is opted to trigger the handover. If either the current network or the target network has low MOS, but still above the threshold, the MH will not be able to fully utilize the current existing networks when the RSS of current connected network is below the RSS threshold. For example, considering a network as shown in figure 1.4, let the MOS level of AP1 be 4.5 and the MOS level of AP 2 be 3.2, it can be clearly seen that AP 1 has better level of MOS. Nevertheless, if the MOS threshold is set to 3.1, the MH will immediately connect to AP2 due to this threshold, instead of staying at AP1 until the RSS value is deemed too low.

This becomes more apparent if an overlay topology is considered. Let the MH be connected to a network with a large coverage area as shown in figure 1.5(b), and it is moving towards the network with the yellow coverage labelled 3G. Let the MOS level of the current connected network be 4.5 and the 3G network be 3.2 as the previous scenario using the same threshold of 3.1 will cause the MH to handover to the 3G network even though the current network has better MOS. This will lessen the overall QoE that the user will experience. Thus the proposed work in this study is developed to alleviate such issues, as well as the issues stated in section 1.1.

The proposed ANHA scheme is developed base on the generalized handover structure as shown in figure 1.3, thus the discussion will also follow the same flow starting from the information gathering phase (HIGP). ANHA, like ECHO will consider the received signal strength (RSS) as well as the MOS value of the network to determine the best network. ANHA will also consider the velocity of the MH to dynamically adapt the RSS threshold to avoid losing connection. The RSS information will be collected each time the MH receives a beacon message from the APs. The MOS information, meanwhile, is calculated based on the delay, jitter and packet loss of the wireless network which will be obtained using the agents in the proposed Ant Colony Handover algorithm. Finally, the velocity is assumed to be collected using the onboard accelerometer or GPS device that exists on the MH. The information from the accelerometer and GPS device, however, is sometime unavailable when the chip is in sleep mode to save the MH's battery. In this study, a velocity estimation method using ANN is also proposed to accurately estimate the velocity of the MH based on the RSS and the rate of change of MH's RSS.

Moving on to the HDP, a novel handover decision method based on AntNet [19] (a variant of Ant Colony Optimization (ACO)), also known as the **AntNet-based Handover Algorithm** (ANHA) is proposed in order to decide the best network and to trigger the handover process. The processes done in ANHA can be summarized as follows:

1. The proposed algorithm makes use of two types of ants. The Data Ants (DA) are ants that carry the data (in this case, the VoIP packets), and also convey the information on the time stamp information of the ants. The other ant, the Request Ant (RA) is the ant that sends the exploration request. The MH carries the probabilistic table for the networks that MH is connected to (see figure 1.6).
2. When the MH moves into the coverage of two or more APs, the MH establishes connections to the new APs, and sends a Request Ant (RA) towards the CH via the currently used main AP to inform the CH of the existence of alternative paths to MH (see figure 1.7).
3. The ant will pass the notification to the CH and die. When the CH receives the notification, it will start to generate Data Ants (DA) to explore the alternative paths available, and at the same time the DA carries data (see figure 1.8).
4. When the DA arrives at the MH, the ant transfers the timestamp information to the MH and dies (see figure 1.9).
5. The MH will keep track of a path context and pheromone table, and updates the information in the tables with the information from the DA.
6. The amount of deposited “pheromone” depends on the evaluation of the path traversed by the DAs.

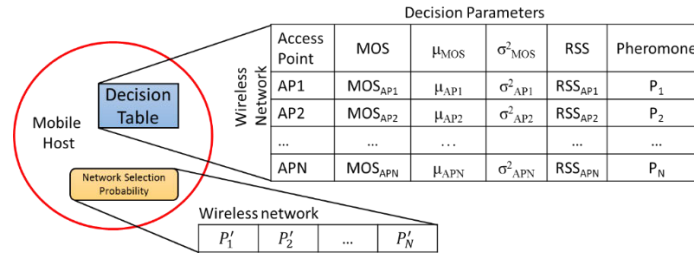


Figure 1.6 Data structures used in MH used to update and decide the handover trigger.

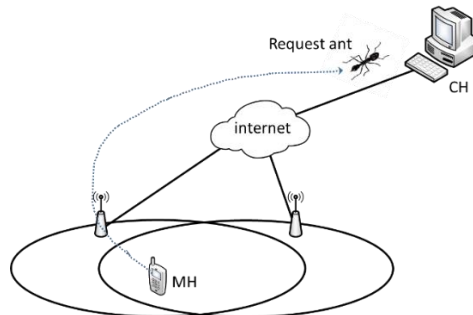


Figure 1.7 Request Ant sent by the MH to inform CH of the new available network and request for duplication packet.

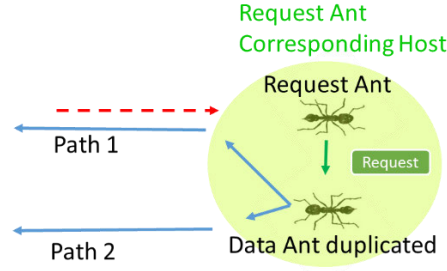


Figure 1.8 The request ant passes the request to the destination node (CH) and dies. The CH starts duplicating packets and sends the data via both paths.

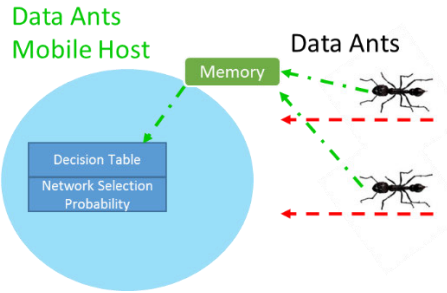


Figure 1.9 The data ants transfer the timestamp information to the MH and die.

Through this process the information needed to calculate the MOS level (delay, jitter and packet loss) can be obtained and updated. Then using the information from the decision parameters, target network will be selected and the handover will be triggered when the desirability of an alternative network, $P'_{alternative}$ becomes higher than the current network's desirability, $P'_{current}$. More details on this algorithm will be discussed in Chapter 5.

Finally, in the execution phase (HEP), ANHA opts the mobile controlled handover (MCHO) management approach to avoid or minimize the network modifications, an end-to-end approach is used. As for the mobility management, the ANHA uses the Host Identity Protocol (HIP) to enable the multi-home environment, as well as to realize the endpoint centric paradigm. In previous endpoint centric approach such as SIGMA, ECHO and HaMAT, there are three key protocol features needed to enable the endpoint centric handover approach: **first**, it requires minimal to none network modification; the **second** is the feature to enable multi-home environment which lets the MH to connect to two or more networks at the same time; and the **third** is the function to dynamically update address change and location change in an end-to-end manner. Host Identity Protocol perfectly fits these descriptions.

The overall architecture of the proposed handover approach is illustrated in figure 1.10 which shows the components used for the development of the handover approach based on the three tiers in figure 1.3. The first tier is the information gathering phase, where RSS, MOS and velocity are collected. Then this information is passed to the second tier which is the handover decision phase, where ACHA processes the information and decides the best network available

and decides when the handover should be initiated. Then finally, the handover process is executed in the third tier, which is the execution phase.

The overall architecture of the proposed handover approach is illustrated in figure 1.10 which shows the components used for the development of the handover approach based on the three tiers in figure 1.3. The first tier is the information gathering phase, where RSS, MOS and velocity are collected. Then this information is passed to the second tier which is the handover decision phase, where ANHA processes the information and decides the best network available and decides when the handover should be initiated. Then finally, the handover process is executed in the third tier, which is the execution phase.

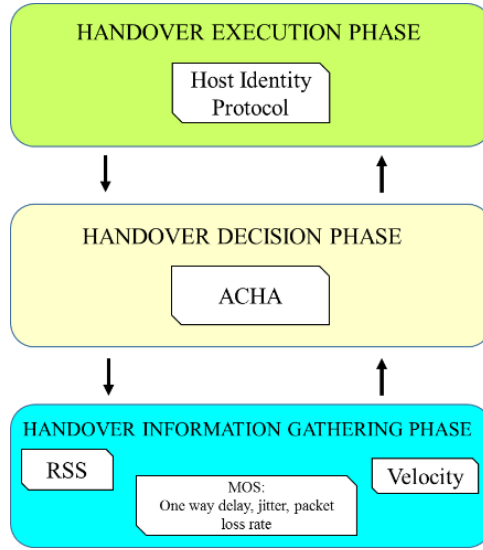


Figure 1.10 The overall architecture of the proposed handover approach with specific components for each phase.

1.2.3 ANN-based Velocity Awareness

ANHA is the main component that will trigger the handover based on the RSS and MOS. However, there are still possibilities of failure, disruptions or waste of network resources due to unnecessary handover [5]. In the future, it is expected that the number of devices connected to the Internet will increase drastically [117]. These devices consists of wireless sensor networks, smartphones, appliances and a plethora of devices; not all of these devices are equipped with the capability to detect its velocity (e.g. via accelerometer or GPS). Thus, a suitable method to detect the velocity of these MH is needed since the velocity information is useful in avoiding unnecessary handover and handover failures.

In this research, an ANN-based Velocity Estimation Method (ANN-VEM) is proposed, where the ANN circuit will be trained with the RSS and rate of RSS change (Δ RSS) to estimate the velocity of the MH and several estimations will be averaged to obtain the final estimation. Then the accuracy of the estimation is further increased via the proposed Warm-up Time, where the estimated velocity during the start of the estimation process is ignored to reduce the overall

estimation error due to inaccurate instantaneous estimations. Then, this information is used to adapt the RSS threshold which will be the safe margin for the MH. This safe margin is to ensure that the MH has enough time to complete the handover process in order to avoid handover failures and disruptions during the handover process.

The complete details of this study is discussed in chapter 6.

1.3 Contributions and Thesis Organization

The key contributions of this thesis is to propose a notable handover approach that utilizes the RSS, latency and velocity information to trigger the handover. Furthermore, the handover and overall quantitative and qualitative performance experienced by the users are tackled (with the use of MOS for the qualitative part). The main contributions of this dissertation can be summarized as follows:

1. A notable handover method is proposed to enhance the handover performance for multi-homed environment with consideration of the user satisfaction. More concretely, contributions in this subject are as follows:
 - a. A new bio-inspired algorithm, the ANHA, based on AntNet is proposed in deciding the target network selection. This algorithm uses the normalized RSS value as well as the network latency information to determine the optimal handover triggering instance for the MH. This method improves the overall performance experienced by mobile users
 - b. ANHA also provides the mechanism to obtain the network status information. This mechanism calculates and smoothes the end-to-end delay and jitter, and also collects the packet loss rate of the network available to the MH for the MOS level estimation. This MOS estimation is used as one of the parameters for the handover triggering. This method adds the qualitative aspect in the handover decision to maintain the quality of experience (QoE) for users (especially for VoIP application).
 - c. An alternative velocity estimation approach using ANN to estimate the MH's velocity is proposed. This method is quite simple for obtaining the velocity information and enhances the accuracy of the velocity when the GPS device is not available.
2. A solution for handover in high velocity scenarios is investigated.
 - a. The effects of velocity to the overall performance is investigated, and an approach to adapt to the effects of MH velocity is proposed. This feature improves the handover performance and avoids connection breakdowns.

The other contributions made in this thesis can be summarized as follows:

- The implementation of ECHO module into simulation using OMNeT++ and its INET module.
- Implementation of the proposed handover approach in simulation using OMNeT++ and INET module.

The organization of this thesis is graphically illustrated in figure 1.11 and is described as follows:

Chapter 1: Introduction. The introduction to the background and the motivation of this research were described in this chapter. A brief overview of the proposed method was also introduced in this chapter. The primary contributions of this research were also concretely summarized in this chapter.

Chapter 2: Literature Review. This chapter reviews the existing mobility protocols and handover approaches that are related to the work in this thesis.

HIP feasibility study:

Chapter 3: Effectiveness of using multi-home environment. This chapter discusses the features of HIP and compares the base performance and potential of HIP as a mobility platform. Host Identity Protocol is compared with existing protocols and systems such as MIP and its variants, and the performance and justification of using HIP is clarified.

Chapter 4: Signaling cost comparison between endpoint centric and network centric approaches. This chapter compares the signaling costs of endpoint centric approach (HIP) and network centric approach (HMIPv6). The effectiveness of using endpoint centric approach compared to network centric approach is clarified here.

Proposed Methods:

Chapter 5: Ant Colony-based Handover Approach. In this chapter, the proposed ANHA will be discussed thoroughly. The explanations and discussions on the probing mechanism and the handover triggering mechanism are included here. Furthermore, the performance of the proposed ANHA will be compared to ECHO here.

Chapter 6: Mobile Host Velocity-Awareness. The ANN-based velocity estimation method is discussed and how it can be used to avoid unnecessary handover and avoid handover failure is included.

\Chapter 7: Discussions, Conclusions and Future Works. This chapter discusses the integration

of both ANHA and the ANN-based velocity awareness approach including the use-case scenario. Then this thesis is summarized and concluded with the advantages of the proposed method, its limitations and its remaining difficulties. Finally, research directions of great interest for the future work are drawn out.

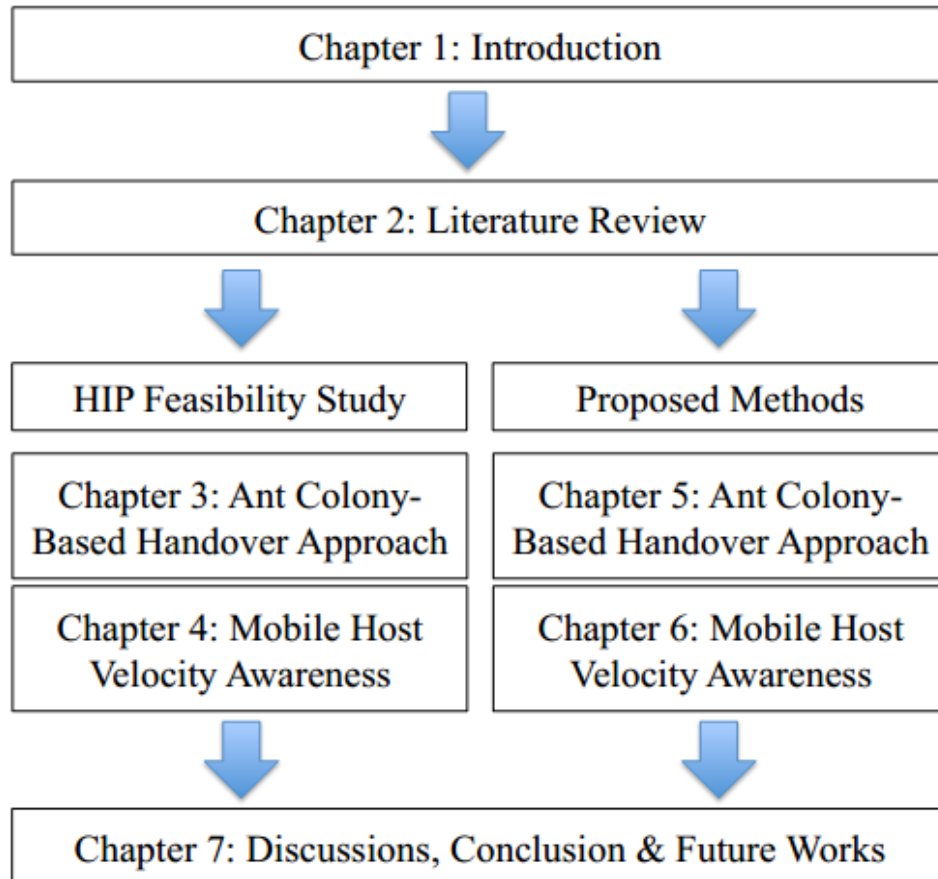


Figure 1.11 The organization of the thesis.

CHAPTER 2

Literature Review

The current available handover mechanisms and protocols with mobility management support will be described in the first section of this chapter. A more detailed description of the most referable works, which are SIGMA and ECHO, is also included in the second section. Then the discussion will move on to the current available bio-inspired algorithms as well as a detailed description of the algorithms adapted to this study, which is the Ant Colony Optimization. At the end of this chapter, some important information (VoIP and Mean Opinion Score) which are related to this work is also included.

2.1 Network Layer Solutions

2.1.1 Mobile IP (MIP)

Conventionally, the IP address of a node is defined uniquely by the network that the node is connected to and also by the physical location of the network. This poses a problem to Mobile Hosts (MH) since an MH may roam between multiple networks, requiring a change of IP address every time the MH connects to a point of attachment. Consequently, this IP address change causes the connections of the MH to be dropped, since higher layer protocols such as TCP rely on the hosts having fixed IP addresses. Thus MIP was developed to alleviate such problems [20-24]. The message flow and the architecture are illustrated in figure 2.1.

A MH in a MIP model has two addresses: (i) the home address (HAddr) and (ii) the care-or address (CoAddr). The HAddr is an address that belongs to the MH home network, which is the IP sub-network that the MH primarily belongs to. This address does not change even though the location of the MH changes (static address). On the other hand, the CoAddr is an address assigned to the MH within a foreign network temporarily. The use of these two types of addresses allows the MH to change its point of attachment, and thus to maintain the reachability through the static IP address (HAddr). It means that the connection will not be disrupted as the MH roams between networks because the IP address changes becomes transparent to the protocols in the upper layers.

For this protocol to work, specialized MIP routers identified as Home Agent (HA) and Foreign Agent (FA) are required in each network where the MH might visit including the MH's home network. These routers are the main entities necessary to map between the MH's HAddr and the CoAddr in the foreign network via binding tables. The MH receives data addressed to the

HAddr through the HA when it is inside its home network. When the MH moves into a foreign network, it will attain a CoAddr in the agent advertisement messages broadcasted by the FA as defined in RFC 1256 [25]. This CoAddr is registered with the HA via the registration request message. Whenever a packet addressed to the HAddr of the MH reaches the HA, the HA investigates if the MH is currently located in a foreign network. When this happens, the HA encapsulates the packet into an IP packet (this is also known as tunneling; IP-in-IP) addressed to the FA. When the FA receives this packet, the FA de-capsulates the packet and forwards the packet to the MH via the CoAddr. The MH still sends back packets via normal IP routing, even if MH is located within foreign network.

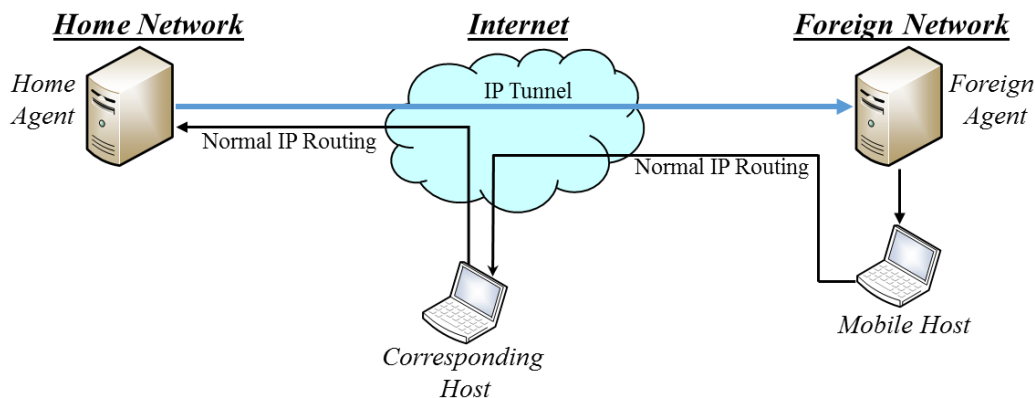


Figure 2.1 MIP architecture and message flow.

There are two ways for a MH to obtain a CoAddr. One way is to wait for an agent advertisement message from the FA as stated previously. The movement of the MH is also detected by checking these periodic messages transmitted by the FA. Another way to obtain a CoAddr is to request in the foreign network by a solicitation message broadcasted by the MH itself, to which the FA can answer with a CoAddr, rather than waiting for an agent advertisement message. When the MH enters its home network again, it sends a deregistration message to the HA, deleting any bindings and CoAddrs, since no mobility support is needed in its home network.

MIP is the most widely accepted IP mobility solution available currently. There is a lot of interest in the research community relating to MIP and many other proposed mobility solutions that utilize MIP functionality. However, there are still a number of problems with it. The most significant problem is related to its deployment. Installation of new network nodes and upgrading or replacing of existing network infrastructure are required. Recent network infrastructure deployed are equipped with MIP functionality, but in most cases, these features are not enabled. This is mainly due to the existing legacy equipment that cannot support MIP, thus it must be replaced before MIP is widely used. Despite the widely available MIP capable network components, very little deployment with the technology has been seen in real situation. The most significant problem is to inflict a very high cost for upgrading a large number of networks when a

network oriented solution such as MIP can become useful. Therefore, there is very little motivation for service providers or companies to use MIP within their networks.

Another major problem is the so called “triangulation problem”. This problem occurs due to the redirection based mechanism, where the packets from a CH (Corresponding Host) has to go through the HA which in turn tunnels the packet to the MH. This triangulation may impose substantial increase in end-to-end transmission delay, being especially inefficient when the MH is receiving data originated from the foreign network where it is currently located. Furthermore, larger packet overheads are induced due to the encapsulation when triangular routing is used; which becomes more critical especially in the case of sending small packets such as the ones used in VoIP.

Finally, MIP also faces the issue of scalability when deploying on a large scale. This is mainly due to the regular binding updates to the HA in order to maintain reachability of each MIP node (i.e. HA, FA, MH and CH). As the number of these nodes increases in a network, the amount of signaling traffic unfavorably affects the network performance, especially in networks with limited bandwidth capabilities such as wireless networks. This proves that MIP does not scale well with growing numbers of MIP nodes.

2.1.2 Mobile IPv6 (MIPv6)

Mobile IPv6 [26, 27] provides some features that improve MIP performance. One of the main enhancements is the routing optimization feature:

Routing Optimization – The triangular routing problem is addressed with this feature. When a MH roams into a foreign network, it directly sends a binding update message to the CH, instead of sending it via the HA, containing its current foreign network particulars such as the CoAddr. When the binding update is completed, the CH can directly send packets to the MH without any support from the HA. This feature improves the scalability of MIP by reducing end-to-end delay and reducing traffic load in the HA.

Another advantage of MIPv6 is the introduction of Stateless Address Auto-configuration (SAA) in IPv6. This feature gives a MH to automatically obtain a CoAddr without the need for FAs. Therefore, less network infrastructure modifications are needed, and thus greatly reducing the barrier to deployment of MIP.

The main disadvantage of using MIP as a mobility solution for real time applications such as VoIP is the handover delays. The delays can be induced when restoring the communication path after a handover, since MIP is a path update protocol rather than a handover management protocol [28]. This delay, which is in the range of a few hundred milliseconds is the time taken for the auto configuration procedure and binding updates to complete, making MIP insufficient for fast handoff support.

In [29], a fast handover mechanism for MIPv6 to address the latency of the auto configuration mechanism has been proposed by the Internet Engineering Task Force (IETF). Link layer trigger, which allows a MH to be informed when a new AP (Access Point) becomes available, is used to improve the auto configuration mechanism. Furthermore, make before break solution is realized through the usage of predictive information about the MHs next point of attachment. In each AP, a neighbor discovery support is implemented giving the MH the ability to obtain new CoAddr via the information on neighboring networks before breaking the connection with the current AP. The MH will be able to use the new CoAddr immediately when connected with the new AP, thereby reducing handover delay as the MH already has the new CoAddr.

With the fast handover support, MIPv6 can fulfil the delay and loss requirements of VoIP, nevertheless, it still faces the deployment problem for future heterogeneous network. For MIP to be useful, a wide deployment of MIP supported devices must be available in a large number of existing networks. This issue has subdued the growth and rollout of MIP.

2.1.3 Mobile IP Regional Registration

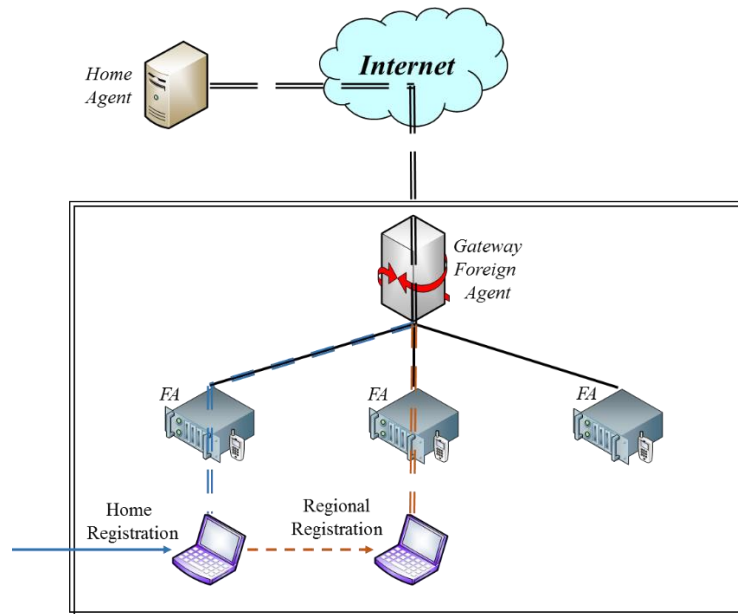


Figure 2.2 Mobile IP Regional Registration.

In MIP, when an MH enters a new access network, it obtains a new CoAddr and must register this address with its HA. However, if the visited network and the home network are geographically far apart, the signaling latency between the two networks can be very large. Thus, the end-to-end delay between the MH and its CH is a significant issue in triangular routing.

By adding an extra layer of hierarchy into the MIP infrastructure, Mobile IP Regional Registration (MIP-RR) [33] aims at solving this problem. A new network entity called the Gateway Foreign Agent (GFA) is introduced at the top level of the hierarchy, where the GFA connects a number of FAs within a predefined region (see figure 2.2). This approach is not limited to only two levels of hierarchy; more levels can be added accordingly to minimize the signaling overhead in large domains.

When a MH moves into a region controlled by a GFA, it will send a regional registration request via the current FA, which has its current CoAddr, to the GFA. This message is then relayed to the MH's HA. As the MH handovers between FAs in the current region, the MH only has to register to the GFA. The HA will continue to use the GFA as the MH's home address while the MH changes its CoAddr within the scope of the GFA.

This approach reduces the number of signaling required by a MH as it roams in foreign networks. The main advantage of using MIP-RR is the decrease in the handoff delay due to the time required to update the HA with a new CoAddr each time the MH performs a handover. It is especially useful when the MH is roaming in networks that are very far away geographically from its home network. Nonetheless, the improvements from MIP-RR come at the cost of adding more

complexity and infrastructure into the network, and thus this approach faces problems with low levels of deployment and slow implementation.

2.1.4 Hierarchical Mobile IPv6 (HMIPv6)

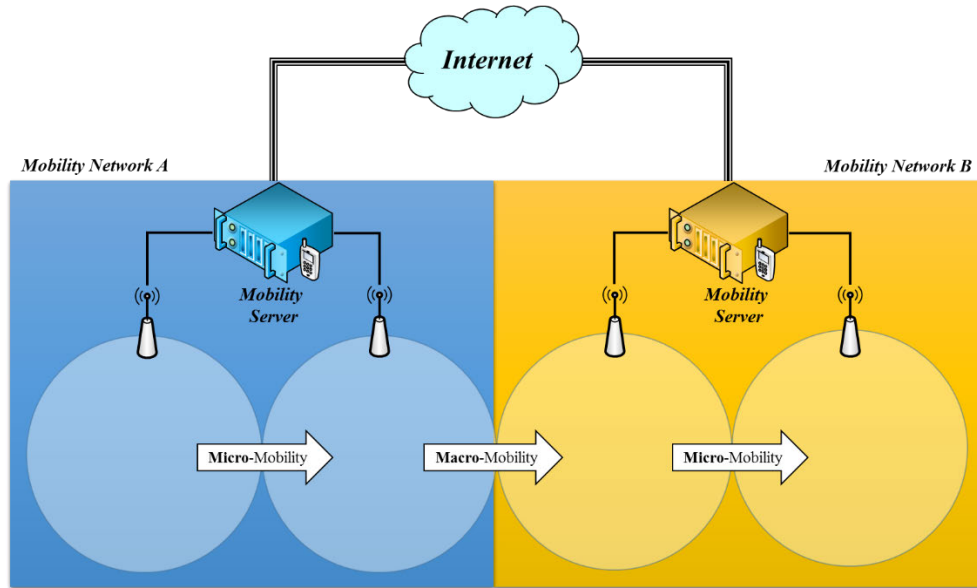


Figure 2.3 Hierarchical Mobile IP architecture.

As discussed in the two previous subsections, MIP faces the problem of scalability. The definition of a scalable mobility mechanism here is that it must be capable of having growing numbers of users without overly loading the network. Thus, Hierarchical Mobile IP (HMIP) [30, 31] is introduced to address this issue. HMIP allows for both macro-mobility and micro-mobility via varying levels of hierarchy, normally two levels, but it can be extended to further levels of hierarchy.

Binding updates are periodically sent to both HAs and CNs by MIP nodes. Besides that, binding updates are also sent when the point of attachment of MH changes. Thus, when the number of MH increases, the signaling and processing load of each network will also increase. Around 69% of the user's mobility is confined within their local home network, according to [30]. The HMIPv6 differentiates between local and global mobility by breaking the network down into sub-networks. This method can greatly reduce the signaling load due to the mobility management and improve the handover performance.

The concept of Mobility Network (MNet), a network defined address space for MHs roaming in the site, is introduced in this protocol, where a site is defined as an address space under the control of a mobility server (Mobility Anchor Point). Each MNet must have at least one Mobility Server (MS). An MS has a similar role as a FA in the MNet which is a router that maintains bindings for each MH roaming in the site. Figure 2.3 depicts an example of a HMIP architecture with two MNet and two levels of hierarchy.

In HMIP, each MH has three IP addresses which are, HAddr, a Virtual CoAddr (VCoA) and a Private CoAddr (PCoA). A MH obtains a VCoA and PCoA using SAA when it detects a newly available access network. Then, the MH registers its VCoA with both its HA and any CHs that it is currently communicating with. The MS maintains the bindings between each of the MHs

IP addresses. HMIPv6 provides two types of mobility modes: (i) Inter-site mobility and (ii) Intra-site mobility.

Inter-site handover: When an MH moves into a new MNet, it will obtain a new PCoA, and a CoAddr with the AP that the MH is currently associated with, and a new VCoA, and CoAddr with the MS of that MNet. To make it clearer:

- a) PCoA – AP's CoAddr
- b) VCoA – MNet's CoAddr

During this process, multiple binding updates (BU)s are sent:

- a) **VCoA** and **PCoA** mapping is sent to the **MS**,
- b) **HAddr** and **VCoA** mapping is sent to the **HA** and to all **CHs**.
- c) **HAddr** and **PCoA** mapping is sent to all **CHs**.

Intra-site handover: When the MH roams within a single MNet, the VCoA does not change and only the PCoA changes. This means that less BUs are sent during an intra-site mobility and none are sent across the Internet as all handovers are only done locally. In this mode, the information that the BU sends are as follows:

- a) **HAddr** and **PCoA** mapping is sent to the **local CHs** (CHs within the same MNet, since VCoA does not change, CHs outside the MH's MNet vicinity do not need to know this change).
- b) **VCoA** and **new PCoA** mapping is sent to the **MS**.

HMIP can be extended further to support different levels of hierarchy. The structure is build according to a tree hierarchy, where an MS is needed at each branch of the hierarchy, and the MS at the topmost branch will manage the mobility, which is the Macro-Mobility (Intra-site handover).

Even though HMIP significantly improves the scalability and performance of MIP, it requires even drastic modifications to the infrastructure, inflicting an even greater rollout problems compared to the original MIP.

2.1.5 Fast Handover for MIPv6 (FHMIP)

The main objective the Fast handovers for MIPv6 (FHMIP) [29, 32] is to alleviate the QoS degradation experienced by real time applications during the handover process when using MIP. RFC 4068 [29] describes two different tools to achieve FHMIP; (i) Anticipated Handover approach and (ii) tunnel based handover approach.

Anticipated Handover approach: in this approach, a make before break solution (it means that a MH establishes a connection to the targeted new network and obtains an IP address for that network first before breaking its connection to its current network), by using layer 2 triggers to allow the MN to anticipate a layer 3 handover. The main layer 2 triggers used here are:

- a) Link Up – A connection with a new AP has been established.
- b) Link Down – A connection with an existing AP has been lost.
- c) Layer 2 Handoff Start – Layer 2 handover to a new AP has started.

The MH needs a valid IP address before it can begin transmitting and receiving packets when performing a layer 2 handover. Using the anticipated handover approach, the MH is able to obtain a new CoAddr on the new AP before moving to that AP. When the information on its next point of attachment of the MH is obtained via layer 2 triggers, the MH will then send a router solicitation message to its old AP containing the information about the new AP to which it wants to handover. On receiving this message, the old AP builds a new CoAddr that will be used by the MH in the new AP, and sends this information to both the MH and the targeted new AP to be validated. If the new CoAddr is validated, the MH can complete the layer 2 handover to the new AP and begin using the new CoAddr immediately. To minimize packet losses during and immediately after the handover process, the old AP will redirect and forward all packets addressed to the MH to the new AP. The handover process is completed when the MN informs it's HA of the new CoAddr.

Tunnel Based approach: This approach is the opposite of the anticipated handover approach. The MH does not obtain a new CoAddr when it enters a new access network, but continues to use its existing CoAddr. This is achieved by a bidirectional tunnel between the old AP and the new AP through which packets can be forwarded. When the MH enters a new access network, it performs a layer 2 handoff. Once the MH becomes a resident of the new AP it will send a request for a new CoA along with transmitting other data. The advantage of this approach is that no layer 3 handover is needed and the handover can be completed quickly. However, this approach adds more complexity to the network infrastructure and the APs. Another drawback is that the handover is layer 2 specific. It means that it cannot support a layer 2 independent approach to mobility. It is for this downsides that most work to date has focused on the anticipated handover mechanism of FHMIP.

The handoff latency and packet loss performance of FMIPv6 was studied using NS2 in [31]. This study has proven that FMIPv6 can achieve delay values under 150ms with loss rates of

less than 1%. Nevertheless, this approach still requires extensive modifications similar to the traditional MIP.

2.1.6 Cellular IP (CIP)

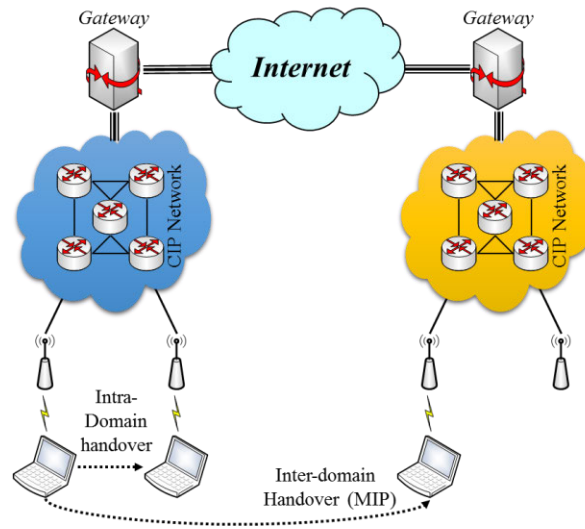


Figure 2.4 Cellular IP Network Architecture.

Cellular IP [34-36] is another approach to IP mobility based on the hierarchical architecture. It incorporates some cellular system design features such as network topology, mobility management, passive connectivity and handoff control, but is based on IP. Cellular IP provides fast and seamless handover control in a limited area or domain for micro-mobility, whilst still using MIP to manage macro-mobility. Some cellular features, which are traditionally located in Mobile Switching Centre (MSC) and Base Station Controller (BSC), are incorporated in this protocol, where each AP behaves like a Base Station (BS).

Cellular IP (CIP) makes use of layer-2 information regarding access point signal strength in order to predict handover, allowing a MH to trigger layer-3 procedures earlier. Unlike HAWAII (refer the next sub-section), in which the MH run MIP, in CIP they must implement CIP procedures. The architecture of CIP is as depicted in figure 2.4. Each CIP domain is composed of a number of CIP nodes structured in a tree topology, having a MIP gateway as the root node. CIP nodes can route IP packets inside the CIP network and communicate with MHs through wireless interface.

The CIP nodes maintain routing and paging caches. The routing caches are used to locate roaming MHs, being updated by IP packets transmitted by the MH. Throughout the CIP node, a chain of temporary cached records is created to provide information on downlink path of packets destined to the MH. After a successful roaming procedure, a CIP node can temporarily have several mappings for the same mobile terminal, leading to different interfaces. Whenever a packet arrives at the CIP node destined to the MH, that packet is sent to all interfaces mapped on the routing cache. Cached mappings must be refreshed periodically by the MH, otherwise they expire and are deleted.

The paging caches are maintained by paging-update packets sent to the nearest access point each time the MH moves. These records are created by MHs that do not send or receive packets frequently.

Within the CIP domain, when the MH approaches a new access point, it redirects its outgoing packets from the old access point to the new access point, updating the routing caches all the way to the gateway. All packets destined to the MH are forwarded to both access points during a time interval equal to the routing cache timeout. After the old path expires, the packets destined to the MH are only forwarded to the new path. As such, when the MH has no packets to send during handover, it has to generate route-update messages in order to allow correct updating of the routing caches. Between CIP domains, normal MIP procedures are used for macro-mobility. It should be noted that in CIP all packets generated within the CIP domain must be routed by the gateway, even if the destination is located in a position adjacent to the source.

CIP provides a mobility solution that incorporates the features of cellular network into the IP environment. The main problem using the CIP for VoIP mobility solution is the presence of packet loss during the hard handoff, and also the delay of using MIP during inter-domain mobility. Also, same as HMIP and HAWAII, major modifications needed on the network to support CIP is also one of the big factors hindering the deployment of such protocol. Further, the CIP requires a large amount of network planning and design, and the need for layer 2 information might restrict the protocol from technology independent approach, which is the current trend of heterogeneous wireless networks.

2.1.7 Handoff-Aware Wireless Access Internet Infrastructure (HAWAII)

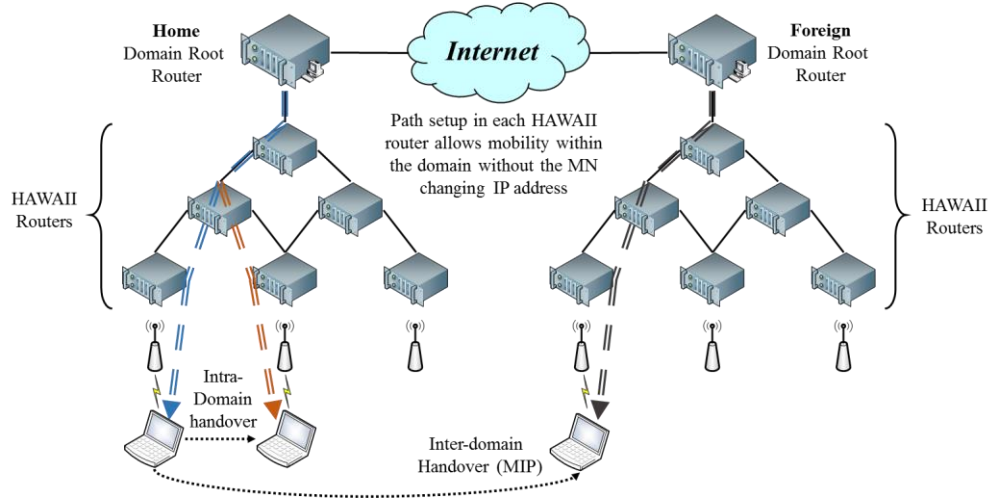


Figure 2.5 HAWAII network architecture.

The Handoff-Aware Wireless Access Internet Infrastructure (HAWAII) [37] was developed to resolve the efficiency and QoS issues of MIP. Mobile IP is implemented in the model with special forwarding entries installed on specific routers, making them aware of specific terminals' location. As such, Inter-domain handover is performed as in MIP, whilst intra-domain handover is performed per-terminal using direct routes (i.e.; the terminal keeps its HAddr as before without any triangulation or IP tunneling). The network architecture of HAWAII is as depicted in figure 2.5.

Each domain in HAWAII consists of a hierarchy of nodes, forming a logical tree. A root gateway called the domain root router [38, 39] takes the role of HA in each domain. Every MH has a home domain and an IP address. The MH's IP address is retained when it moves within a home domain. Packets addressed to the MH are routed to the home domain root router based on the IP subnet address of the domain in the normal way. Then, this packet will be forwarded to the MH via special dynamically established paths. Since each AP behaves as a different FA, the establishment of these paths is triggered when the MH moves between two APs via normal MIP registration messages. These messages create direct routing entries in the intermediate nodes they cross within the home domain. When the MH migrates to a foreign domain, the usual MIP procedure is used, where the FA is now represented by the foreign domain root router, which will assign a CoAddr of the MH and forwards packets to or from the MH.

There are three important characteristics of the path setup and MH specific route in HAWAII:

Firstly, the path setup for intra-domain micro mobility is built from the path setup message sent by the MH, addressed to the domain root router, when it first enters an access network in a domain. When the current domain root router receives this message, it creates a forwarding entry for the MH and forwards the message to its next hop neighbor, which in turn will repeat the same

process. As the message propagates towards the root router, through the network, the propagated message creates a host specific route on the root router and all routers it has gone through. Take note that not every router in the domain will have a record on the MH; only routers on the host specific path has it; thus any routers that does not have route information for the MH will have to send any messages for the MH to the root router.

Secondly, the router within the HAWAII hierarchy that is closest to the MH and is the intersection between two links is known as a crossover router. There are two types of crossover router behaviors during the handover process which are forwarding and non-forwarding. In the case of forwarding, packets are forwarded from the old base station (BS) to the new BS, whilst in the non-forwarding case, the packets are diverted at the cross over router instead.

Finally, the MH routing entries are in a soft state. It means that when the route is inactive, it will be deleted. When the MH moves within a domain, it sends path setup update messages to update routing table entries (that is when moving from one access point to another access point which is still within the same domain). Then the MH sends path refresh messages (while connecting to one access point) to maintain its routing table entries, because if the path entry reaches its timeout, that path will be deleted.

The main difference between HAWAII and cellular IP is that the source to destination number of hops in cellular IP is static, since all routing update packets must reach the gateway. This is not true in HAWAII, because the nodes above the crossover router are not involved during the handover process. Because of this feature, there is less overhead in terms of signaling in the nodes above the crossover router in HAWAII, thus making it more scalable than cellular IP.

There are some problems when considering HAWAII for mobility solution. Firstly, for the protocol to work, all network nodes need to be upgraded to be HAWAII capable. Secondly, although the scalability of HAWAII is far better than MIP, when the number of MHs is large, the root router has to maintain the path state and other host specific information which can cause a bottleneck at the root router. Finally, HAWAII still depends on MIP for inter-domain handover, hence, it suffers the same problems as MIP when it makes a handover.

2.1.8 Terminal Independent Mobility for IP (TIMIP)

Terminal Independent Mobility for IP (TIMIP) [80] is a mobility protocol that aims to avoid the need for any specialized IP layer signaling. For example, in all the IETF protocols for IP mobility, MHs are required to have mobility-aware protocol stack before any kind of mobility management features can be implemented. Thus, legacy devices that are not equipped with such protocol stack will not be able to benefit from the mobility protocols. TIMIP takes a reverse approach, instead of having a mobility protocol stack on the MHs, it aims to provide a mobility approach where the MH does not need any modifications to have the mobility management abilities.

The most important entity in this approach is the Access Network Gateway (ANG), which is similar to HAWAII's domain root router and CMIP's domain gateway. For a MH to be recognized by the network, it must be registered offline (like registering for a new phone from the service provider) with specific information such as the MAC address, IP address, MIP capabilities, IP address of the MIP home agent, authentication key and authentication option. This information is kept by ANG. The MIP capability specifies whether MIP is required; if the legacy MH does not have MIP capabilities, it will be surrogated by the ANG (surrogate MIP), if it does have a MIP stack, then the MIP will be implemented on the MH itself. Once this data is configured at the ANG, it is forwarded to the APs, so that they will be able to know the IP address of the newly associated terminals based on their MAC address provided by layer-2.

Figure 2.6 shows the establishment of routing path in a TIMIP domain. The routing path establishment is reminiscent of HAWAII's route establishment. (1) When a MH first enters an AP within a TIMIP's domain, it will register its MAC Address, where (2) the AP in turn will search for a matching IP address from the registration information broadcasted by the ANG. Then (3 – 5), the registration message is propagated towards the ANG and the route establishment is completed when the message reaches the ANG.

Meanwhile, Figure 2.7 shows the micro-mobility procedure. When the MH moves from the old AP to the new AP within the same ANG domain, the first four steps are the same as the routing path establishment, where the MH registers to the AP, and the registration message is propagated. However the following steps are a bit different. When the message arrives at a cross over access router (AR) (the AR that belongs simultaneously to the old path and to the new path), the path establishment is completed, and the old path must be deleted. The cross over AR will send a RoutingUpdate message to the AR on the old path that will continue to propagate the message until it reaches the old AP. As the message propagates, the old path's entry is also deleted at each AR and at the AP of the old path.

As for the macro-mobility, TIMIP is similar to HAWAII and CIP, where it relies on MIP for the handover process. The ANG becomes the HA for MHs whose home network is the ANG. For foreign legacy MHs, the ANG implements surrogate MIP whilst for MHs that supports MIP, ANG works as the FA.

From the description of TIMIP, the performance mostly will be similar to HAWAII, since they have similar registration and handover processes. The only difference is the support for legacy

devices that does not have MIP protocol stack. Moreover, the drawbacks are also the same as HAWAII and CIP; due to the reliance on MIP, this approach will also face the drawbacks of using MIP for macro mobility. Furthermore, the deployment problem will also persist since a lot of network nodes that need to be implemented before this method can be used.

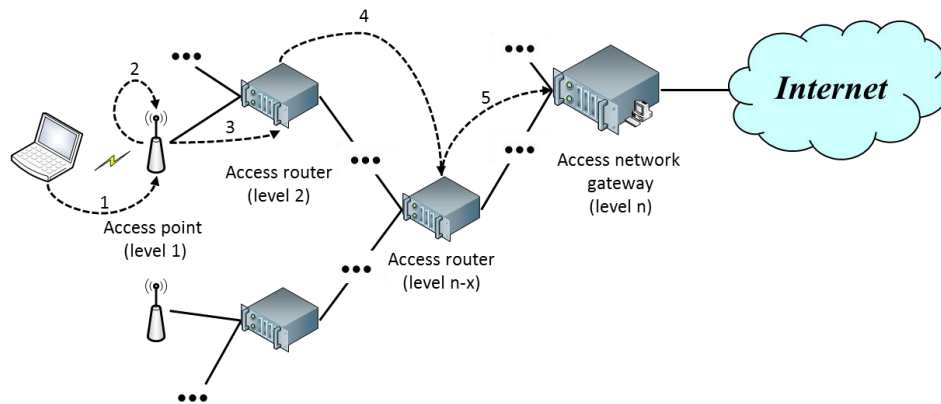


Figure 2.6 Establishment of routing path in a TIMIP domain.

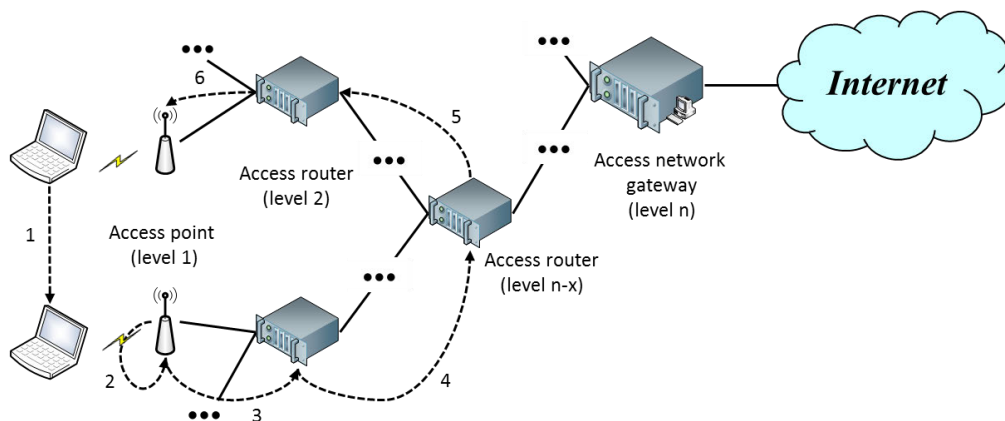


Figure 2.7 Routing configuration during handover process.

2.2 End-to-end Solutions

Many of the existing mobility solutions proposed require a combination of architectural and infrastructural upgrades before they can be applied in the real world. This hinders the deployment of such solutions as can be seen in the case of slow rollout of MIP.

To circumvent this issue, there have been an increasing interest in approaches that requires no network support, namely, end-to-end approaches. The end-to-end methods shifts the complexity from the network to the mobile hosts, providing the advantages of requiring less or no network modifications and making the deployment of such methods easier (compared to network based approach). Most of these endpoint centric solutions operate above the network layer (such as transport layer), thus it requires only end host protocol updates to be implemented.

A number of advantages such as route optimization, integration of multiple interfaces to provide seamless handovers, and independence of additional network infrastructure can be gained [40]. However, the major disadvantage of using higher layer is the lack of inherent support for location management; but this can be solved exploiting higher layer network infrastructure such as the Domain Name System (DNS).

Although there are several candidates for implementing higher layer mobility; the most promising approaches are transport layer based mechanisms such as TCP based or SCTP based. However, recently another protocol which is introduced between the network and transport layers, which is known as Host Identity Protocol (HIP), is becoming more popular due to its unique features, such as the separation of the duality of the current IP address function and added security features [55]. In terms of its performance, it has comparable performance as SCTP based approaches and enhanced network based approaches such as CIP [76-78]. The advantage of HIP is further solidified by the study done by Mugga et al. [79], where in their research, the re-homing time (considering a multi-home environment with two paths between MH and CH, re-homing time is the time needed to recover to the alternative path when the primary path fails) is half of that obtained using SCTP. Thus, it is in the interest of the current research to implement the handover algorithm on HIP.

2.2.1 TCP Migrate

TCP Migrate is one of the approaches towards end-to-end Internet host mobility. The main focus of this approach is on the session continuity without reestablishment of the TCP session due to the change in the IP address for a handover. This solution leverages DNS support of dynamic updates to achieve location management.

Basically, a TCP connection is identified by the n-tuple, which consists of the IP address pair and the port pair at the source and the destination (usually written as [Source IP | Source Port | Destination IP | Destination Port]). The TCP connection will be unreachable if any of these four values changes.

TCP Migrate introduces a new option in the SYN packet that identifies the packet as part of a previous TCP connection. During the TCP connection establishment initialization, a token is negotiated using a migrate permitted option. This setting allows a TCP connection to be identified by the source address, source port and the token.

An open TCP connection can be resumed from a new point of attachment with a new IP address with these parameters. This feat is achieved by sending a migrate TCP packet containing the token to the server or other endpoint. This endpoint will then use the source port/address pair as the new destination port/address pair for the continuation of the TCP connection. The TCP connection can then be resumed while retaining the connection states (e.g., transmission sequence number and control window).

The DNS is updated dynamically via the secure DNS update protocol to allow host location tracking. The MH sends a DNS update message to its home server after it changes its point of attachment and has successfully transferred the TCP connection, thereby enabling it to be located at the new IP address.

The location information present at the DNS can become stale. It means that a node is no longer available at the address specified by the DNS. This is especially true for MHs that change their point of attachment frequently. For this reason, the Time to Live (TTL) field for all DNS name that address mappings for HMs is set to zero. It means that all mapping information is not cacheable by any foreign DNS, therefore eliminating stale bindings.

Currently this approach is only applicable with TCP, rendering it unusable with real time applications such as VoIP. Further, this solution can suffer from significant handover delays due to DNS migration and update delays. On the bright side, it does not require any network infrastructure modifications where only the TCP stack has to be upgraded.

2.2.2 TCP Multi-Home (TCP-MH)

In the current era, most devices are equipped with multiple Network Interface Card (NIC) which gives a MH the ability to connect to more than two different networks. Nevertheless, many higher layer protocols do not directly support multi-homing. TCP Multi-Home (TCP-MH) [43] was developed to address TCP's inability to support multiple IP addresses per TCP session. This is realized by introducing a new set of TCP options which require minimum modification to the existing standard without compromising the existing features.

During the TCP setup session between two nodes, SYN packets containing an MH-Permitted option are exchanged between the two nodes to advertise the multi-homing capability to each other. After successfully establishing connection, a MH-ADD option can be used to add multiple IP addresses as valid addresses for the current TCP session. Then when an address is not reachable anymore, or the connection to a network is down, that address can be removed from the session using the MH-Delete option.

The IP address used to initially establish the TCP session is used as the primary IP address for communication. The alternate address (if any) will be used after multiple Retransmission Time Out (RTO) due to connection loss or network failure on the primary IP's network.

The most notable advantage of TCP-MH is its direct compatibility with existing applications using TCP. Furthermore, because it is a well-known protocol, the developers can avoid the process of learning a new protocol, which in turn can speed up the development of applications using the new features of the protocol. Unfortunately, TCP-MH does not support mobility, since the target of the authors was simply to add the multi-homing ability with no emphasis on mobility. Moreover, TCP-MH does not handover, but rather fails over after multiple RTO, which can lead to significant packet loss and delay, thus not very suitable for real time applications.

2.2.3 Session IP diversity based Generalized Mobility Architecture (SIGMA)

SIGMA [44-47], which utilizes IP diversity to perform seamless handovers between wireless networks, is a transport layer handover scheme. In this scheme, a MH utilizes two wireless interfaces; one as the primary interface to communicate with the current network, whilst the second interface is used to connect to a new network when one becomes available. The IP address of these two different networks are bounded to each interface, respectively. Furthermore, in this scheme the RSS information of the networks is used to trigger the handover, that is to say, when the RSS of the target network becomes higher than the current network, the handover is initiated.

The Stream Control Transmission Protocol (SCTP) [48] is used in this method to enable the multi-homing feature in the hosts. In SCTP, a single association can span multiple IP interfaces regardless of the underlying technology. When an SCTP association is established between two endpoints, one IP address (one Interface) will be determined as the primary address. All communications between these endpoints will be routed through this address until a failure is detected or an upper layer specifically requests the use of an alternate IP interface.

Using the SCTP Dynamic Address Reconfiguration (DAR) [49] addresses can be dynamically added and deleted from the association. The DAR extension provides SCTP with a new message type called an Address Configuration Change (ASCONF). Any of the endpoints can inform its peer about the IP addresses from where it is reachable. This can be done dynamically during an active association and is the main feature that enables SCTP to support seamless handovers.

SIGMA Handover Mechanism can be summarized in the following 4 steps (refer to figure 2.8 and 2.9, the numbers corresponds to the steps below):

1. Acquire new IP – When MH moves into a new wireless access network like the one in figure 2.8, the MH can establish a connection to the new AP (AP2) using its second Network Interface Card (NIC) and obtain a new IP address to be bound with that NIC.
2. Then, when the IP address is acquired, it will be added to the association by informing the CN of the new IP address using the ASCONF chunk.
3. The handover decision in SIGMA is based on the RSS comparison between the current serving network and the new network. When the RSS of the new AP (AP2) becomes greater than the existing AP (AP1), the handover will be triggered. To perform a handover, SIGMA has to redirect the data flow between the MH and the CH via the new AP (AP2). The handover is done by sending a Set Primary ASCONF message to the CH containing the new Primary address
4. Finally when the MH moves out of the coverage area of AP1, an ASCONF message is transmitted to the CH containing the Delete IP parameter. This is to delete the old IP address from the association, so that no data is transmitted to the MH via the old access which may no longer be available.

Although SIGMA can meet the handover delay requirements for real time traffic such as VoIP, the handover decision only considers the RSS of each available access network. The RSS does not reflect the quality that will be achieved once a handover has taken place. This results in making non-optimal handover decision in heterogeneous networks.

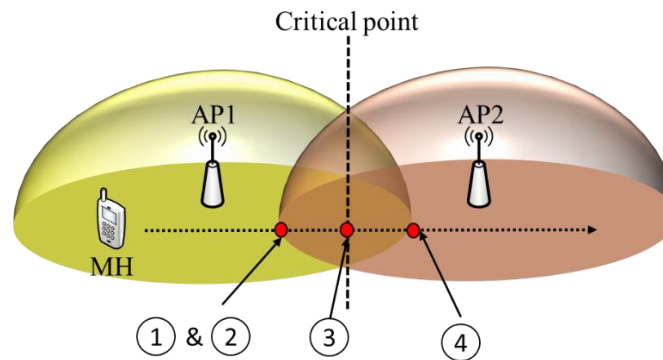


Figure 2.8 Handover procedure for SIGMA in four steps

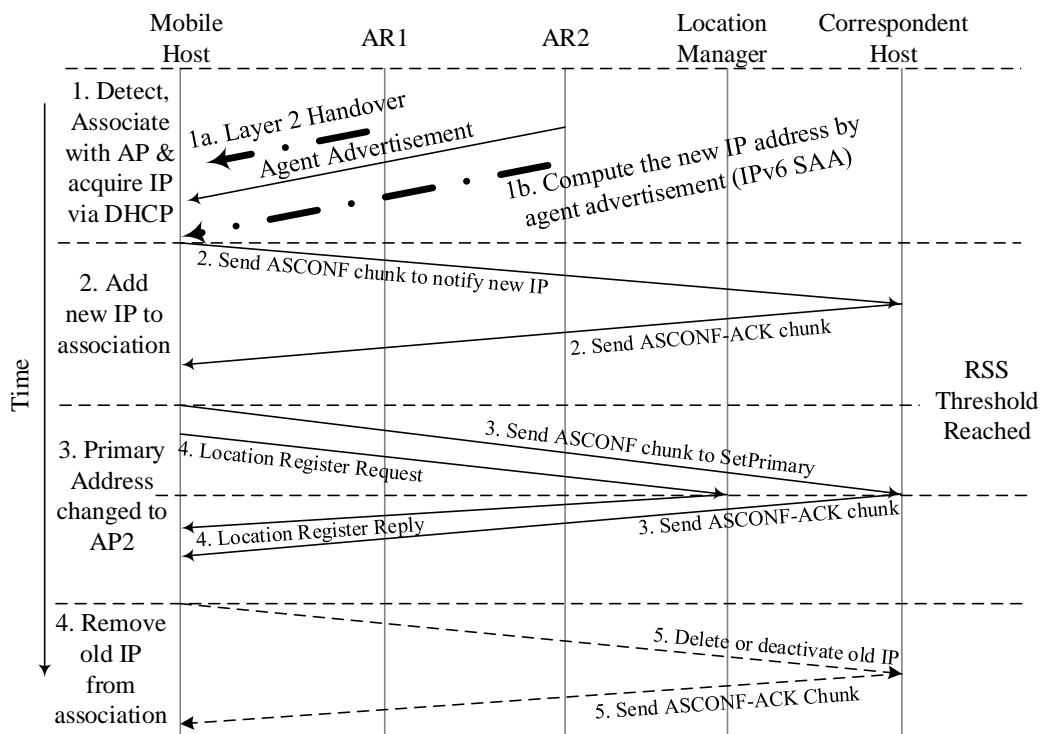


Figure 2.9 Signaling Diagram of SIGMA

2.2.4 MSCTP

MSCTP [50,51] utilizes the IP diversity achieved through SCTP's multi-homing capabilities, having the same handover approach as SIGMA. A MH obtains a new IP address using DHCP or IPv6 SAA when it enters the overlapping area between two access networks. Then, the new IP address is notified to the CH via the ASCONF chunk from the SCTP DAR extensions. Once the RSS level of the current network becomes lower than the new network, the handover is triggered, where the primary address of the MH is changed at the CH. Then when the MH moves out of the old network's coverage area, the old IP address is then deleted. The same process is repeated when the MH moves through wireless network coverage areas.

The main difference between SIGMA and MSCTP is the location management method. MSCTP suggests the use of SIP or MIP to deal with location management but implements a solution using MIP. MSCTP establish an association using MIP, but when the association is established, MIP is no longer used, and only normal SCTP is used; it means that MSCTP can avoid the triangular routing issues or handover latencies of MIP. The CH sends the setup message to the MH via the HA to start the association. The setup message carries the list of all the IP addresses to reach the CH. When the MH receives this message, instead of going through the HA, MH directly sends the acknowledgement message to the CH. This message carries the CoAddr of the MH, through which the CH can directly communicate with the MH without the need for redirection through the HA.

Although MIP is only needed in the initialization phase, MSCTP still relies on the concept of a home network. It means that it still depends on the same network entities as MIP; thus facing the same problems as MIP. Additionally, it also depends solely on the RSS to trigger the handover, like SIGMA.

2.2.5 Cellular SCTP

Cellular SCTP (CSCTP) [52, 53] is an SCTP based mobility scheme developed to offer better performance than MSCTP and SIGMA. This feat is attained using the concurrent multipath transfer feature of SCTP where data is transferred via all the networks currently available in the SCTP association during the handover process.

In CSTP, a MH must have three components: (i) a host agent component to communicate with Access Routers (AR)s, (ii) a CSCTP component (which is actually SCTP with DAR extensions), and (iii) a component to implement the CSCTP handover procedure. The location and mobility management is done using a SIP agent on the application layer of both MH and CH.

The CSCTP, has nearly similar steps to MSCTP in the handover process. The difference is the utilized mechanisms. The MH starts off single-homed. Then, when the MH moves into the coverage area of a new AR, the host agent of MH communicates with the AR using neighbor discovery protocol and obtains a new IP address via DHCP or IPv6 SAA from the AR. Then the host agent informs the CSCTP agent of the new IP address, who, in turn, triggers a handover immediately. The CSCTP creates a new Boolean variable per association called handoff mode, which is set to true when a new IP address is obtained. The MH then uses the DAR extension to add the new IP address to the CH via the Add IP message. When the CH receives this Add IP message the CH also sets its handoff mode variable to true and adds the new IP address to the association.

The most significant difference of CSCTP compared to other SCTP based mobility methods is that at this point, both IP addresses are considered as the primary addresses. The CH duplicates all packets and sends them to both available addresses. Hence, the MH can receive the same data from both currently connected ARs, thereby reducing the probability of packet loss and delays during the handover process. Once the MH decides that the old IP address is no needed or inactive, it sets the handoff mode to false and sends a Delete IP address message to the CH containing the old IP address. The CH, in turn will remove that IP address from the SCTP association, turn the handoff mode to false, and use only the new IP address as the primary.

Despite CSCTPs ability to reduce the probability of lost packets during handover, the handover is initiated without considering any kind of performance indicator. It means that the handover will start without considering how the new network will perform. This is quite risky, and the performance cannot be maximized. For example, if the old network has better performance than the new one, the performance immediately drops when the new network is detected, instead of prolonging the connection time with the old network.

2.2.6 Endpoint Centric Handover (ECHO)

ECHO [13-15, 54] is an end-to-end handover mechanism that is based on SIGMA. The main difference between SIGMA and ECHO is that, ECHO is tailored more towards VoIP. ECHO implements the partial reliable extension and unordered packet delivery of SCTP (PR-SCTP). These are the features which makes the packet sending control more similar to how UDP works, where no retransmission is done when there are packet drops, while maintaining the multi-homing capability and congestion control mechanisms. Furthermore, it also considers the Mean Opinion Score [16], which is a qualitative measurement that signifies the quality of experience received by VoIP users.

The network establishment and the handover process are similar to SIGMA (as shown in figure 2.9). However, the handover triggering decision is different from other SCTP based methods as discussed in the previous sub-chapters. A flow chart of the handover triggering decision is shown in figure 2.10. Referring to the network topology in figure 2.8, ECHO starts checking the MOS level when the RSS level of the current network becomes lower than the alternative network (after point 3 in figure 2.8), then checks if the MOS of the alternative network (AP2) is higher than the MOS threshold (the threshold of 3.1 is selected in [13] which is the second lowest item in table 2.3). If this condition is met, the handover is initiated. If not, no handover occurs and the MH will continue to monitor the RSS of the current main network. Then, if it falls below the RSS threshold (in [13] it is set to -79dBm) a handover is triggered, even though the target network does not support VoIP call. The disadvantage in this flowchart is that the handover will only occur after the RSS of the current network falls below the target network, and the handover is triggered according to the threshold regardless of the situation when the current network still has better MOS level compared to the target network. This is partly due to the late probing and the threshold as well as the fact that the MOS and RSS are considered separately.

Besides that, static threshold might pose a problem; for example, the RSS threshold might be suitable for MHs that are moving at 5 meters per hour, but when the MHs moves at higher velocity, the MH might not be able to complete the handover process before leaving the coverage of the old access point. This will cause packet losses and high handover latency due to the connection loss. Thus, the velocity is one of the important parameters that should be considered in mobile networks.

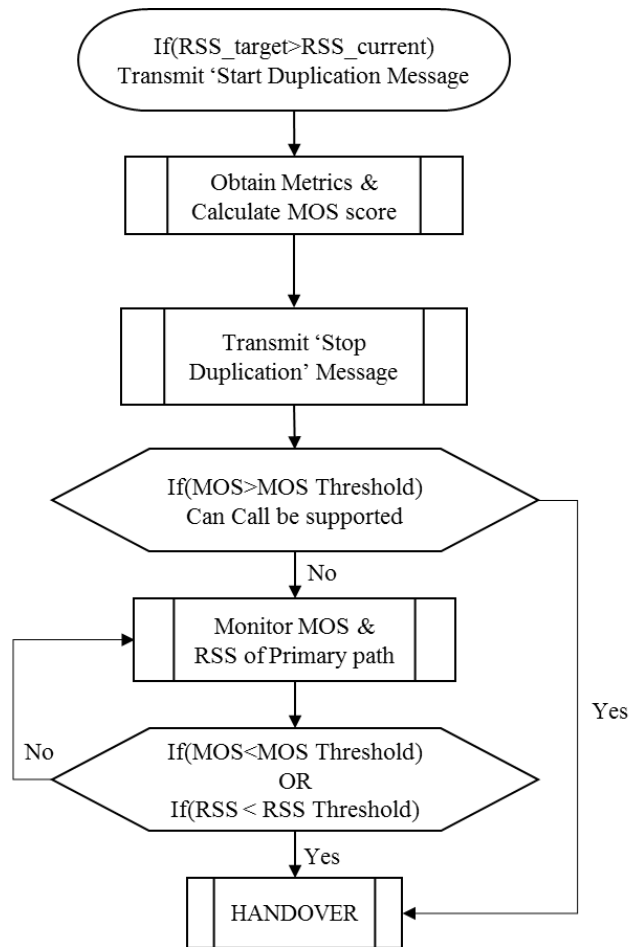


Figure 2.10 The handover triggering flowchart of ECHO.

2.2.7 Host Identity Protocol (HIP)

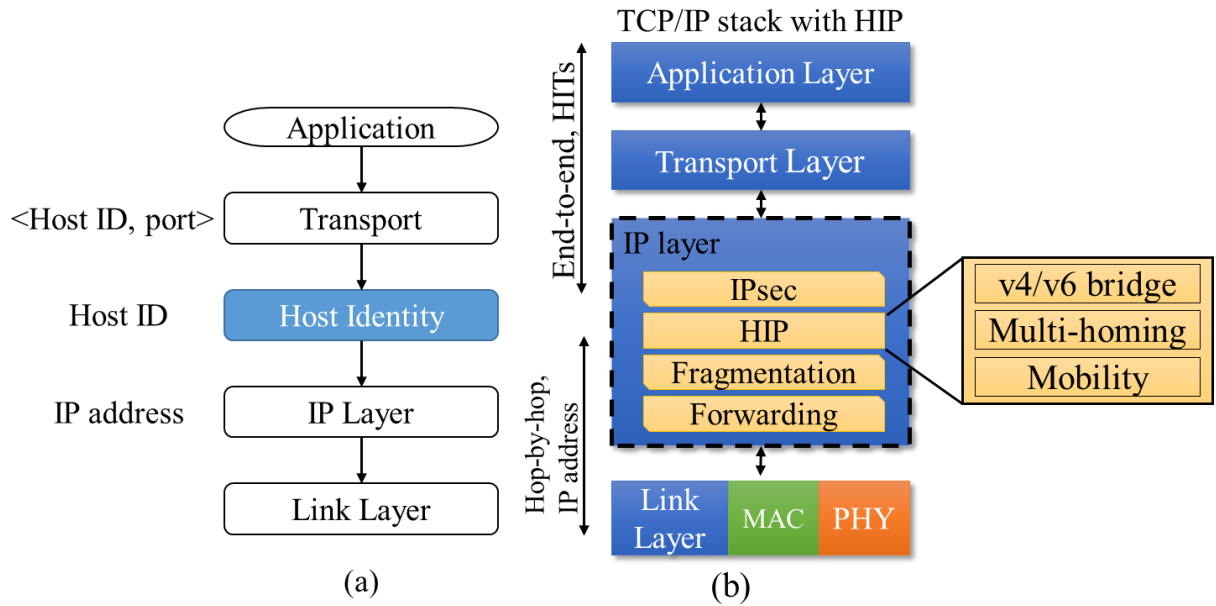


Figure 2.11 The Host Identity Protocol: (a) The idea of HIP, Host Identity is introduced as the identifier. (b) Detailed layering of HIP, provides Multi-homing and Mobility.

The Host Identity Protocol is a quite new protocol which was proposed by Moskowitz *et al.* [55]. In the current Internet system, the IP address not only works as the location address (locators), but also works as the identity or the ID of the end-point hosts (identifiers). As illustrated in Fig. 2.11(a), HIP introduces a new Name Space of Host Identifiers (HI) so that sockets of applications or transport are bound to these HIs instead of the IP addresses. The introduction of HI separates the function of the IP address, where the HI will work as the identifier whilst the IP address will only work as the locator. This is useful in mobile communications because the identifier does not change even though the MH changes its IP address when moving into a new network; making the layers below HIP to become transparent to the upper layers (transport, applications and *etc.*). Furthermore, HIP enables the MH to use multi-homing. It means that the MH will be able to connect to two or more networks at the same time depending on the number of Network Interface Cards (NIC). Besides that, HIP supports mobility management via the rendezvous server (which can easily be the DNS server) and the UPDATE message [56].

The update sequence used by HIP is as shown in figure 2.12. Referring to this figure, when the MH moves into an overlapped area between AP1 (access point 1) and AP2, the second NIC of the MH will associate with the new network. Subsequently, when the IP address for the new network is obtained, the first UPDATE message will be sent to the CH to update the additional IP address. When the CH receives this message, it can quickly start communicating via the new network location but with some restrictions according to [57]. The CH will send back an UPDATE message to acknowledge the first message including an echo request so that when the MH receives this message it will send another message to the CH to confirm the connection to the new network.

When the MH receives the UPDATE from the CH, it can start sending IP packets via AP2. In the proposed method, the multi-homing and the UPDATE message are utilized to obtain the MOS values of the networks that the MH is currently connected to; this will be discussed in detail in chapter 5.

Basically, HIP has all the features for effective mobility management as available in SIGMA and ECHO (e.g., multi-homing, location manager), with added advantages, such as separation of the IP address function. Furthermore, because HIP resides below the transport layer, any type of transport layer can be used with HIP, for example, the UDP can be used for VoIP applications.

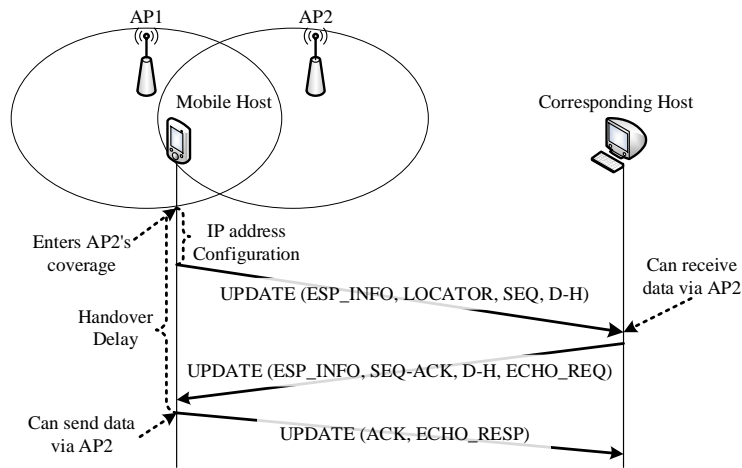


Figure 2.12 HIP update sequence (three-way handshake). The CH can start receiving via AP2 after receiving the first UPDATE. The MH can start sending via AP2 after receiving the UPDATE from CH.

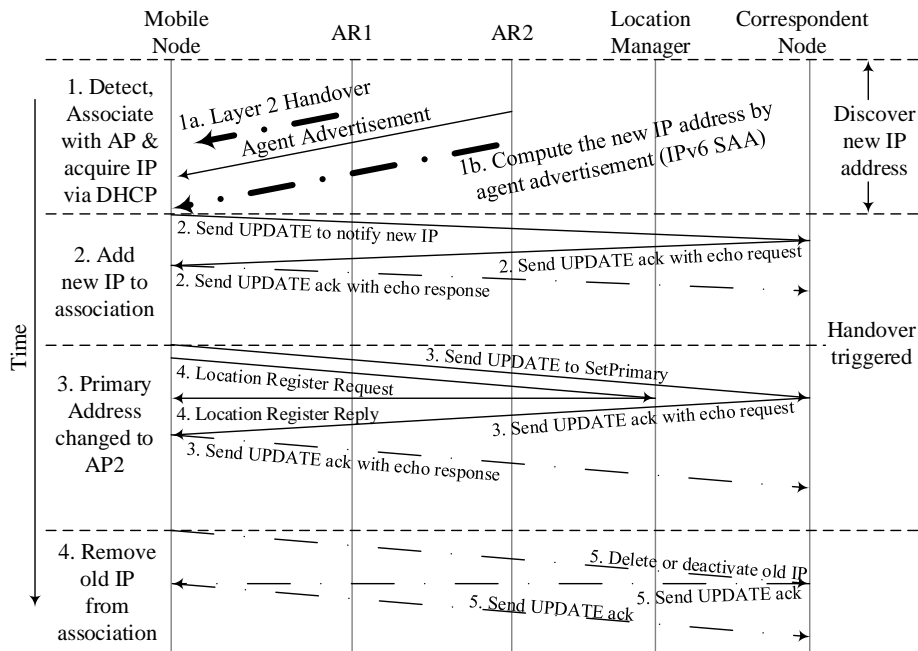


Figure 2.13 Signaling diagram of HIP

2.3 Bio-Inspired Algorithms

Handover decision is one of the most important issues (including handover management, resource allocation and mobility support) related to the heterogeneous wireless networks and should be efficiently addressed for the realization of the envisioned next generation communication system [58]. However, there exist some significant challenges in developing the essential functional components and designing the corresponding algorithms for handover decision such as impracticality of centralized control, dynamic nature, resources constraint, and heterogeneity in the heterogeneous environment [5].

On the other hand, bio-inspired model offers a new approach to design novel powerful solutions for many engineering problems [59, 60]. Comparable challenges faced by the heterogeneous wireless communication system surfacing from dynamic nature, system complexity, heterogeneous architectures and absence of centralized control have been well addressed by the biological system [61]. Many biological mechanisms such as the adaptability to environmental changes, inherent robustness to external perturbation, and self-optimization are appealing to be introduced in the handover decision solution to deal with those aforementioned significant challenges. As some interdisciplinary studies have argued, many biological mechanisms resulting from the evolution of nature over millions of years always go far away beyond the traditional technologies so that they are promising to be used to settle some complex engineering problems [62-64].

2.3.1 Basic Ant Colony concept based on stigmergy

The Ant Colony concept was first introduced by Pierre-Paul Grasse [65] who observed the behaviour of termites which he later called as stigmergy, which is how the insect that produced a stimuli known as pheromones and how other insects from the colony reacts to it. Stigmergy describes this type of communications where the “workers are stimulated by the performance they have achieved” (in this case, by how intense the pheromone is on a specific path).

Later, Deneubourg et al. [66] used an experiment recognized as the “double bridge experiment” to thoroughly investigate the pheromone deploying and following behaviour of ants (see figure 2.14). In this experiment, an ant nest was connected to a food source via two bridges of the same length (refer to figure 2.14). In this scenario, the ants will start to search for food in its surrounding areas. At first, the ants chooses the bridge randomly so that both bridges are used to arrive at the food source. After some time, due to random fluctuations, the pheromone intensity of one of the bridges becomes higher compared to the other, and pulls more ants to go through this bridge. This phenomena further increases the amount of pheromone of that bridge, making it more attractive; hence, after certain amount of time, the whole colony converges to use the same bridge to go to the food source. This colony-level behaviour, which exploits positive feedback, can be used by ants to obtain the shortest route from the nest to the food source.

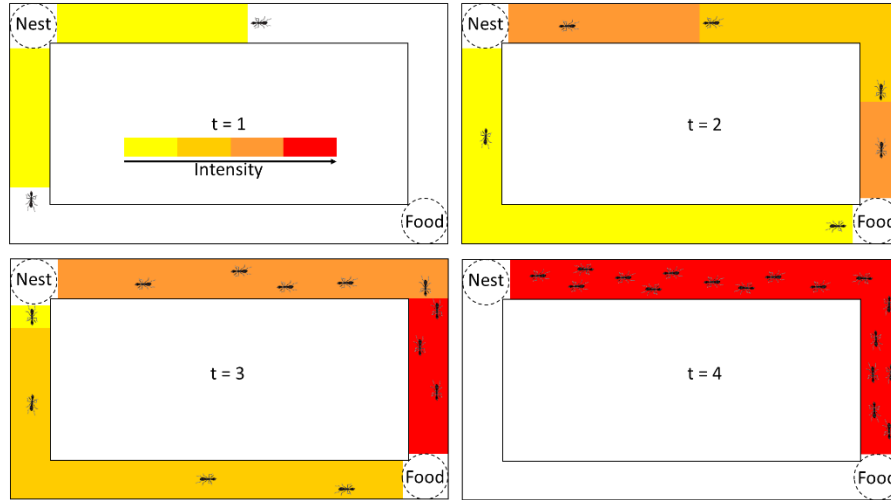


Figure 2.14 The double bridge experiment used by Deneubourg et al. where the lengths of the bridges are equal. As time progresses, due to the random fluctuation of the ants' path selection stochastic, the ants will converge to only one of the bridges.

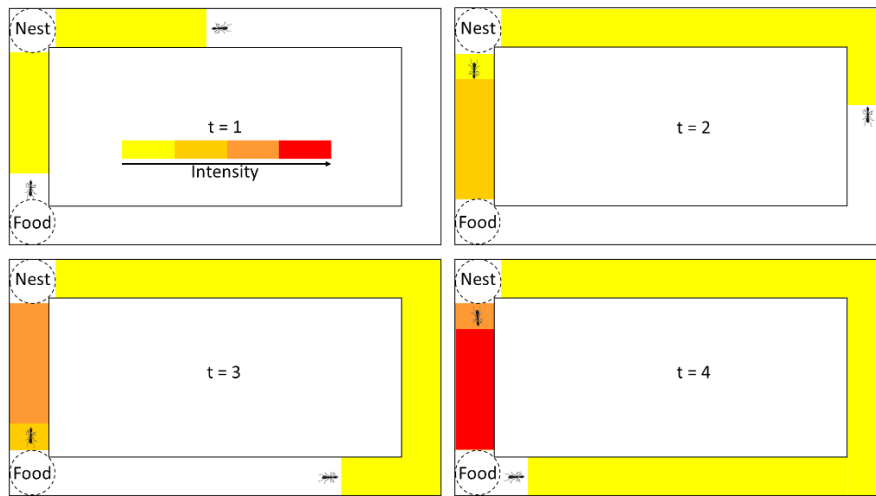


Figure 2.15 The modified double bridge experiment developed by Goss et al. where one of the bridges have different lengths. The two main difference from the original experiment as in figure 1 are, (i) the initial stochastic fluctuation in the initial bridge selection is greatly reduced and (ii)

This experiment was further extended by Goss et al. [67], where they modified the double bridge experiment so that one of the bridges is significantly shorter than the other (as shown in figure 2.15). In this case, the pheromone deployment becomes affected by, (i) the random fluctuations are reduced, and (ii) the ants that go through the first path will arrive at the nest much faster. Thus, increasing the pheromone level of the first path rapidly and increasing its attractiveness much faster than the longer path. Based on the behaviour observed, Goss et al. developed a model to represent the probability of one path which will be chosen by the ants. Let at a moment in time, m_1 ants have used the first bridge and m_2 the second bridge, the probability p_1 of an ant to choose the first bridge is:

$$p_1 = \frac{(m_1 + k)^h}{(m_1 + k)^h + (m_2 + k)^h} \quad (2.1)$$

Where h and k should be fitted to the experimental data; obviously, $p_2 = 1 - p_1$. From [65], a good fit for $k \approx 20$ and $h \approx 2$. The work by Goss et al. is the main source of inspiration in the development of the Ant Colony Optimization by Dorigo et al. [65-74].

2.3.2 Ant Colony Optimization

Ant Colony Optimization (ACO) is a metaheuristic method that utilizes artificial ants to build solutions for the considered optimization problem and the information of these solutions are shared or exchanged using a communication scheme that is very similar to how ants share their information (using pheromones to label desirable paths to food source). The first and original ACO is known as the Ant System (AS), which was developed by Dorigo et al. [68-70].

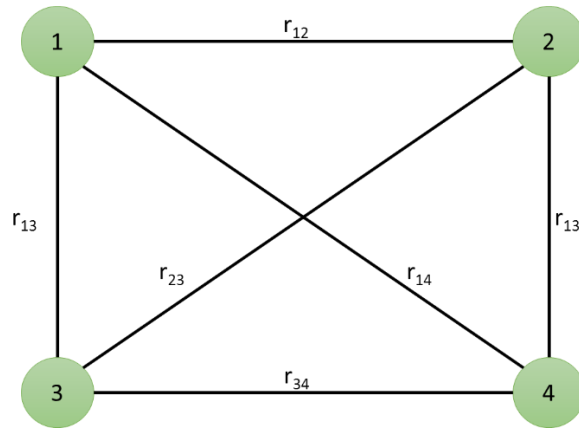


Figure 2.16 Four-city TSP construction graph. In this graph, there are four cities, and each cities are interconnected with each other with different costs (in this case the distance), and the distance between these cities may be symmetric or asymmetric (i.e. r_{12} can have different cost than r_{21}).

Table 2.1 Algorithm 1 – The ACO metaheuristic algorithm.

Algorithm 1 The Ant Colony Optimization Metaheuristic [65]

Set parameters, initialize pheromone trails

while termination condition not met **do**

 ConstructAntSolutions

 ApplyLocalSearch(optional)

 UpdatePheromones

endwhile

Let an area have four cities that are interconnected with each other creating a directed graph as in figure 2.16. Let the green circles be the cities, C_k , that are connected to its neighbouring cities via roads r_{ij} and the distance of r_{ij} and r_{ji} may be symmetric or asymmetric. To easily understand how ACO works, the Traveling Salesman Problem (TSP) is one of the most suitable model, because the problem itself is very similar to the ants' foraging behaviour. The main objective of TSP is to find the shortest "Hamiltonian tour". A Hamiltonian tour is the path that goes through a city once and only once; for example, starting from city 1, the salesman or the artificial ant moves to city 2, 3 and 4 via r_{12} , r_{23} and r_{34} to complete one Hamiltonian tour in the directed graph shown in figure 2.16 (such tour will be known as ant solutions).

In ACO, artificial ant (will be known as only ants after this) builds solution by traversing the graph in figure 2.16 iteratively. Each iteration goes through three phases as shown by the algorithm in table 2.1. After the initialization process, several solutions are constructed by the artificial ants, which then are improved through a process known as local search (this process is optional), and finally the pheromone is updated. Then the process is repeated again in the next iteration and the following ants will be influenced by the change in the pheromone in previous iterations. The pheromone update process governs how the ants generated in the following iterations choose the neighbouring cities during the solution building process. To understand this process in more detail, let us consider the pheromone update process in AS.

In AS, the pheromone values are updated by all of the ants generated for that iteration (which is not necessarily the same in other variants of ACO algorithms). Still referring to Figure 2.16, the pheromone, τ_{ij} , is associated with the road that is connecting the cities i and j . Its value is updated according to the following rules:

$$\tau_{ij} \leftarrow (1 - \rho) \cdot \tau_{ij} + \sum_{k=1}^m \Delta\tau_{ij}^k \quad (2.2)$$

which is a combination of the evaporation process and the pheromone reinforcing components. ρ is the evaporation rate, m is the number of generated ants and $\Delta\tau_{ij}^k$ is the amount of pheromone deployed on road (i,j) by ant k , which is calculated using the following expression:

$$\Delta\tau_{ij}^k = \begin{cases} \frac{Q}{L_k} & \text{if ant } k \text{ used road } (i,j) \text{ in its tour} \\ 0 & \text{otherwise} \end{cases} \quad (2.3)$$

where Q is a constant, and L_k is the length of the tour constructed by ant k .

The expression in (2.2) consists of two important processes in ACO, which are the pheromone evaporation (represented by expression: $(1-\rho) \cdot \tau_{ij}$) and the pheromone deployment (represented by expression: $\sum_{k=1}^m \Delta\tau_{ij}^k$). The pheromone evaporation process is to reduce the attractiveness of paths with low quality solutions (in the case of TSP, the tour with longer distance) and also to avoid the algorithm from being trapped in the local optimal by reducing the attractiveness of all available paths. On the other hand, the pheromone deployment is to increase

the attractiveness of the paths with high quality solutions. These processes are reminiscent of the scheme used by real ants.

From equation (2.3) it can be clearly seen that the roads in shorter tours (smaller L_k) will gain more pheromone compared to roads with longer tours. Thus, iteration by iteration, the ants will converge to the roads with higher pheromone level, which will become the optimal Hamiltonian tour.

2.3.3 AntNet

In this sub-chapter, AntNet [75] is discussed since the proposed method is an adaptation of this algorithm. As have been referred in [65], AntNet is one of the variant of ACO by Dorigo et al. in tackling dynamic optimization problems. AntNet was developed to solve the routing problem in communication networks. The core concepts of this algorithm are: (i) population of artificial ants generate new solutions to the problems by carrying out simulations concurrently and repeatedly, (ii) the ants build the solutions in an incremental way by probabilistic local search, and (iii) the information of previous simulations are exploited to direct future search to better solutions.

In summary, AntNet utilizes two types of ants; the forward ants (FA) and the backward ants (BA). At regular intervals, each network nodes will deploy an FA to a random destination, concurrent with the data traffic. These FAs will move hop by hop from its source node to the destination node and at each node, the FA will select its next hop according to the information stored in each node in order to search for path with minimum cost. At the same time, the ants will also collect the time length, the congestion status and the node identifiers of the followed path (FA goes through the same queuing process of normal data packets making it possible to obtain the congestion status of the node). When the FA reaches its destination, it will transform into a BA and retrace the path that it has gone through to return to its source node (BA moves via high priority queue to quickly update the nodes in its path with the latest information). While moving back, it will update each node in its path with the information of the tour that had done and of its goodness. Finally, when the ant reaches its source node the ant will die.

In detail, AntNet has several core mechanisms:

1. The data structures used in each Network nodes.
2. The probability for a node to create an ant.
3. The probability of choosing the neighbouring node, given a destination node.
4. The pheromone reinforcement and evaporation scheme.

The first one is the data structure contained in the network nodes as shown in figure 2.17. Each network node stores a routing table which contains the probability value expressing the goodness of choosing a neighbour node given a specific destination node, and an array containing the traffic statistics as seen by the node. In simple words, the routing table stores the goodness information of choosing a neighbour node whilst the array stores the information on the duration it takes for the packet to arrive at the destination from the current node. These two data structures is updated according to the information collected by the ants in the FA BA round trip.

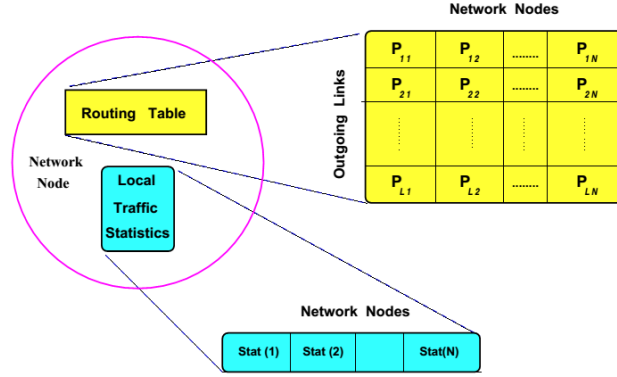


Figure 2.17 Data structures used and updated by ants in AntNet for the case of a node with L neighbours and a network with N nodes. The routing table contains the probability values expressing the goodness (desirability) of choosing a next node for a specific destination. The array contains the traffic statistics of the network from the viewpoint of the network node.

$$P_d = \frac{f_{sd}}{\sum_{d'=1}^N f_{sd'}} \quad (2.4)$$

The second mechanism determines how a node decides to create an ant. Each node creates an ant based on the probabilistic expression shown by (2.4). The probability of a node creating an ant is represented by P_d ; f_{sd} is a measure (in bits or number of packets) of data flow from the source node to the destination node (data flow of $s \rightarrow d$, s for source and d for destination). It means that, the more packets sent from the node to a destination, the higher the probability of an ant to be sent towards that destination.

Next is the mechanism that governs how an ant at node k and moving to destination node d chooses the neighbour node n to move to. The neighbour n is selected using the goodness (probability) P'_{nd} which is a normalized sum of probabilistic entry P_{nd} of the routing table with a heuristic correction factor l_n (the queue length of the n -th link queue of the current node k). P_{nd} is obtained from the following expression:

$$P'_{nd} = \frac{P_{nd} + \alpha l_n}{1 + \alpha(|N_k| - 1)} \quad (2.5)$$

α is the importance weight of the heuristic compared to the pheromone (P'_{nd}). l_n is a $[0,1]$ normalized value proportional to the queue length of the node k with its neighbour n :

$$l_n = 1 - \frac{q_n}{\sum_{n'=1}^{|N_k|} q_{n'}} \quad (2.6)$$

Finally, the pheromone is updated via the evaporation and the reinforcement process. When the FA arrives at the destination node d , it will retrace the nodes that it has passed through. While retracing hop by hop, the BA will update the two data structures in that node with the latest

information about the goodness of the choosing the previous node and the traffic statistics information. The pheromone level of choosing node f (which is the node that the BA comes from) when ant is at node k , $P_{fd'}$ is updated according to the following rule:

$$P_{fd'} \leftarrow P_{fd'} + r(1 - P_{fd'}) \quad (2.7)$$

where r is the reinforcement component, which corresponds to the goodness of the trip time and calculated as:

$$r = c_1 \left(\frac{W_{best}}{T} \right) + c_2 \left(\frac{I_{sup} - I_{inf}}{(I_{sup} - I_{inf}) + (T - I_{inf})} \right) \quad (2.8)$$

T is the ant's trip time, W_{best} is the best trip time experienced by the ants travelling toward destination d , I_{inf} and I_{sup} are estimates of the mean ant trip time. The first term in equation (2.8) evaluates the ration between the current trip and the best trip time observed in the current node (within a specific interval). This term is corrected by the second term that evaluates the stability in the latest trip times. Coefficients c_1 and c_2 weight the importance of each term.

At the same time, the pheromone of other neighbours of node k , nodes n , goes through the evaporation process in accord with the following expression:

$$P_{nd'} \leftarrow P_{nd'} - rP_{nd'}, \quad n \in N_k, n \neq f \quad (2.9)$$

From equations (2.7) and (2.9), it can be seen that every path discovered by the ants receives a positive reinforcement in it pheromone and the reinforcement is a non-linear function of the goodness of the path estimated from ant's trip time. This way, the usage of r is twofold; explicitly, r increases the goodness of the pheromone whilst implicitly, the r also signifies the ant's arrival rate. It means that paths with high reinforcements, independent of their frequency, or the paths with low and frequent reinforcement are trusted as the optimal paths. With these mechanisms, by applying the same concepts as the original ACO algorithm (as shown by the algorithm in table 2.1) the AntNet has shown better performance in network routing optimization compared to other existing methods [75].

2.4 Other Important Information related to this work

2.4.1. VoIP Codecs

VoIP Coder/decoders (Codecs) are algorithms used to encode and decode voice signals between the analogue and the digital domains. There are a variety of different VoIP codecs offering varying levels of compression, bandwidth efficiency and quality. However, all codecs, particularly ones with a high level of compression take time, which adds extra delay between the users.

Table 2.2 shows a comparison of the most common VoIP codecs. The MOS value shown in the table is an indication of the maximum attainable quality of each codec and will be discussed in the following section along with table 2.4 showing how the MOS value relates to user perceived quality.

Table 2.2 Comparison of VoIP Codecs

Codec	Frame Size	Default Data Rate	Sampling Rate	MOS
G.711 (PCM)	0.125ms	64kb/s	8 kHz	4.3
G.723.1	30ms	5.3kb/s	8 kHz	3.65
G.726	Sampling	16/24/32/40kb/s	8 kHz	3.85
G.729	10ms	8kb/s	10 kHz	3.92
GSM FR	22.5ms	12.2kb/s	8 kHz	3.5
GSM EFR	22.5ms	12.2kb/s	8 kHz	3.8

The frame size represents the amount of audio (in terms of time) that is present in each transmitted packet and also how often each packet will be transmitted. The data rate represents the amount of VoIP data that is transmitted per second. The data rate and frame size determine the size in bytes of each VoIP frame.

The sampling rate defines the number of samples of the analogue voice signal taken per second. The Nyquist rate [81] is the minimum sampling rate required to avoid aliasing and it is equal to twice the maximum frequency component contained in the signal being sampled. Voice signals in the PSTN are bandpassed from 100 Hz to 4 kHz as the majority of information of human speech is confined within this range. Based on this, the minimum sampling rate required by VoIP codecs to achieve the same performance as the PSTN is 8 kHz. It is for this reason that the majority of codecs use this rate, some codecs, however, use a higher sampling rate to achieve better clarity.

The most commonly used codec is G.711 (which is also the codec considered for the study in this research) which was standardized by the ITU-T in 1972 [82]. It is the dominant codec used in the PSTN and is supported by all the VoIP systems. G.711 samples voice signals at a rate of 8kHz using logarithmic Pulse Code Modulation (PCM) and does not perform any compression. Other codecs such as G.729 [83] achieve much lower bit rates by creating dependencies between consecutive packets. Although these codecs are more bandwidth efficient, they are more sensitive to loss due to the inter-packet dependencies in their methodology.

Another mechanism employed by some VoIP systems for increasing bandwidth efficiency is silence suppression. Using silence suppression no data is transmitted from a party's handset when they are not speaking. This is achieved by implementing Voice Activity Detection (VAD) algorithms [84]. These algorithms continuously monitor background noise allowing them to determine when a party is speaking. This offers an efficient way for reducing bandwidth usage as during a normal call only one party is speaking at any one time. There are, however, a few problems with using silence suppression. If the receiver hears no background noise they may assume that the call has dropped and the line is dead. To address this many implementations transmit a very low data rate approximation of the background noise. VAD algorithms must also be able to operate quickly in order to prevent clipping. It should be noted that the most commonly used VoIP service, Skype, does not use silence suppression for two reasons. Firstly, transmitting silent packets maintains UDP bindings at NAT. Secondly, if data is being transmitted over TCP the silent packets prevent a reduction in the congestion window size during the silent period [85].

2.4.2. E-Model Mean Opinion Score

E-Model is a model standardized as G.107 [16] by the International Telecommunication Union – Telecommunication Standardization Sector (ITU-T) to estimate the subjective quality of VoIP calls. The score is basically calculated using the equation (2.10), where R is known as the R-factor which signifies the quality level of the call; Ro is the Signal to noise ratio; Is indicates the impairments simultaneous to the voice signal transmission; Id specifies the impairments caused by delay metrics; the Ie is the effects of equipment which is affected by the packet loss rate, and finally A is the advantage factor which is determined by the caller expectations.

$$R = Ro - Is - Id - Ie + A \quad (2.10)$$

The equation can be further simplified by using the default values for Ro , and Is from [16] and A is assumed to be zero, it means that the R is purely affected by the delay context and the packet loss rate. The simplified version of the R-factor equation is as shown by equation (2.11). The Id can be estimated using the method introduced in [23]. The method uses the real time protocol (RTP) one-way delay (obtained by subtracting the timestamp of the packet with the receiving time at MH) and the jitter which is obtained using the method in table 2.5 [24] and filtered using equation (2.12). The Dj is actually dependent on the jitter buffer used in VoIP devices, and usually a limit is configurable by the user; in equation (2.12), the limit is set to 300ms.

$$R = 93.34 - Id - Ie \quad (2.11)$$

$$Dj = \min(\text{codec_frame_size} + (0.9 * RTP_jitter), 300) \quad (2.12)$$

Meanwhile, the value of Ie is determined according to table 2.3, which is obtained from ITU-T standard G.113 [25]. The packet loss is detected by checking the number of missed transmission sequence number TSN, and the percentage of packet loss is calculated by dividing the accumulated lost packet per the number of the current TSN. Then the value is compared to table 2.3 to determine the value of Ie of the network.

Then the R-factor is converted to MOS using equation (2.13) to rate the MOS level of each path between the MH and the CH [16]. The level of satisfaction by the user is listed as in table 2.4, where the score of 4.34 and above indicates the best call quality experienced, whilst the value of below 3.1 means that the quality experienced is very poor [16]. This MOS value will be used in the proposed method as one of the context when deciding the handover trigger timing.

$$MOS = \begin{cases} 1 & R < 0 \\ 1 + 0.035R + R(R - 60)(100 - R)(7 \times 10^{-6}) & 0 < R < 100 \\ 4.5 & R > 100 \end{cases} \quad (2.13)$$

Table 2.3 The values for factor I_e under packet loss conditions for codecs G.711 with Packet Loss Concealment (PLC) considered [25].

Packet Loss% (Note 2)	G.711 with PLC (Note 1)
	Random Packet Loss
0	0
1	5
2	7
3	10
5	15
7	20
10	25
15	35
20	45

NOTE 1 – Speech packet length: 10ms.
NOTE 2 – Here, packet loss is defined as the effective packet loss as valid for the packet sequence at the entrance of the speech decoder. It describes the packet loss as accumulated due to network packet loss as well as packet loss introduced by jitter buffers.

Table 2.4 The Mean Opinion Score experience classification [16].

R Value (Lower Limit)	MOS (Lower Limit)	User Perception
90	4.34	Very Satisfied
80	4.03	Satisfied
70	3.6	Some Users Dissatisfied
60	3.1	Many Users Dissatisfied
50	2.58	Nearly All Users Dissatisfied

Table 2.5 Method to calculate jitter

```
void jitterCalculation(double rcvdTime, double tmStamp, hoParams *path){
//tmStamp = received packet's creation time (timestamp)
//rcvdTime = local system time when packet was received
double delay = rcvdTime - tmStamp;
double jitter = delay - path->delay;
path->delay = delay;

if(path->firstPktRcvd < 2){
path->cumulJitter = 0.0;
jitter = 0.0;
path->delay = delay;
path->firstPktRcvd++;
}
else{
if (jitter<0) jitter = -jitter;
path->jitter += (1./16.) * (jitter - path->jitter);
// the 1/16 is the moving average window size for jitter smoothing.
}
}
```

CHAPTER 3

EFFECTIVENESS OF USING ENDPOINT CENTRIC, MULTI-HOME ENVIRONMENT

3.1. Introduction

In this chapter, the effectiveness of using multi-home environment, using Host Identity Protocol (HIP) particularly, compared to other non-multi-home environment using existing mobility management protocols that have been discussed in the previous chapter (MIPv6, HMIPv6, HAWAII, CIP and TIMIP) is discussed. In Chapter 2, the characteristics and the features of existing non-multi-home environment have been described, and from their reference in the literature, the methods that incorporate hierarchical structure such as HMIPv6, HAWAII, CIP, and TIMIP have shown great improvements compared to the original MIP, where the packet drop rate and the handover latency could be alleviated efficiently. However, as discussed in the previous chapter, these methods face some issues such as scalability problem, domain configuration optimization (to avoid bottlenecks at the domain servers/gateways) and most importantly deployment problems.

The main factor to deployment problems is the fact that network or hierarchical approaches require major modifications to the networks due to the need of specialized entities for the protocol to work. This issue might not a major hurdle for newly implemented networks; for example, suppose that a new service provider, who currently does not have existing network infrastructure, wants to start to provide LTE services. They cannot avoid at least startup cost to install a new network, thus deploying the latest technology with the best performance is a logical action. Nevertheless, it is a different matter for service providers who already have existing network infrastructure, which is equipped with legacy equipment. To incorporate a new protocol or system that requires specialized equipment, for the new features to be available globally throughout the existing network, major modifications or installation of new devices to the existing network must be done, which consequently cause a very high cost. For most service providers, unless the modification is deemed to be crucial or critical in avoiding a major failure, they would not consider to change the existing networks [7].

As a result, there is a growing interest among researchers in finding an alternative way that can enhance and improve the performance of the network, or the handover process in this case, but without any modification to the network itself. One of the approaches that have gained more attention among researchers is the end-to-end (endpoint centric) handover approach, which does not require modifications to any nodes or equipment in the network. The only modification needed is on the endpoints themselves, which are the mobile device (user's side) and its corresponding hosts (servers or other users). The question is, how this approach could be compared to the network based counterparts? What kind of improvements, what kind of drawbacks or what kind of trade-offs would come with this approach. Thus, in this chapter, the comparison between endpoint centric approach which provides multi-home environment (HIP in particular) and existing network-based approach which only provides single-home environment will be discussed in this chapter.

Additionally, since HIP is chosen as the platform to implement the proposed handover approach, some question on why it should be used instead of existing approaches such as SIGMA and ECHO that use SCTP, will arise. At the end of this chapter, a study by Mugga et al. [79] will be presented to justify the advantages of using HIP.

3.2. Comparison between Multi-Home (HIP) and non-Multi-Home (MIPv6, HMIPv6, HAWAII, CIP and TIMIP)

A simulation study on the effectiveness of HIP, in terms of packet loss rate and handover latency, is done in this work in order to compare the handover performance of HIP with other network-based methods. The simulation topology considered for this study is as shown in figure 3.1.

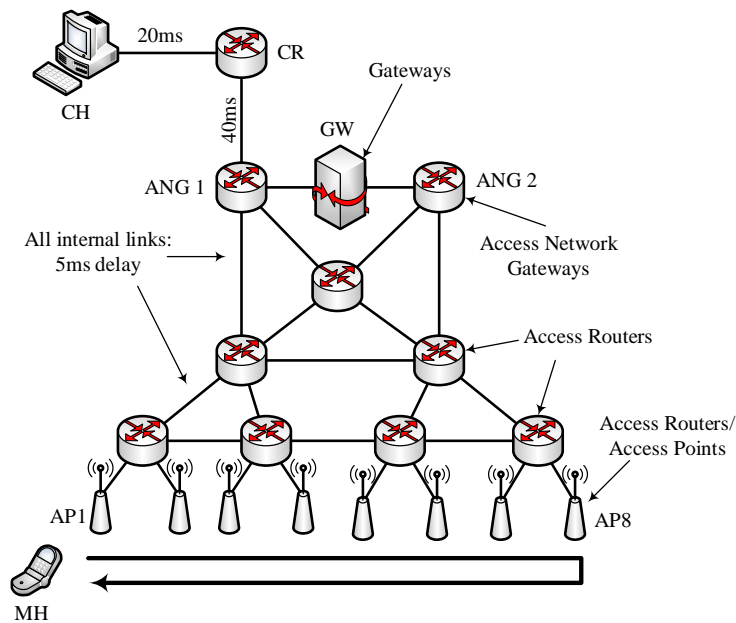


Figure 3.1 Network Topology used for the simulation to evaluate the packet loss rate and handover latency of the considered methods.

The simulation starts by the connection of the MH to AR1. Then the MH moves in a sequential manner, where it will handover from AP1 to AP2, then AP2 to AP3, and so on until it reaches AP8. When it reaches AP8 it will turn back and the sequence of the AP will be reversed until the MH reaches AR1 again. This cycle will repeat according to the simulation time duration selected. The speed of the MH is varied for each simulation run, where the unit of speed used here is in the number of handover per minutes (hpm). It means that, the higher the number of handover per minutes, the higher the MH's velocity. Through mathematical calculation, a speed of 5hpm is equivalent to 10 meter per second (mps). The speed of the MH is configured from 0hpm to 50hpm for each simulation run respectively (0mps to 100mps).

The access points (have the same capabilities of access router; have both layer-2 and layer-3 capabilities) and the access routers are interconnected to the single GW, organized in a hierarchical structure of point-to-point wired links of 10Mbit/s and with sufficient buffer space to force queueing instead of packet dropping. This is done so that any packet loss occurrence is due to the packet loss in the wireless network, not because of buffer overflow in the wired network. The link delay for internal links, which are the connections between ARs, APs and ANGs, is configured to 5ms respectively. The connections outside the MH's network domain have higher delays (e.g. 40ms between ANG1 and CR, 20ms between CR and CH) to simulate inter-domain traffic. As for the wireless part of the network topology, each AP will manage an independent 11Mbit/s IEEE802.11 cell with different channels.

In this simulation, the CH is configured to send UDP packets according to G.711 [82], where 160 bytes packets are sent at a rate of 50 packets per second, to get a sending rate of 64kbps. Only downlink packet flow is considered, since this kind of data flow is easily affected by the handover process. The results of the simulation are discussed in the following sub-chapter.

3.3. Packet loss rate and handover latency comparison results

The packet loss comparison is shown in figure 3.2, where the horizontal axis is the speed of the MH in handover per minutes (hpm) and the vertical axis is the packet drop percentage. The handover latency results are shown in figure 3.3, where horizontal is the same as figure 3.2, in hpm, whilst the vertical axis is the handover latency in seconds. The packet loss considered in this simulation is the percentage of packet loss, which is obtained by equation (3.1). The handover latency is the summation of the durations between the last packet received via the old AP and the first packet of received via the new AP divided by the number of handovers as shown in equation (3.2).

$$PacketLoss(\%) = \frac{\text{Number of packets not received at MH}}{\text{Number of packets sent}} \times 100 \quad (3.1)$$

$$HandoverLatency = \frac{\sum(\text{first packet via new AP} - \text{last packet via old AP})}{\text{number of handovers}} \quad (3.2)$$

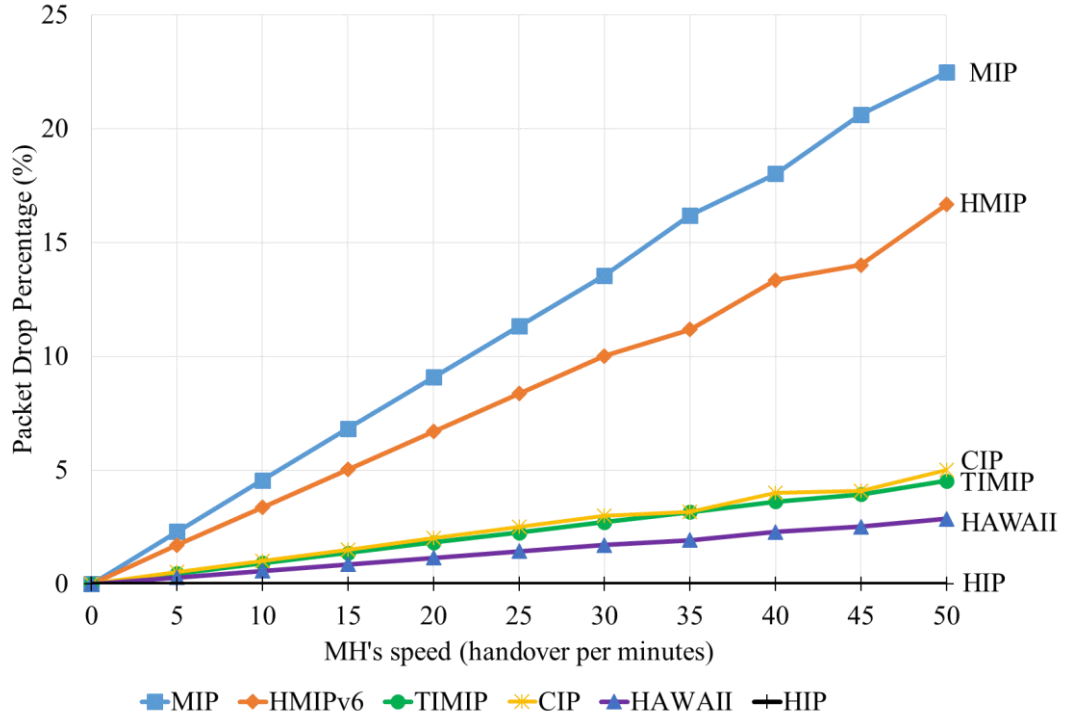


Figure 3.2 Packet drop rate comparison between HIP (multi-home, endpoint centric) and other existing MIP-based protocols (single-homed, network centric).

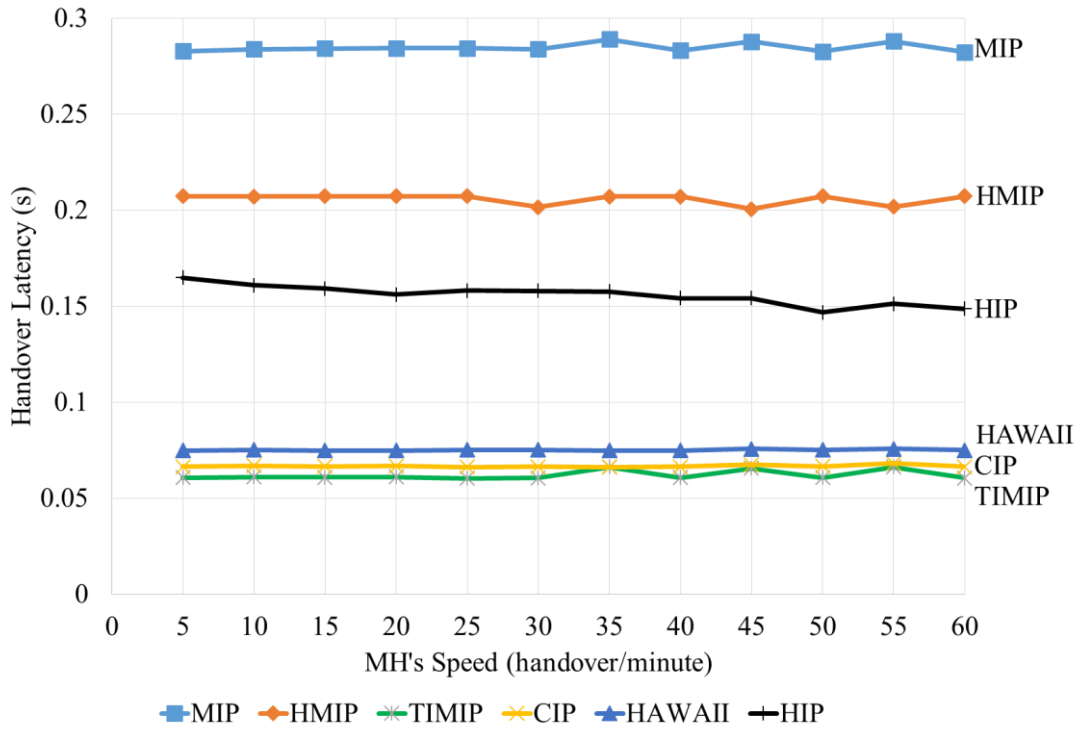


Figure 3.3 Latency comparison between HIP (multi-home, endpoint centric) and other MIP-based protocols (single-home, network centric).

Packet loss comparison: From figure 3.2, it can be seen, obviously, that MIP has the worst result. This is due to the drawbacks that have been discussed in chapter 2, where the auto configuration and binding update poses a very large delay. HMIP follows after MIP, and is followed by TIMIP, CIP and HAWAII. This outcome has been expected from the discussion of the features of these methods in chapter 2. HMIP is better than MIP since the hierarchical architecture incorporated can improve auto configuration and binding updates by implementing a local routing procedure. However the delay is still high since the handover signaling still needs to be communicated to the gateway, which causes a considerably large delay. Then a great jump in the performance can be seen from HMIP to TIMIP, CIP and HAWAII. TIMIP, CIP and HAWAII give nearly the same outcomes, since the implemented approach and architecture are nearly the same. CIP has higher packet loss due to the layer 2 hard handoff during which the data packets may still be sent to the old network, but the MH has already broken off from the old network and is connected to the new network. As for HAWAII and TIMIP, better performance can be seen since the packets are redirected to the new AP during the handover process. Take note that the performance of TIMIP, CIP and HAWAII in terms of packet loss is very good (considering that they are single-home approaches), since they can maintain the packet loss percentage at a very low level. Furthermore, the packet loss rate of these protocols is still within the acceptable range in achieving a high MOS value (refer to table 2.2 and 2.3 in chapter 2). However, the need for major network nodes modification impedes the deployment of such approach in the real world. Finally, from figure 3.2, HIP shows the best results, where it can maintain very low packet loss due to the multi-home environment. This is because, the MH finishes the establishment to the new AP before switching to the new network as the primary communication route. From this observation, it can be clearly seen that the implementation of multi-home environment is advantageous since very low packet loss rate can be maintained even if the MH is moving at high speed.

Handover latency comparison: In figure 3.3, also obviously, MIP has the highest latency, which is due to the binding update process and the auto configuration procedures in the handover process. Next in line is HMIP, also as expected, the delay is due to the binding updates at the gateway for each handover process. Third in line is HIP; due to the end-to-end approach, the handover latency is at least equal to one round trip time between the MH and the CH. The handover latency is dominated by three enhanced hierarchical approach, HAWAII, CIP and TIMIP. HAWAII has the highest latency of the three due to the MH – domain root router route establishment method, not every HAWAII router is aware of the MH's location, thus, some delay might occur during the establishment through the new AP. CIP is better than HAWAII due to the use of layer-2 triggers (RSS) to trigger the handover process, compared to layer-3 trigger used by HAWAII (e.g., router advertisement messages as defined in RFC 1256 [25]). TIMIP is the fastest due to the more structured routing reconfiguration method used during the handover process, where the old route is diverted quickly to the new route at the closest crossover router between the old AP and the new AP. From this study HIP is far better compared to MIP and HMIP, but still has a higher latency compared to HAWAII, CIP and TIMIP. This is the gap between endpoint centric approach and

network based approach. However, the delay for HIP is still within the acceptable range for VoIP (the maximum acceptable latency in VoIP is a round-trip-time of 400ms [90]) depending on the network delay. Furthermore, since in multi-home approach, the MH is connected to both old and new APs during the handover process, the handover latency will only affect the overlapping distance between old and new APs needed for a successful handover process. As long as the overlapping distance is sufficient, multi-home approach can provide a lossless handover experience.

In a nutshell: There is a tradeoff between the packet loss rate and the handover latency. When using an endpoint centric, multi-home environment, the packet loss rate can be minimized compared to network centric, single-home approach. On the contrary, single-home network centric can minimize the handover latency, whereas the endpoint centric approach can only achieve a minimum latency that is equal to a round trip time between the MH and the CH. However, the characteristic of multi-home environment can compensate the handover latency with the fact that the MH is connected to more than one network during the handover process. Moreover, the fact that no network node modification is needed to implement the handover approach is one of the deciding factor in choosing the endpoint centric, multi-home environment as the handover platform.

3.4. Justification on selecting HIP-based approach instead of implementing SCTP-based approach (comparison between endpoint centric multi-home environment)

To justify the reason of choosing HIP instead of SCTP, a study done by Mugga et al. [79] will be discussed in this sub-chapter. In their work they have done a simulation study that compares the rehomining time between HIP and SCTP. Rehomining time is the time taken for each protocol to detect the link failure to the time when data communication is resumed on the alternative path. Take note that in their study, they did not implement any layer-2 triggering (e.g. RSS) information to trigger the rehomining process. The rehomining is triggered purely based on the protocol's capabilities; for HIP the trigger used is the IPv6 neighbor discovery, whilst for SCTP, the rehomining is triggered from the retransmission time out (RTO). The results of their simulation are shown in table 3.1. It can be seen from the results that HIP shows a faster rehomining time compared to SCTP. Thus it is expected that, with the implementation of more effective handover triggering, HIP can achieve faster rehomining time performance.

Table 3.1 Performance comparison between SCTP and HIP.

Protocol	Rehomining Time (s)
SCTP	0.99 ± 0.0100
HIP	0.41 ± 0.0004

3.5. Concluding remarks

From the discussion in this chapter, it can be concluded that HIP is one of the best candidates to be used as the platform for the handover approach implementation. In terms of packet loss, HIP, a candidate for the endpoint centric and multi-home environment, has shown better performance compared to current most enhanced network centric, single-home environment based protocols (HAWAII, CIP, and TIMIP). In terms of handover latency, HIP is far better than MIP and HMIP, but limited by the endpoint centric characteristic; handover control signaling has to be communicated directly between the MH and the CH, restricting the handover latency performance to a least one round trip time. However, the availability of redundant paths for failure resilience, the avoidance of network modification for fast deployment of the handover method, and also the ability to achieve zero packet loss outweighs the discrepancy of the handover latency. Furthermore, the performance on rehomeing time using HIP is better than the one using SCTP, which has been justified with the presentation of an existing study by Mugga et al. [79].

CHAPTER 4

SIGNALING COST COMPARISON BETWEEN ENDPOINT CENTRIC AND NETWORK CENTRIC APPROACHES

In this chapter, the signaling cost of endpoint centric approach, using Host Identity Protocol (HIP) particularly, is compared to network centric approach represented by HMIPv6. The main target of this chapter is to compare the signaling cost of using end point centric approach, HIP in particular, with a network centric approach. HMIPv6 is chosen as the benchmark for network centric approach, because the analytical model is widely available in the literature [45, 47, 91-93] and because of the fact that MHIPv6 is the base model for more enhanced MIP-based hierarchical protocols such as HAWAII, CIP and TIMIP. In terms of signaling, HAWAII, CIP and TIMIP are based on HMIPv6; the only difference and advantage of the enhanced protocols are: HAWAII – this protocol improves the handover signaling through multiple layers of HAWAII routers and the use of domain root router; TIMIP – nearly the same as HAWAII; and CIP also through the use of specialized routers and through the use of layer-2 triggers. If the topology is simplified to just only one level below the domain root router for HAWAII or the domain gateway in CIP, then the signaling cost will probably be similar to HMIPv6. The effectiveness of using endpoint centric approach compared to using network centric approach will be discussed via analytical models of the HMIPv6 and HIP.

4.1. Network structure considered in the signaling cost comparison

In this section, the particulars on the considered network structure, the analytical model used to calculate signaling cost of HIP and HMIPv6, and the comparison results will be presented. The considered network structure considered in for the analytical model is shown in figure 4.1. A two dimensional subnet arrangement where $AR_{1,1}, \dots, AR_{m,n}$ represents the access routers. There is one location manager (LM) (which is the home agent (HA) in the case of HMIPv6) and a number of CHs connected to the Internet. The MHs are connected to one or more CHs, and they roam within the subnets covered by $AR_{1,1}, \dots, AR_{m,n}$. Intermittent data transfer between a pair of MH and CH is assumed to occur caused by communication between MH and CH via applications like

HTTP and VoIP. This active transfer period during the whole MH-CH interactivity is known as a session.

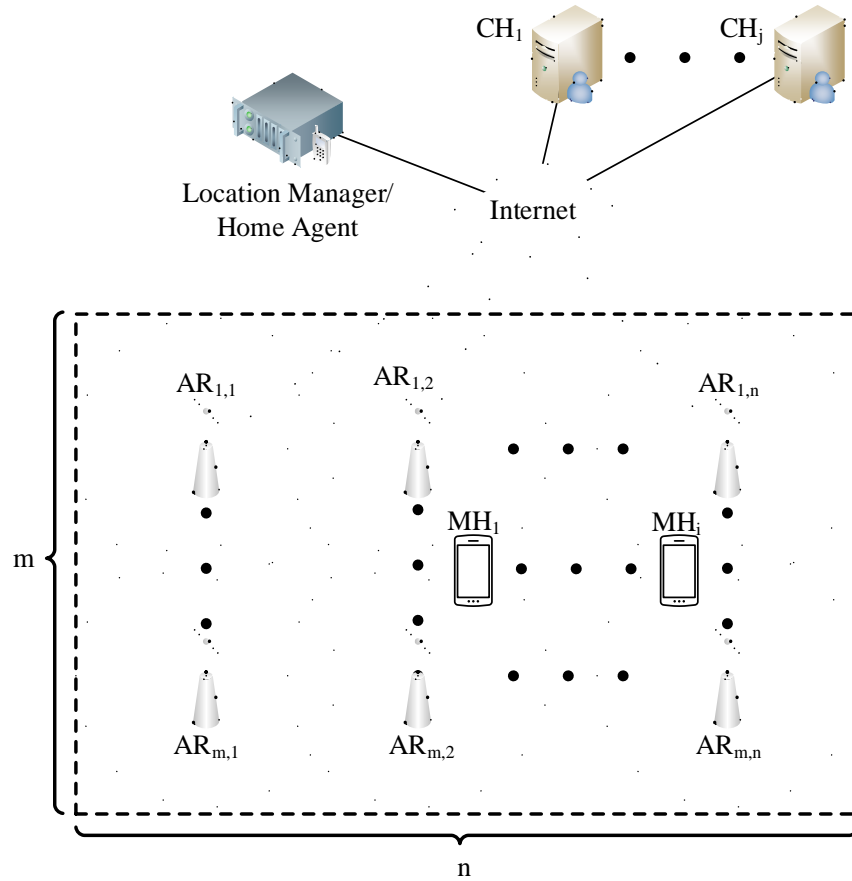


Figure 4.1 Network structure considered for the signal cost analysis. The number of access routers (AR)s, in other words the number of subnets existing in the topology is equal to $m \times n$. A total number of i MH and j CH is considered. The location manager (LM) (for endpoint centric approach) is the same as HA in the case of HMIPv6.

4.2. Notations

The notations used to develop the analytical models of HIP and HMIPv6 are given in this subsection. The notations are divided into three parts; the first part are the notations applicable to both HIP and HMIPv6, the second part are the notations that apply to HIP only, and the third part are the notations for HMIPv6 only. For the sake of consistency, the notations for HMIPv6 modeling are similar to those used in [47].

4.2.1. Notations that apply to both HIP and HMIPv6 signaling cost modeling

N_{mh}	Total number of MHs.
N_{ch}	The average number of CHs communicating with an MH.
T_r	MH's residence time in a subnet. This determines how long an MH stays in one subnet.
S	Number of session during an MH-CH transport layer connection (association) time.
λ_{sa}	Average session arrival rate.
λ_{pa}	Average packet arrival rate.
ϕ	Session-mobility ratio defined as $\lambda_{sa} \times T_r$.

4.2.2. Notations that apply to only HIP signaling cost modeling

l_{ml}	Average distance between MH and location manager in hops.
l_{mc}	Average distance between MH and CH in hops.
LU_{ml}	Transmission cost of one location update from MH to location manager
γ_l	Processing cost at location manager for each location update
v_l	Location database lookup cost per second for each transport layer association at LM
BU_{mc}	Transmission cost of one binding update between MH and CH
Ψ_{LU}^{HIP}	HIP location update cost per second for the whole system, including transmission cost and process cost incurred by location update of all MHs
Ψ_{BU}^{HIP}	HIP binding update cost per second between MHs and CHs for the whole system
Ψ_{PD}^{HIP}	HIP packet delivery cost per second from CNs to MNs for the whole system
Ψ_{Tot}^{HIP}	HIP total signaling cost per second for the whole system including location update cost, binding update cost and packet delivery cost, $\Psi_{Tot}^{HIP} = \Psi_{LU}^{HIP} + \Psi_{BU}^{HIP} + \Psi_{PD}^{HIP}$

4.2.3. Notations that apply to only HMIPv6 signaling cost modeling

l_{mh}	Average distance between MAP and HA in hops
l_{mc}	Average distance between MH and MAP in hops
LU_{mh}	Transmission cost of one location update from MH to HA
LU_{mm}	Transmission cost of one location update from MH to MAP
γ_h, γ_m	Processing cost for each location update at HA and MAP, respectively
v_h, v_m	Processing cost for each data packet at HA and MAP, respectively
C_{mh}	Registration cost of one location update from MH to HA, including transmission cost and processing cost
C_{mm}	Registration cost of one location update from MH to MAP, including transmission cost and processing cost
R	Number of subnets under a MAP
M	Average number of subnet crossings that will cause a HA registration in HMIPv6, i.e. MH moves out of region covered by a MAP
Ψ_{LU}^H	HMIPv6 location update cost per second for the whole system which includes transmission cost and processing cost incurred by location update of all MHs to their HA and/or MAP
Ψ_{PD}^H	HMIPv6 packet delivery cost per second for the whole system from CHs to MHs, including the encapsulation/decapsulation processing cost at mobile agents
Ψ_{Tot}^H	Total HMIPv6 signaling cost per second for the whole system including location update cost, binding update cost and packet delivery cost, $\Psi_{Tot}^H = \Psi_{LU}^H + \Psi_{PD}^H$

4.3. Signaling Cost Analysis of HIP

In this section, the signaling cost of HIP is analyzed. This model is derived from the model developed by [47] which describes the signaling model of SIGMA. Since both SIGMA and HIP implements the same system architecture, where the communication between the MH and CH is done in an end-to-end way, and the location information is maintained by a location manager (which could be a DNS server), very much like the network depicted in figure 4.1. The only difference between HIP and SIGMA is that, SIGMA incorporates SCTP as the entity to enable the multi-home environment. Thus, the model used here is similar to the one developed for SIGMA, with some difference in the binding update cost.

The cost for HIP is divided into three parts, which are the HIP Location Update (LU) cost, the HIP Binding Update (BU) cost and the HIP Packet Delivery (PD) cost. The LU is the cost incurred when the MH sends a location update packet to the LM each time it moves from the old subnet to a new subnet (subnet crossing). The BU is the cost incurred when MH updates its location

with the CH during subnet crossing. Finally the PD is the cost incurred by packets exchange between MH and CH during a session.

4.3.1. HIP Location Update cost

In HIP, when the MH migrates from one wireless network to another (which happens every T_r seconds, which is the dwell time of the MN in one network), The MH will send a location update signal to the location manager, which will cause a transmission cost, LU_{ml} , and a processing cost, γ_l . Since the location update is sent only once, and only to the LM per network migration, the number of CHs that the MHs is currently communicating with will not affect the location update cost. Consequently, the average cost of location update per second for the whole system can be estimated as the multiplication of the total number of MNs with the location update cost for each MN per the MN's residence time in the network, as denoted by equation 4.1:

$$\Psi_{LU}^{HIP} = N_{mh} \frac{LU_{ml} + \gamma_l}{T_r} \quad (4.1)$$

The location update cost is divided into two components, which are the wired part of the network which is denoted as $(l_{ml} - 1)$, which represents the number of wired hops and the wireless part which is denoted by the authors in [47] as the proportionality constant, ρ . This constant is considered because it is generally known that the cost incurred by a wireless hop is greater than that of wired, due to the retransmissions and medium access contentions at the data link layer of the wireless links. When the per-hop location update transmission cost is δ_U , the location update transmission cost, LU_{ml} for a round trip between the location manager and the MH can be obtained with equation 4.2:

$$LU_{ml} = 2(l_{ml} - 1 + \rho)\delta_U \quad (4.2)$$

Therefore the total location update cost is as follows:

$$\Psi_{LU}^{HIP} = N_{mh} \frac{2(l_{ml} - 1 + \rho)\delta_U + \gamma_l}{T_r} \quad (4.3)$$

4.3.2. HIP Binding Update cost

For the binding update costs analysis, the costs at the endpoints (MH and CH) is not considered, because the processing cost of MH and CH does not contribute to the network load. Because this analysis is focusing on the cost impacts to the network performance, by omitting those processing costs, more focus can be concentrated on the impact of the handover protocol on the network performance. The same assumptions have been made by other works [45, 47, 91-93].

The binding update transmission cost, BU_{mc} , is the cost due to MH's binding update at the CH that happens at every subnet crossing. Since HIP is an endpoint centric approach, the binding update has to be sent to each CH currently connected to MH after each handover. Thus the average binding update cost can be estimated as:

$$\Psi_{BU}^{HIP} = N_{mh}N_{ch} \frac{BU_{mc}}{T_r} \quad (4.4)$$

Take note that there is no processing cost due to earlier assumption, only the transmission cost is considered here. When the per-hop binding update transmission cost is δ_B , the binding update transmission cost, BU_{mc} for a binding update for HIP can be obtained by equation 4.4:

$$BU_{mc} = 3(l_{ml} - 1 + \rho)\delta_B \quad (4.5)$$

Equation 4.2 used in this analysis is different from the one done in [47] since the protocol used is different. The work in [47] uses SCTP which implements a two-way handshake for the binding update between MH and CH. Since HIP is used in this analysis, the binding update transmission cost is multiplied by 3 due to the three-way handshake incorporated by HIP (refer to subsections 2.2.3 and 2.2.7 in chapter 2 for more information). Thus the binding update cost per second for the whole system can be obtained by equation 4.6:

$$\Psi_{LU}^{HIP} = N_{mh}N_{ch} \frac{3(l_{ml} - 1 + \rho)\delta_B}{T_r} \quad (4.6)$$

4.3.3. HIP Packet Delivery cost

In HIP a location database lookup is executed each time the MH start an association with the CN. This lookup is done via the LM, or in more detail the Rendezvous Server (RVS) [57]. The HIP Base Exchange (session initiation) using a RVS server is shown in figure 4.2. In the MH will first send an I1 message to the RVS to initiate a session with the CH. The I1 some information: source IP address, destination IP address, Host Identity Tag (HIT) of the initiator (MH in this case) and HIT of the target responder (CH in this case). The RVS will then forwards the I1 to CH with some extra information telling the CH that the I1 packet is from MH and informing the CH about the RVS' Hashed Message Authentication Code (HMAC) (a security feature in HIP). Then, the Base Exchange is completed with R1, I2 and R2 messages. The lookup process inside the RVS server is considered as the location database lookup cost, in this case.

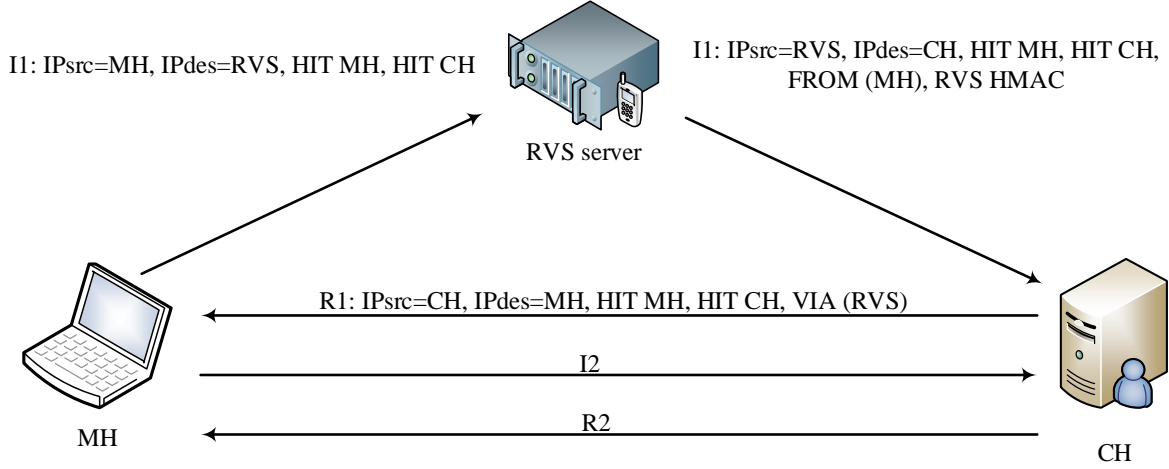


Figure 4.2 HIP Base Exchange using an RVS server.

Furthermore, the same assumptions as in [47] is taken here, where the data packet transmission cost, IP routing table searching cost and bandwidth allocation cost is not considered, since they are not particularly related to mobility protocols. Instead, only the location data base lookup cost at the RVS and tunneling related costs are considered to clearly show the impact of mobility protocol on the overall network load.

Assuming that each session duration time is independent from each other, the association setup event can be estimated to happen every S/λ_{sa} seconds. Further assuming that the database lookup cost has a linear relationship with N_{mh} , and ϕ_l be the per location database lookup cost, whilst ψ be the linear coefficient at LM, then the per-second per-association lookup cost v_l can be calculated as:

$$v_l = \frac{\phi_l \lambda_{sa}}{S} = \frac{\psi N_{mh} \lambda_{sa}}{S} \quad (4.7)$$

As there are no packet encapsulation or decapsulation in HIP, no processing cost is considered in the intermediate routers. Thus, the delivery cost can be expressed by equation 4.8, where the calculation only includes the location database lookup cost.

$$\Psi_{PD}^{HIP} = N_{mh} N_{ch} v_l = N_{mh}^2 N_{ch} \frac{\psi \lambda_{sa}}{S} \quad (4.8)$$

4.3.4. Total Signaling cost of HIP

Summing up the costs discussed previously (location update, binding update and packet delivery costs), the total signaling cost of SIGMA can be defined as:

$$\Psi_{TOT}^{HIP} = \Psi_{LU}^{HIP} + \Psi_{BU}^{HIP} + \Psi_{PD}^{HIP} \quad (4.9)$$

4.4. Signaling Cost Analysis of HMIPv6

In this section, the signaling cost of HMIPv6 will be discussed. The model that is presented in this section was developed by Fu et al. [47], and is included here for ease of discussion. Since the MH does not send a binding update to the CH in HMIPv6, there will be no binding update cost. The total signaling cost will only consider the location update cost and the packet delivery cost.

4.4.1. HMIPv6 location update cost

In HMIPv6, the mobile node registers only with the (Mobility Anchor Point) MAP, as long as it is still in the region covered by the MAP. If it moves out of that region, a registration with the HA will be done. A registration to a MAP will incur a transmission cost to MAP (LU_{mm}) and a processing cost of the location update message at the MAP (γ_m). Therefore, the cost of update to MAP (C_{mm}) becomes:

$$C_{mm} = LU_{mm} + \gamma_m \quad (4.10)$$

where LU_{mm} is defined by equation 4.11.

$$LU_{mm} = 2(l_{mm} - 1 + \rho)\delta_U \quad (4.11)$$

When the MH crosses to another subnet (happens every $M \times T_r$ seconds), that MH will have to register with HA. This process will cause two types of costs which are the transmission cost to HA (LU_{mh}) and the processing cost at the HA (γ_h), and also the processing cost at MAP ($2\gamma_m$, since MAP needs to process both registration request and reply messages). Thus, the cost of update to FA (C_{mh}) becomes:

$$C_{mh} = LU_{mh} + \gamma_h + 2\gamma_m \quad (4.12)$$

where LU_{mh} is defined as follows:

$$LU_{mh} = 2(l_{mm} + l_{mh} - 1 + \rho)\delta_U \quad (4.13)$$

Similar to HIP, the MH also does not send update messages to CN in HMIPv6, hence the cost will only be affected by the number of MN in the system. The average location update cost per second for the whole system can be estimated using equation (4.14). The expectation that the MH will move out of a MAP regional network is denoted by M and is calculated by equation (4.15), where $m \times n$ is the total number of subnets and R is the number of subnets under a MAP.

$$\Psi_{LU}^H = N_{mn} \frac{MC_{mm} + C_{mh}}{MT_r} \quad (4.14)$$

$$M = \frac{mn - 1}{mn - R} \quad (4.15)$$

The total signaling cost for HMIP location update is shown by equation 4.16:

$$\Psi_{LU}^H = N_{mh} \left[\frac{2(l_{mm} - 1 + \rho)\delta_U + \gamma_m}{T_r} + \frac{2(l_{mm} + l_{mh} - 1 + \rho)\delta_U + \gamma_h + 2\gamma_m}{T_r} \times \frac{mn - R}{mn - 1} \right] \quad (4.16)$$

4.4.2. HMIPv6 packet delivery cost

For the HMIPv6 packet delivery cost, the same assumptions as in section 4.3.3 is taken, the only cost considered are the location database lookup cost and tunneling related costs at HA and MAP. For each packet sent from CH to MH, the processing cost incurred are:

Process for lookup at HA: (i) one location database lookup, (ii) one encapsulation at HA.

Process for lookup at MAP: (i) one location database lookup, (ii) one decapsulation, and (iii) one encapsulation at MAP

Let τ be the per encapsulation/ decapsulation cost at HA or MAP, there will be τ for the lookup at HA due to one encapsulation process, whilst for the lookup at MAP, there will be 2τ due to the decapsulation and encapsulation process. Let ψ to be the linear constant for location database lookup as defined in equation (4.7). Thus the processing cost of each data packet at HA and at MAP can be defined by equations (4.17) and (4.18) respectively.

$$v_h = \varphi_h + \tau = (\psi N_{mh} + \tau) \quad (4.17)$$

$$v_m = \varphi_m + 2\tau = \left(\psi \frac{N_{mh}R}{mn} \right) + 2\tau \quad (4.18)$$

Another component needed to calculate the packet deliver cost of HMIPv6 is the packet arrival rate λ_{pa} . The packet arrival rate has to be considered here because unlike HIP – where session arrival rate is sufficient because no tunneling is required for the packet delivery and no per-packet delivery cost component – the tunneling cost for each packet sent from CH to MH has to be considered. Thus, the packet arrival rate can be calculated by equation 4.19. F is the file size being transferred by the session, and PMTU is the path MTU between CH and MH.

$$\lambda_{pa} = \lambda_{sa} \frac{F}{PMTU} \quad (4.19)$$

Finally, the packet delivery cost can be obtained by summing up both HA cost components and MAP cost components as shown by equation 4.20.

$$\Psi_{PD}^H = N_{mh}N_{ch}\lambda_{pa}(v_h + v_m) = N_{mh}N_{ch}\lambda_{pa} \left(\psi N_{mh} \frac{mn + R}{mn} 3\tau \right) \quad (4.20)$$

4.4.3. HMIPv6 total signaling cost

Based on the analysis on the packet delivery cost and the location update cost, the total signaling cost for HMIPv6 is:

$$\Psi_{TOT}^H = \Psi_{LU}^H + \Psi_{PD}^H \quad (4.9)$$

4.5. Results and signaling cost comparison of HIP and HMIPv6

In this section, results that shows the impact of various input parameters on HIP total signaling cost. In all the numerical analysis, the following parameter values obtained from [47] are used:

$\gamma_l = 30$	$\gamma_h = 30$	$\gamma_m = 20$	$\psi = 0.3$	$\rho = 10$
$l_{ml} = 35$	$l_{mc} = 35$	$l_{mh} = 25$	$l_{mm} = 10$	$\lambda_{sa} = 0.01$
$S = 10$	$m = 10$	$n = 8$	$\tau = 0.5$	$R = 10$
$F = 10$ kbytes	$PMTU = 576$ bytes			

γ_l , γ_h and γ_m are the values of lookup process at LM, HA and MAP respectively. The values used here have been calculated by the authors in [47] based on the complexity of the process. Since LM and HA cover larger number of MH entries compared to MAP, their processing cost is logically higher than MAP. The same reasoning is applied for l_{ml} , l_{mc} , l_{mh} , and l_{mm} which corresponds to the distance between MH and LM, distance between MH and CH, distance between MAP and HA, and distance between MH and MAP in hops respectively. Both LM and CH are outside the network topology, and has to go through the core network or the internet to be reached, thus the large value of 35 hops is configured for their distance from MH. As for the distance between MAP and HA is slightly lower, configured at 25, since MAP is located at the edge of the network. As for the distance between MH and MAP, the number of hops is smallest due to the fact that MH is within the region of MAP; and the number adds up, e.g. distance from MH to LM (35) is equals to the sum of the distance between MH-MAP and MAP-LM ($10 + 25 = 35$). Other parameters such as ψ , ρ , F and $PMTU$ are values obtained by the authors in [47] based on user traffic and mobility models in [94, 95].

An example scenario to visualize the network topology is shown in figure 4.3. The different vertical squares with different colors represent network domains which is supported by one MAP each (in the case of HMIPv6). The black arrow shows an example of how an MH crosses subnets (handover). A subnet crossing within the same colored box is the handover within the same domain, whilst crossings between different colored boxes are intra-domain handover.

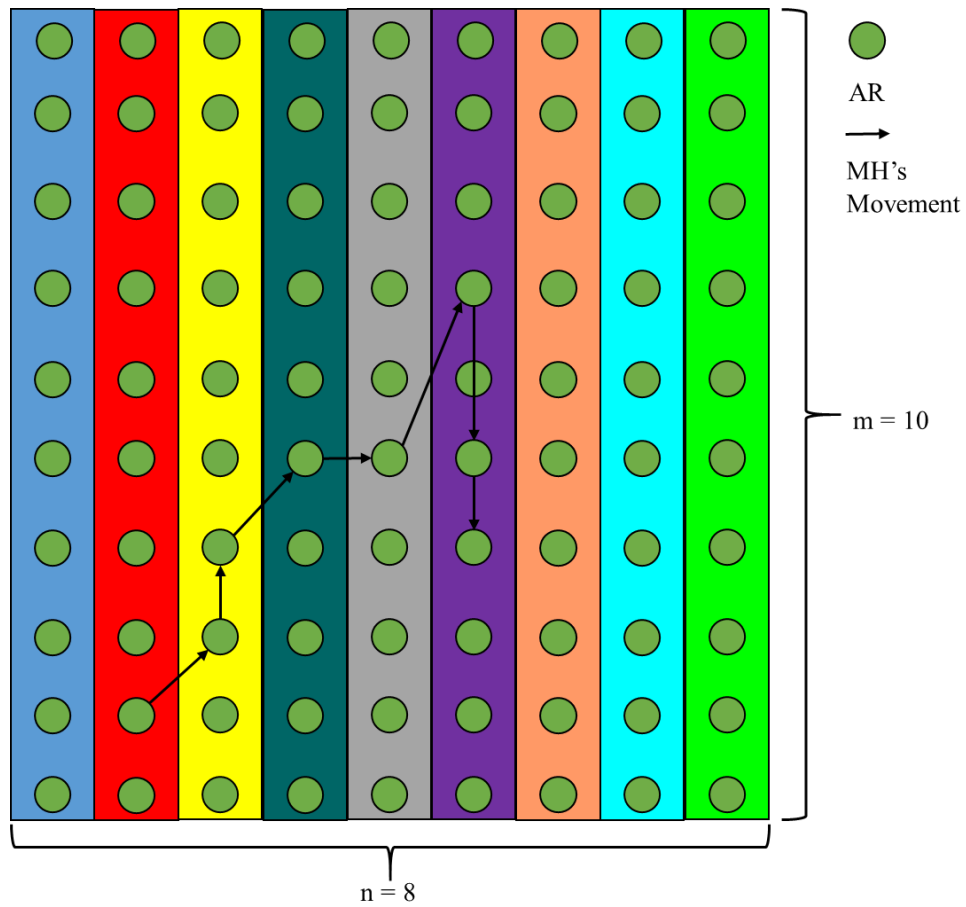


Figure 4.3 An example scenario for better understanding of the network topology and the MH's movement. This figure corresponds to the network shown in figure 4.1.

4.5.1. Impact of number of MHs for different network (subnet) residence times

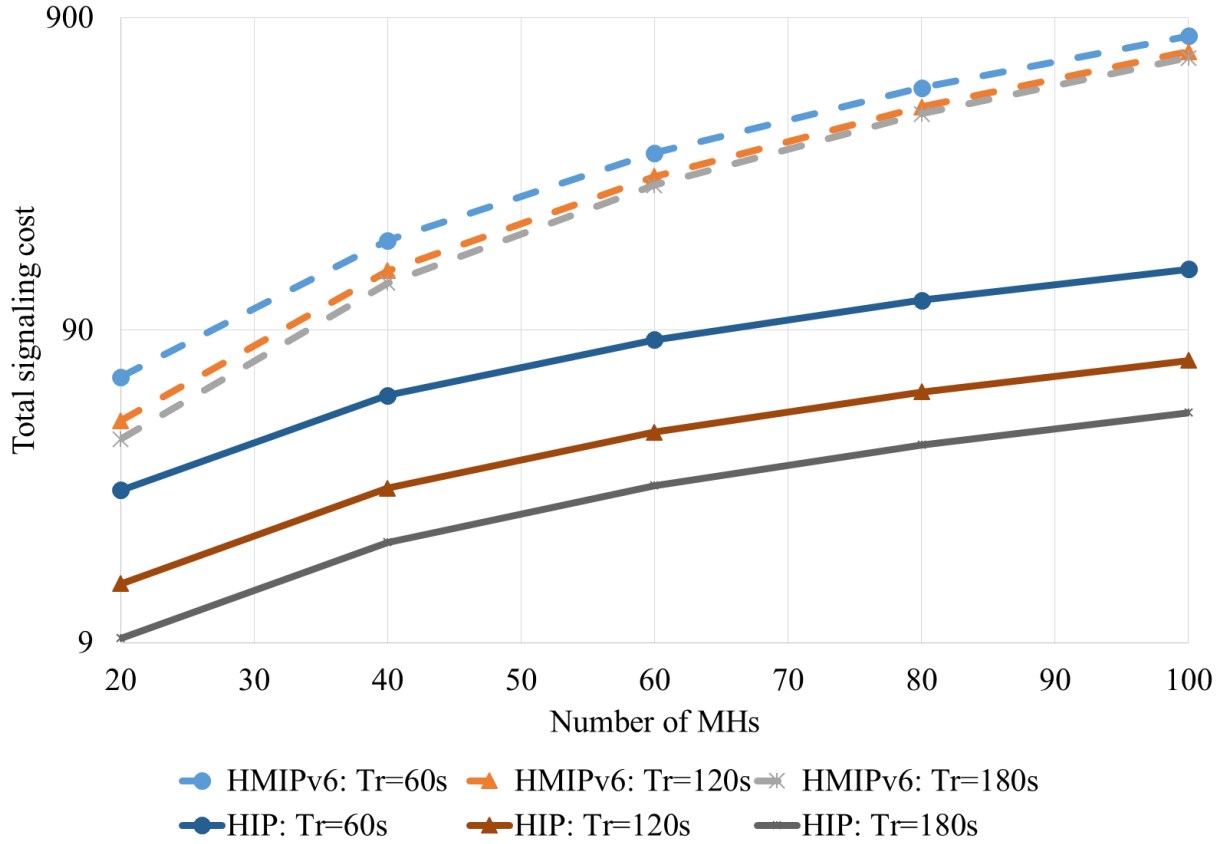


Figure 4.4 Impact of number of mobile hosts (MHs) under different subnet residence times

In this analysis, the number of mobile nodes (N_{mh}) are varied from 20 to 100 nodes and the residence times (T_r) are varied from 60s to 180s, whilst the number of CN (N_{ch}) is configured to 1 and the transmission costs $\delta_U = \delta_B = 0.2$. Figure 4.4 depicts the total signalling cost of HMIPv6 and HIP. From the figure, it can be seen that higher MN speed (smaller residence time, T_r) will incur a higher cost due to the increased number of location and binding updates signals. Furthermore, it can be seen that HIP has lower costs compared to HMIPv6 in this settings; this is mainly due to the small location and binding update transmission costs, δ_U and δ_B respectively. Consequently, the location updates and binding updates becomes low. The high packet delivery cost experienced by HMIPv6 causes the signalling cost of HMIPv6 to become considerably higher than that of HIP.

4.5.2. Impact of average number of communicating CH and location update transmission cost

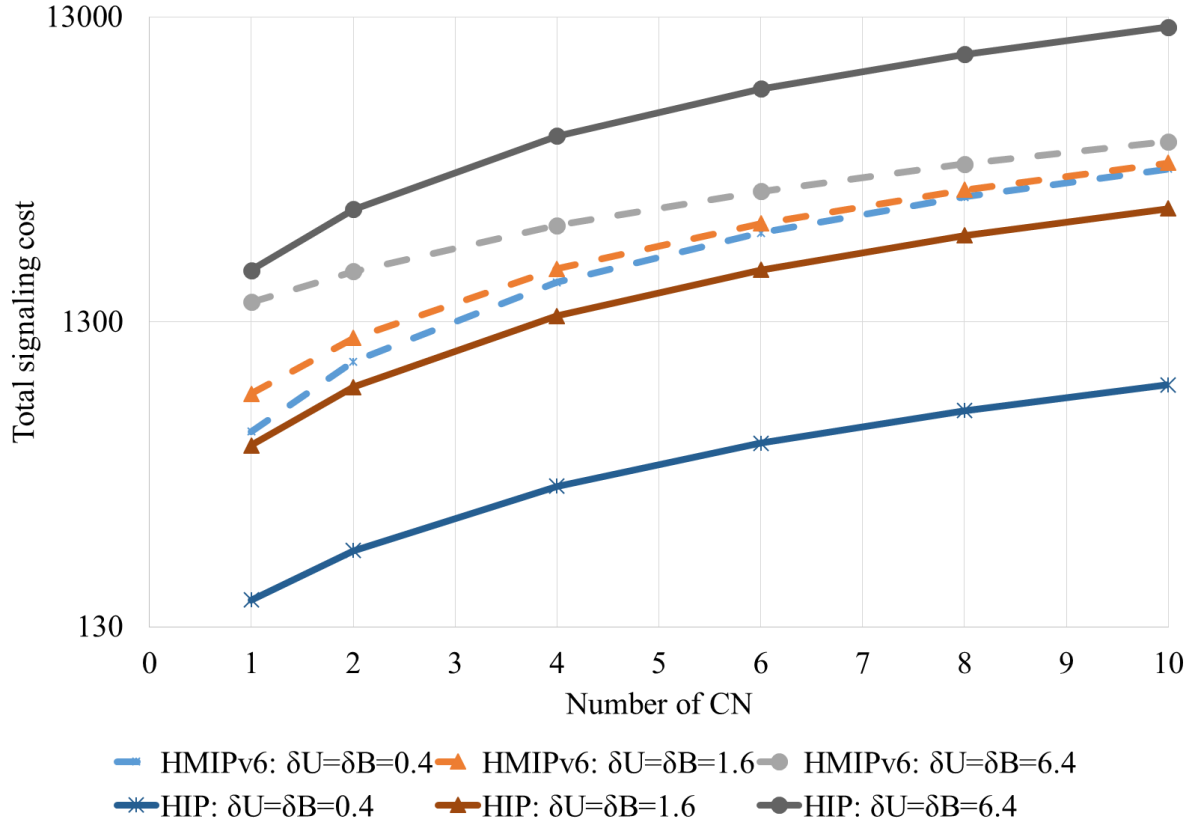


Figure 4.5 Impact of number of CHs and per-hop binding update transmission cost.

For the next analysis, the residence time, T_r , and the number of MHs, N_{mn} , are configured to 60s and 80 respectively. The number of average CHs that an MH communicates with, the per-hop location update cost (δ_U) and binding update (δ_B) are varied, and the impact of doing so is shown in figure 4.5. From this figure, it is clear that an increase in these three values will result in the increase of total signalling cost for both HMIPv6 and HIP. However, it can be seen that the change in δ_U has a smaller impact to HMIPv6, due to its hierarchical structure. Furthermore, HMIPv6 is not affected by δ_B since it does not perform binding updates with the CH. From figure 4.5, it can be seen that the total signalling cost for HIP is well below HMIPv6 for lower δ_U values (0.4, 1.6). However, at higher values of δ_U , it can be seen that the total costs of HIP exceeds that of HMIPv6. This is due to the fact that HIP needs a higher signalling cost due to frequent location and CH binding update for each subnet crossing compared to HMIPv6.

4.5.3. Session to Mobility Ratio

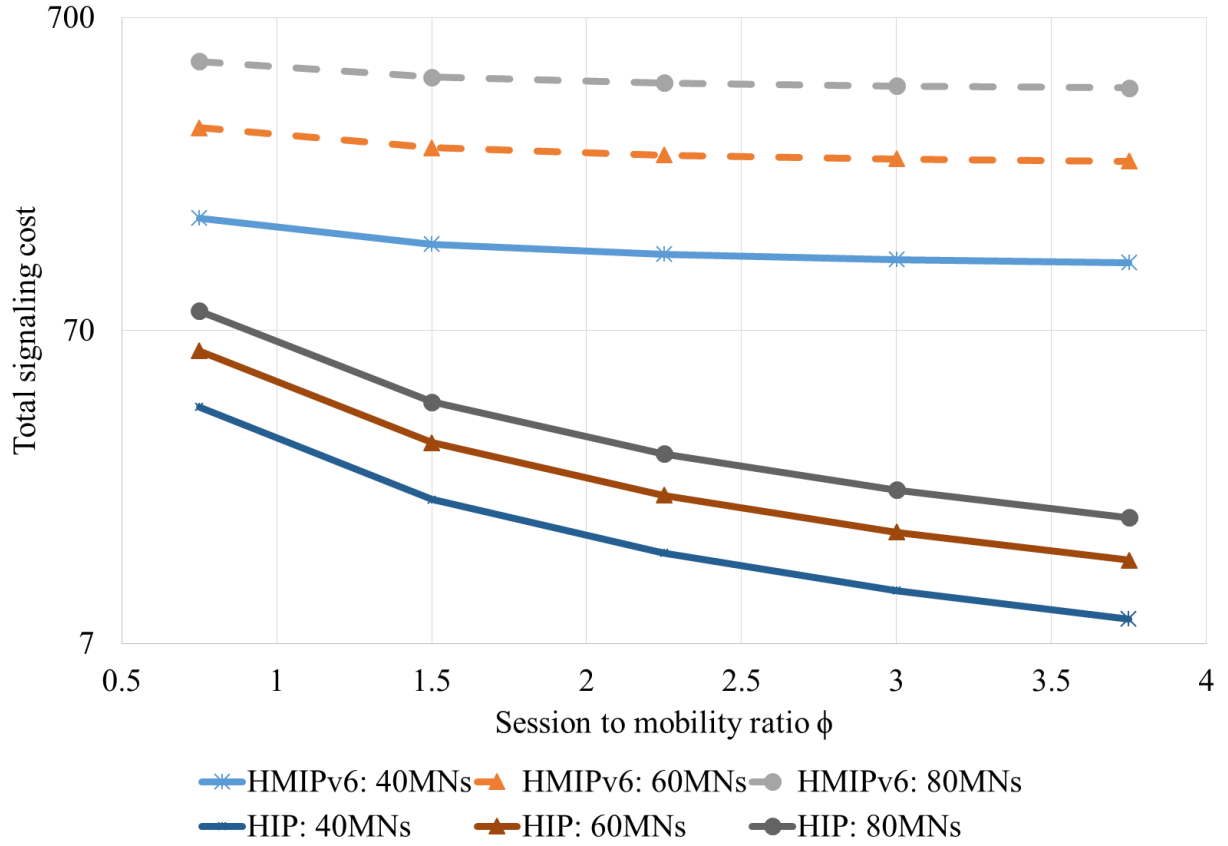


Figure 4.6 Impact of session mobility ratio on total signaling cost for different N_{mh} .

Session mobility ratio is the ratio between the MH's mobility (affected by T_r) compared to the session arrival rate. In this analysis, the residence time is varied from 75s to 375s whilst λ_{sa} is fixed to 0.01, which gives an SMR of 0.75 to 3.75. The graph clearly shows that the total cost will increase when the number of MN is increased. Furthermore, the higher the SMR, the lower the mobility of MN, that is why from the graph, it can be seen that the total signalling cost becomes lower, due to lesser location and binding updates. Additionally, it can be seen that the slope HMIPv6's graph is not as steep as the ones for HIP. This is again due to the hierarchical structure of HMIPv6, which reduces the MN's mobility's influence on the signalling cost.

4.6. Concluding Remarks

To main objective of this chapter is to compare the effectiveness of opting the endpoint centric approach, compared to the network centric approach. The candidate for endpoint centric here is

HIP, whilst the candidate for network centric is HMIPv6. From the analytical analyses done in this chapter, it can be concluded in most cases, endpoint centric approach has lower signaling cost compared to network centric approach. The only time that endpoint centric has worse performance is when the binding update cost coefficient, δ_U is high, which is actually one of the worst case scenarios. However, under normal conditions, endpoint centric approach (HIP) has shown far superior results in terms of signaling cost. Thus, the study in this chapter reaffirm the feasibility of using HIP as the platform to implement the proposed handover approach.

CHAPTER 5

Ant Colony-Based Handover Approach

As discussed in chapter 1, currently there exist many protocols and systems to support for handover and mobility management. The most popular approaches to mobility management are based on MIP. However, these methods face deployment problems due to the need for major modifications on the existing network nodes. Thus, more researchers are becoming interested in finding ways to avoid such problems. One of the methods that have gained a lot of interest is the endpoint centric approach. This approach is very useful in avoiding the deployment issues, since only the endpoints, in this case the users MH and the service or application servers, the CH namely, has to be modified. Thus, the deployment problems can be alleviated.

However, how good is the performance of endpoint centric approaches compared to existing network-based approaches (MIP and its variants)? In chapters 3 and 4 the efficiency and the performance of the endpoint centric approach, HIP in particular, has been clarified. HIP has shown better performance especially in term of packet loss rate and signaling cost. Even though the latency of the endpoint-based approach may be higher than the network-based approach, its multi-home feature can cover this discrepancy (e.g. HIP). Thus, it has also been clarified that endpoint-based handover approach can outperform network-based handover approach.

Therefore, the study in this research has been directed towards endpoint centric approach with protocols and systems that can provide multi-home environment in order to enhance the performance for MH, in terms of the handover performance. Currently, there exist several approaches that have shown good handover performance. Most related to this work are two such methods, which are SIGMA and ECHO. SIGMA utilizes the RSS information to trigger the handover process. When considering real time applications such as VoIP, such kind of trigger cannot ensure the sustainability of the application's quality. Consequently, ECHO was developed in order to alleviate the discrepancy in SIGMA by incorporating the E-Model Mean Opinion Score (MOS). However, the approach in SIGMA and ECHO has some drawbacks, where the handover is triggered based on static threshold. This approach can reduce the effectiveness of the method.

Thus, in this chapter a new handover approach known as the AntNet-Based Handover Algorithm (ANHA) is proposed to alleviate the drawbacks in ECHO. ANHA utilizes the same parameters as ECHO (MOS and RSS). The main difference between ANHA and ECHO is that,

ANHA incorporates an Ant Colony-based algorithm to process the MOS and RSS information and decide the handover trigger effectively. Then the handover is executed using HIP.

5.1. AntNet-based Handover Algorithm (ANHA)

A brief introduction of the proposed method has been disclosed in the chapter 1. For ease of discussion, the summary of ANHA is repeated here and the specific components will be elaborated in the following subsections.

Firstly the architecture of ANHA is, as shown in figure 1.10 (in chapter 1), is divided into three main parts, (i) the information gathering, (ii) the decision-making, and (iii) the handover execution. Basically, the whole handover process is as follows (for ease of discussion the same figures as in chapter 1 are included here):

1. ANHA makes use of two types of ants. The Data Ants (DA) are ants that carry the data (in this case, the VoIP packets), and also convey the information on the time stamp of the ants creation. The other ant, the Request Ant (RA) is the ant that sends the exploration request. The MH carries the probabilistic table with information on the quality of the networks that MH is currently connected (see figure 5.1).
2. When the MH moves into the coverage of two or more APs, the MH establishes connections to the new APs, and sends a Request Ant (RA) towards the CH via the currently used main AP to inform the CH of the existence of alternative paths to MH (see figure 5.2).
3. The ant will pass the notification to the CH and die. When the CH receives the notification, it will start to generate Data Ants (DA) to explore the alternative paths available, and at the same time the DA carries data (see figure 5.3).
4. When the DA arrives at the MH, the ant transfers the timestamp information to the MH and dies (see figure 5.4).
5. The MH will keep track of a path context and pheromone table, and updates the information in the tables with the information from the DA.
6. The amount of deposited “pheromone” depends on the evaluation of the path traversed by the DAs.

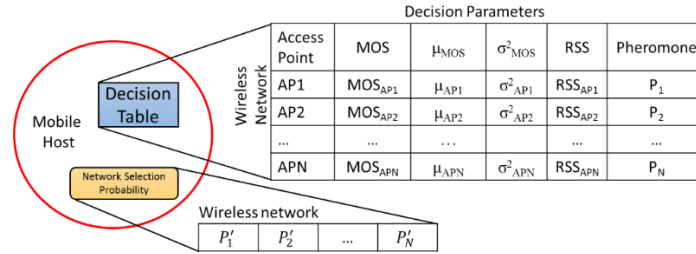


Figure 5.1 Data structures used in MH used to update and decide the handover trigger.

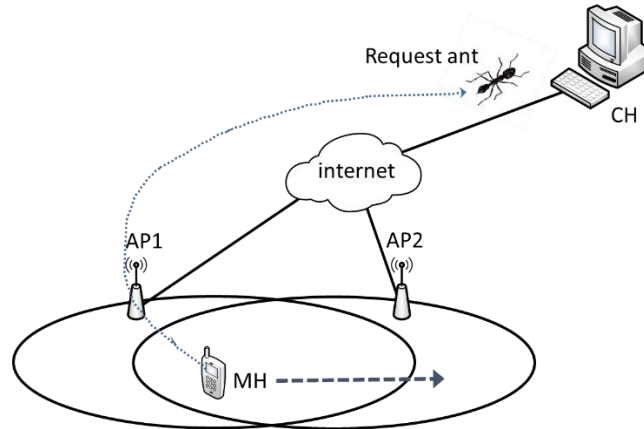


Figure 5.2 Request Ant sent by the MH to inform CH of the new available network and request for duplication packet.

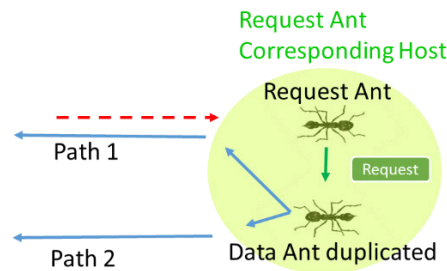


Figure 5.3 The request ant passes the request to the destination node (CH) and dies. The CH starts duplicating packets and sends the data via both paths.

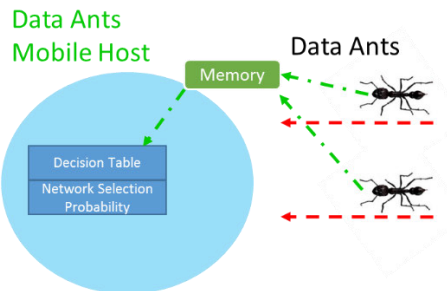


Figure 5.4 The data ants transfer the timestamp information to the MH and die.

5.1.1. Ant deployment and Data collection by Ants

Referring to figure 5.2, at first the MH is only connected to AP1. During this time, MH only receives data carried by the Data Ants (DA). The DA, as its name suggests, carries data (in this case VoIP packets) from the CH to the MH. Additionally, it also carries the time stamp of its creation, which is utilized by the MH to calculate the end-to-end delay, jitter and the packet loss rate of the path that it traverses. Then, when the MH detects a new network, AP2 in this case the MH will send a Request Ant (RA) to the CH to carry the information about the new network (an alternative path between MH and CH). When the CH receives this, it will start the packet duplication process, where it starts to send duplicated packets to the MH. From the DA arriving from these two different paths, MH is able to monitor the path condition based on its MOS value obtained from the packet delay, packet losses and jitter information. Since both paths are receiving the same packet size at the same data rate, the difference in MOS can determine the quality of each path. Then, the path quality information is used in the handover triggering mechanism.

Conceptually, these ants are reminiscent to the ants used in AntNet (refer to chapter 2 for more details). Technically, these ant agents are the control messages in HIP. The DA and RA are created using the modification on the HIP header during packet transmission, where the HIP header is modified to incorporate the timestamp of the packets and the packet duplication option (Packet Dup) as shown in figure 5.5 (in the original HIP header, these two entries does not exists). Using these as the representation of the ants is useful in collecting the information for the handover decision. Furthermore, since the information on the paths is obtained from the same packets as the data, no additional overhead is incurred.

Original HIP Header Configuration

Packet Type	SRC HIT	DST HIT
-------------	---------	---------

Added new header value



Modified HIP Header Configuration

Packet Type	SRC HIT	DST HIT	Time Stamp	Packet Dup
-------------	---------	---------	------------	------------

Figure 5.5 New HIP header values for the time stamp information and the packet duplication start and stop

Table 5.1 Handover Decision Table

Access Point AP_N	MOS_N	μ_N	σ^2_N	RSS_N	Pheromone, P_N
AP_1	MOS_1	μ_1	σ^2_1	RSS_1	P_1
AP_2	MOS_2	μ_2	σ^2_2	RSS_2	P_2
...
AP_N	MOS_N	μ_N	σ^2_N	RSS_N	P_N

5.1.2. Decision Parameter Calculation

As shown in figure 5.1, the MH keeps the parameters collected from the DA within the decision parameter table. Table 5.1 gives a clearer view of the contents of the decision parameter table. The entries of this table are associated with the AP connected to the MH's NIC, and each of these associations will be known as path from here on. MOS_N is the Mean opinion score of a path available to the MH. μ_N is the average value of MOS. σ^2_N is the variance of MOS. RSS_N is the RSS of the AP for the specific path. Pheromone, P_N is the pheromone value of each path.

In ANHA the main parameters used to trigger the handover is the Mean Opinion Score (MOS) [16] and the RSS value. The MOS is calculated using the end-to-end delay, jitter and the packet loss rate of a path. This subsection will discuss on how RSS, path end-to-end delay, path jitter, path packet loss rate are obtained and on how these parameters are processed. After that, the discussion will move on to how these parameters are used to obtain the MOS of the available paths between MH and CH.

Received signal strength (RSS): the received signal strength is collected from the physical layer of the MH. This information is obtained every time the MH receives any packets from the access point (AP).

Path End-to-end delay calculation: Path end-to-end delay (ETED) is calculated at the HIP layer based on the packets brought in by the Data Ants. Each time a DA is transmitted, the local system time stamp is added to the HIP header (via the modification discussed earlier). At the receiver, these timestamps can be used to estimate the ETED. The ETED is calculated by equation (5.1):

$$ETED = CurrentTime - TimeDACreated \quad (5.1)$$

Then the value of the ETED is smoothed using the round trip time formula from the Jacobson's algorithm as shown by equation (5.2). This algorithm essentially creates a kind of moving average effect giving some immunity to the variability of the individual ETED measurement. This is the same model used to estimate the retransmission timeouts in the TCP protocol [116].

The smoothed end-to-end delay value of each NIC (paths) will be used for the MOS calculation. Alpha is a constant to control the moving average window size that determines the number of accumulated instantaneous ETED that will affect the average ETED value. In this study, 0.125 is used.

$$ETED_{accumulated} = ETED_{accumulated} + \alpha(ETED_{accumulated} - ETED_{current}) \quad (5.2)$$

Path Jitter calculation: Packet jitter is also calculated at the HIP layer using the ETED information. Jitter can be defined as the difference between the relative transmit time between two packets; relative transmit time here refers to the ETED. The inter-arrival jitter is a smoothed value of the jitter values. In this study, the inter-arrival jitter is calculated using the E-Model recommended Real-Time Transport Protocol (RTP) jitter algorithm from RFC1889, as shown by equation 5.3 [96].

$$jitter_{inter-arrival} = jitter_{inter-arrival} + \alpha(jitter_{inter-arrival} - jitter_{current}) \quad (5.3)$$

Packet loss rate calculation: Packet loss is detected by checking the number of missed transmission sequence number (TSN), and the percentage of packet loss is calculated by dividing the accumulated lost packet per the total packet received as shown in equation (5.4).

$$Packet\ loss(\%) = \frac{number\ of\ packet\ lost}{total\ packet\ received} \times 100 \quad (5.4)$$

MOS calculation: The e-model has been explained in detail at the end of chapter 2. Before MOS is obtained, the R-factor of a path has to be calculated based on the R-factor equation as shown by equation (5.5). Using the standard values from [16], equation (5.5) can be simplified into equation (5.6), thus reducing the variables to Id and Ie.

$$R = Ro - Is - Id - Ie + A \quad (5.5)$$

$$R = 93.34 - Id - Ie \quad (5.6)$$

The Id component is calculated based on the delay and jitter information. The jitter information obtained from equation (5.3) is filtered first by equation 5.7 [87]. Then the values of ETED and Dj are used to calculate the Id component of equation 5.6. The full explanation and equations to calculate Id is included in [87].

$$Dj = \min(codec_frame_size + (0.9 * RTP_jitter), 300) \quad (5.7)$$

Meanwhile, for the I_e component: the packet loss rate information is mapped to table 2.3 in chapter 2, in order to get the values of I_e based on the packet loss percentage of a path. Then finally the value of R can be calculated from the values of I_d and I_e using equation (5.6). After obtaining the value of R -factor, MOS can be obtained by equation 5.8.

$$MOS = \begin{cases} 1 & R < 0 \\ 1 + 0.035R + R(R - 60)(100 - R)(7 \times 10^{-6}) & 0 < R < 100 \\ 4.5 & R > 100 \end{cases} \quad (5.8)$$

These values are important for the decision process, which will be discussed in the next subsection. All the information calculated here is kept inside the decision table (see table 5.1 for a clearer view).

Up till this point the, information collection phase as in figure 1.10 in chapter 1 has been thoroughly discussed. In the next section ANHA will be discussed in detail.

5.1.3. The proposed ANHA

ANHA is developed based on one of the most successful ACO algorithms, which is called AntNet [75]. AntNet was developed for route optimization in communication networks. The features of AntNet have been discussed in chapter 2. ANHA is an adaptation of AntNet into the handover scenario. In AntNet, the source node and the destination nodes are usually static nodes; however, in this study one of the nodes, which is the MH, is moving and the number of path between the source node (CH) and the destination node (MH) will change as the MH moves from one network to another. Hence, in adapting the algorithm of AntNet in this work some modifications are needed.

In AntNet ant agents are deployed periodically and these ants search for the path from the node sending the ant to some specific destination. However, in this case, the available path is limited to the number of NIC available on the MH and also the number of network currently available to the MH. Thus in the proposed approach on implementing ACO concept, the ants are modified to be directed ants, where in the previous section, the RA are used to request for duplication. Then, the DA will affect the change of pheromone in each path available to the MH. The level of pheromone will then determine the path selected by the MH.

There are several crucial AntNet components that are adapted to ANHA. These components are:

1. The Local Traffic Model
2. The Path Goodness Measurement
3. The Handover Decision table updating rule
4. The Local Decision Policy
5. The evaluation of the DA's path

The Local traffic Model: this component is the local estimate of the network status from the MH's point of view. The information processed here plays an important role in the evaluation of the ant traversed paths and in the statistical filtering of the ant collected information. This component implements a moving exponential estimate of mean and dispersion (basically the Jacobson's Algorithm [116]) for the observed ant trip time. The ant trip time concept from AntNet here is adapted to the MOS information collected from the DA. In ANHA, this component actually estimates the local MOS estimate from the MH's point of view. This component is calculated using equations (5.9) and (5.10). μ_{MOS} and σ_{MOS}^2 are the mean and the variance of MOS respectively. η is the sampling rate used to calculate the mean and variance of MOS. For example, if $\eta = 0.1$, means that only the last 50 samples of MOS will affect the value of the mean and variance of MOS. These equations are adapted from the AntNet local traffic model as shown in table 5.2. In the original AntNet model the round trip time (rtt) is used to measure the traffic condition of the path between the source node and the end node. In this research, rtt alone is considered insufficient since it does not portray the QoE of the measured path. Thus, MOS is opted to replace the rtt.

$$\mu_N \leftarrow \mu_N + \eta(MOS_{current} - \mu_N) \quad (5.9)$$

$$\sigma_N^2 \leftarrow \sigma_N^2 + \eta((MOS_{current} - \mu_N)^2 - \sigma_N^2) \quad (5.10)$$

Table 5.2 ANHA adaption from AntNet local traffic model.

AntNet local traffic model: $\mu_N \leftarrow \mu_N + \eta(rtt_{current} - \mu_N)$ $\sigma_N^2 \leftarrow \sigma_N^2 + \eta((rtt_{current} - \mu_N)^2 - \sigma_N^2)$	ANHA adapted local traffic model: $\mu_N \leftarrow \mu_N + \eta(MOS_{current} - \mu_N)$ $\sigma_N^2 \leftarrow \sigma_N^2 + \eta((MOS_{current} - \mu_N)^2 - \sigma_N^2)$
---	---

The Path Goodness measurement: The goodness of the path between MH and CH cannot be obtained directly, but should be dependent on the status and the changes in each path. In ANHA, the value of MOS is associated to a goodness measure of r where the value of r is between 0 and 1. r can be calculated using equation 5.11.

$$r = c_1 \left(\frac{MOS_{current}}{MOS_{best}} \right) + c_2 \left(\frac{MOS_{best} - I_{sup}}{(MOS_{best} - I_{sup}) + (MOS_{best} - MOS_{current})} \right) \quad (5.11)$$

I_{sup} is calculated using equation 5.12:

$$I_{sup} = \mu_{MOS} + z \left(\frac{\sigma_{MOS}}{\sqrt{W}} \right) \quad (5.12)$$

Where z determines the confidence interval of the I_{sup} , while W is the sliding window size, meaning how many samples of MOS measurement taken for the calculation. A sigmoid filter processes the value of r , so that the values at the highs and lows of r are retained. The sigmoid equation used here is shown by equation 5.13.

$$r \leftarrow \frac{1}{1 + e^{-16 \times r + 8}} \quad (5.13)$$

The value of z depends on the z -table as shown in table 5.3. For example, when considering 95% confidence interval, the value of z will be 1.65.

Table 5.3 The z -table

The z -table										
z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.9973	0.9974

ANHA Path Goodness Measurement equations are adapted from the AntNet model as shown in table 5.4. Notice that the r equation in AntNet, since lower rtt is better while higher MOS is better, the equation for the rtt is reversed, compared to MOS in the adapted equation. Meanwhile, the calculation for I_{sup} and the sigmoid function are the same.

Table 5.4 ANHA adaptation from AntNet Path Goodness Measurement Model.

<p>AntNet Path Goodness Measurement:</p> $r = c_1 \left(\frac{rtt_{current}}{rtt_{best}} \right) + c_2 \left(\frac{I_{sup} - rtt_{best}}{(I_{sup} - rtt_{best}) + (rtt_{current} - rtt_{best})} \right)$ $I_{sup} = \mu_{rtt} + z \left(\frac{\sigma_{rtt}}{\sqrt{W}} \right)$ $r \leftarrow \frac{1}{1 + e^{-16 \times r + 8}}$
<p>ANHA Path Goodness Measurement:</p> $r = c_1 \left(\frac{MOS_{current}}{MOS_{best}} \right) + c_2 \left(\frac{MOS_{best} - I_{sup}}{(MOS_{best} - I_{sup}) + (MOS_{best} - MOS_{current})} \right)$ $I_{sup} = \mu_{MOS} + z \left(\frac{\sigma_{MOS}}{\sqrt{W}} \right)$ $r \leftarrow \frac{1}{1 + e^{-16 \times r + 8}}$

The Handover Decision Table updating rule: The values of pheromone, P_N (refer table 5.1) of each path determines the attractiveness of a path. In ANHA, the pheromone of each path between MH and CH is given a fair amount of pheromone. For example, if there are two paths, each path is given an initial value of $\frac{1}{2}$. Then the level of pheromone is updated via the reinforcement process. The positive reinforcement is shown in equation 5.14, whilst the negative reinforcement (evaporation) is shown by equation 5.15.

$$P_N \leftarrow P_N + r(1 - P_N) \quad (5.14)$$

$$P_{N'} \leftarrow P_{N'} - rP_{N'} \quad (5.15)$$

Let a MH have two NICs, NIC1 and NIC2. When a DA arrive at NIC1, the pheromone at NIC1 will receive a positive reinforcement, whilst at the same time the pheromone at NIC2 will receive a negative reinforcement. Thus, as time goes by, and DA keeps arriving at both NICs, the NIC with better quality will eventually become bigger rapidly compared to NIC with low quality. This is the ant stigmergy characteristic that has been incorporated in ANHA. This updating rule is directly adapted from AntNet without any changes.

The local decision policy: Then the best path is chosen using the local ant decision policy where the probability of choosing a path is calculated using equation 5.16. This information is kept in the network selection array in figure 5.1.

$$P'_N = \frac{P_N + \alpha l_n}{1 + \alpha(|N_k| - 1)} \quad (5.16)$$

P'_N is the probability of choosing a path. P_N is the pheromone level of each path. N_k is the number of available paths. α is the weight for the heuristic component l_n , where l_n is the normalized value of RSS of each AP connected to MH as shown by equation 5.17. The path (AP) with the highest P'_N is selected as the best path, and in the handover process, this path will be chosen. The calculation for P'_N is directly taken from AntNet, whereas the equation for l_n is adapted by changing the $qLength_N$, which represents the queue length of each path in AntNet to RSS, since the parameter the target of proposing ANHA is to integrate both parameters, which are the RSS and MOS in the handover decision and triggering process. Notice that α is the coefficient that controls how much influence RSS has on the handover decision. The value of α can determine whether the MH triggers the handover process earlier or later; for example, when α is bigger, RSS gives more influence and the handover will be triggered earlier while when the α is smaller, RSS have less influence causing the MH to trigger the handover later.

$$l_N = 1 - \frac{RSS_N}{\sum_{n'=1}^{|N_k|} RSS_{n'}} \quad (5.17)$$

Table 5.5 ANHA adaptation from AntNet local decision policy.

<p>AntNet local decision policy:</p> $P'_N = \frac{P_N + \alpha l_n}{1 + \alpha(N_k - 1)}$ $l_N = 1 - \frac{qLength_N}{\sum_{n'=1}^{ N_k } qLength_{n'}}$	<p>ANHA local decision policy:</p> $P'_N = \frac{P_N + \alpha l_n}{1 + \alpha(N_k - 1)}$ $l_N = 1 - \frac{RSS_N}{\sum_{n'=1}^{ N_k } RSS_{n'}}$
---	---

The evaluation of the DA's path: the path available to MH is evaluated in two kinds of way:

- Implicit way: through the decision table update component.
- Explicit way: through the use of RSS value as the heuristic component.

Up to this point, the features of AntNet adapted to ANHA have been discussed. These five components is the core component in implementing a bio-inspired ACO based handover decision algorithm. The algorithm not only take into account the current state of the network (RSS) it also

considers the long-term memory, P_N (the pheromone that are updated based on the MOS level of a path according to the Handover Decision Table update rule), in deciding the best path.

5.2. How ANHA operates

In this section, the operation of ANHA is discussed. The operation starts with the initialization, then followed by the handover decision process and the triggering strategy. Additionally, the method on how the decision metrics are obtained will also be discussed.

5.2.1. ANHA initialization

The following three steps can express the initialization of ANHA:

- Establish VoIP call over UDP and ANHA-equipped HIP daemon.
- Begin monitoring the downlink VoIP stream.
- Begin monitoring the physical layer parameters (RSS) of each NIC on the MH

The ANHA-equipped HIP daemon is created once the MH establishes a connection to the network and a VoIP call is created between the MH and the CH. Once the daemon is created ANHA starts to monitor the downlink stream VoIP packets via the information brought by DA. According to [14] in 802.11 the bottleneck usually occurs on the downlink side, thus this is one of the reasons for monitoring the downlink VoIP stream. This could also be extended to other types of radio access. Additionally, it is worth noting that both MH and CH are equipped with ANHA, thus both sides are monitoring their own downlink, thus accounting for the full duplex call.

The ETED, jitter and packet loss rate, as discussed previously, are mapped to a MOS score in real time, enabling the MH the capability to monitor the quality of the ongoing call. At the same time, the RSS level is also monitored. Since it is assumed that the MH is equipped with multiple NICs, and is capable of multi-homing, it can scan other available access networks. When it detects another network, it can start associating with the new network.

5.2.2. Handover decision using ANHA

When MH moves into the overlapping area between two networks, and has established a connection to the new network, ANHA will start to monitor the MOS level of each path. As soon as the connection to the new network is stable, ANHA starts to work by sending the RA to the CH. It is at this point that ANHA differs from other mobility mechanisms such as ECHO and SIGMA. SIGMA performs a handover immediately after detecting that the RSS level of the new network is higher than the current network without considering the Quality of Experience (QoE) that will be obtained after handover, whilst ECHO will only start the decision making after the

RSS of the new network becomes higher than the current network. ANHA on the other hand, starts the decision process earlier by starting to monitor the QoE (MOS in this case) from each access network, as well as monitoring the RSS of the available networks in deciding the handover triggering time.

5.2.3. Packet Duplication process

Similar to the approach in ECHO, in order for ANHA to calculate the downlink QoE for each path, the MH must be able to receive data from all available paths between MH and CH. To achieve this, the MH deploys the Request Ant (RA) to the CH to request CH to begin duplicating the Data Ants (DA) and start sending via all the available paths. For example, the MH is currently connected to two wireless networks (conceptually, MH has two paths towards CH), the RA will notify the CH to start deploying DA to both available paths.

The duplication request is implemented technically by modifying the HIP header by introducing a Boolean bit (as shown in figure 5.5). To start the duplication, the Packet Dup bit should be set to 1 and to stop it should be set to 0. As only header information is required, each RA contains no data and thus imposing very little overhead. On the reception of the start duplication request, the CH will start to duplicate the DAs and send the same contents to each path towards the MH. Even though the duplication might affect the traffic load of CH, since both paths will endure the same effect, thus it is still possible to compare the quality of the two paths at the MH. When the MH finishes assessing the paths, it transmits an RA again, but this time with the Packet Dup set to 0. When the CH receives this RA, it will stop the duplication process and continues to send data to the path chosen by the MH that is included in the RA.

This process of transmitting duplicated data may seem to have no benefit, but some benefit can be obtained under certain circumstances. During the duplication process, multiple copies of the data packets are sent to the MH via different paths. The MH then has an option to choose which packets that arrives earliest. This provides enhanced performance even before the handover is executed. This is especially apparent when the primary path used by the MH is suffering from poor network performance maybe due to being at the cell edge or due to congestions in the network.

5.2.4. Overview of handover decision

The flowchart of the handover process done by ANHA is shown in figure 5.5 (b). Included in the figure is the flowchart of ECHO, shown by figure 5.5 (a), for easy comparison between the handover decision approach taken by ANHA and ECHO. As can be seen, in ECHO, packet duplication request is sent after the RSS of the target network is higher than the current network. This approach may reduce the overall QoE received by the users. For example, the current network has a very low MOS (e.g. 3.2) and the target network has a very good quality (e.g. 4.2). In ECHO, even though the MH is already connected to the new network, the better MOS level is

not utilized until the RSS of the current network falls lower than the RSS of the target network. This becomes more apparent when the MH is connected to a network with a large coverage area (e.g. LTE). ANHA on the other hand starts the duplication earlier, alleviating the issue faced in ECHO. The packet duplication lets the MH to measure the QoE of the available paths by converting the network parameter (ETED, jitter and packet loss) into MOS using the E-Model R-Factor mapping. The MOS information and the RSS information collected are then processed by ANHA in finding the best path for the MH.

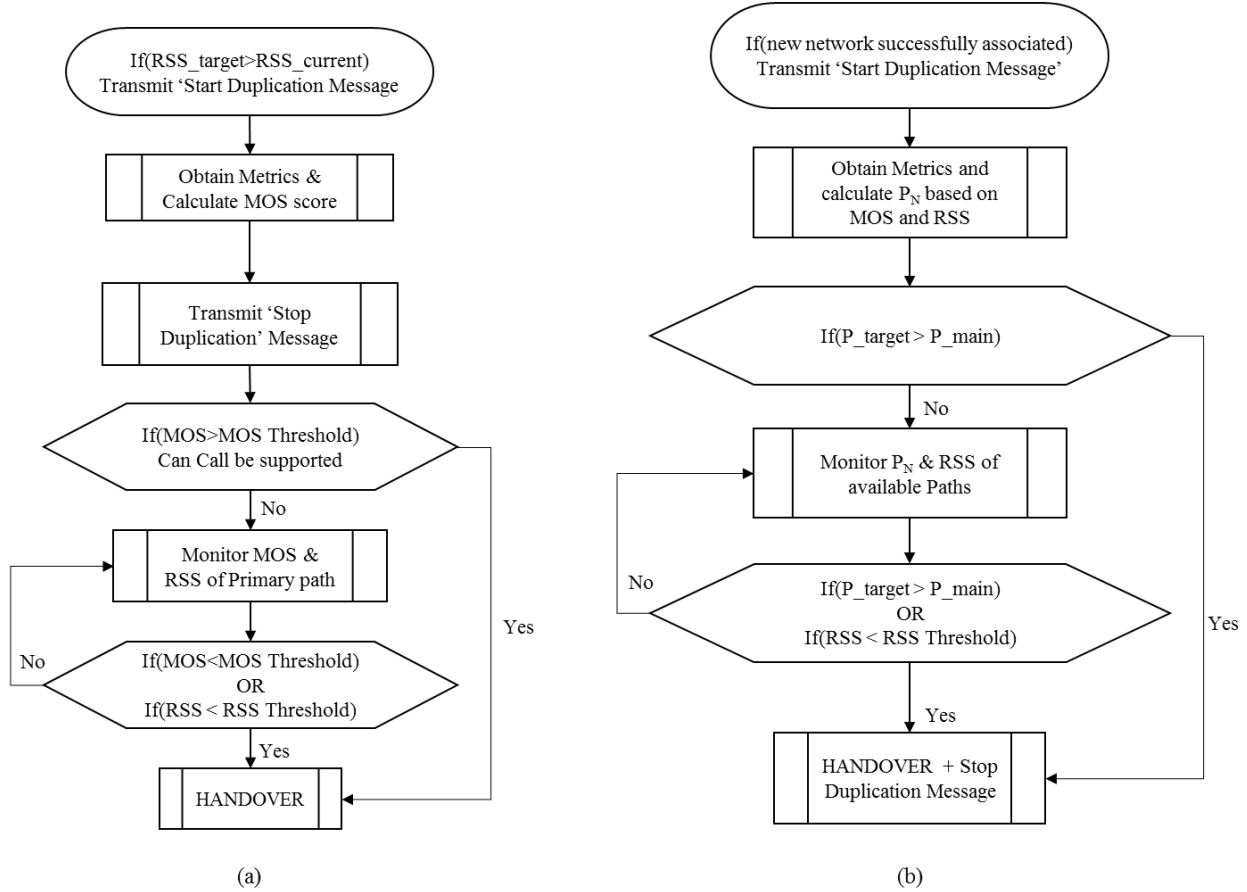


Figure 5.5 (a) The ECHO flow chart, (b) The ANHA flow chart.

Another drawback faced by ECHO is the use of static threshold for MOS and RSS to determine the handover trigger. The use of static MOS threshold might degrade the total average MOS experienced by the users. For example, the current network has a high MOS value (e.g. 4.2) and the target network has much lower MOS value (e.g. 3.2). Let the MOS threshold be 3.1; even though the MH is currently on a network with higher MOS quality, it will automatically handover to the network with lower MOS due to the static threshold settings. Thus in ANHA, the comparison approach is implemented to alleviate this issue. In ANHA's case, the network with the highest value of P'_N (the Probability of choosing a path) will be selected. It means that if the target network has lower MOS than the current network, the handover is not triggered. RSS is

incorporated in the calculation of P'_N , in order to give more control on the handover triggering by adjusting the value of α . Furthermore, a failsafe RSS threshold is also implemented as a safety measure; this topic will be discussed in chapter 6.

5.3. Simulation

The proposed method and ECHO are simulated using OMNeT++ version 4.1 with the HIPSIm++ module [76]. Two network topologies are considered in this research. The first topology is the sequential topology, as shown by figure 5.6. The second topology is the overlapped topology, which is shown by figure 5.7. Since VoIP is considered in this study, the network traffic used in the simulation is in accordance with the G.711 codec [82], where 160 bytes packets are sent every 0.02 second which amounts to 64kbps. The discussion is divided into three main parts, which are:

- **Sequential Topology:** for this simulation, the simulation topology shown by figure 5.6 is considered. The simulation on this topology is divided into two simulation groups.
 - In the first group, the link delay between the Gateway (GW) and AP2 is set to a static value of 10ms and the link delay of GW-AP1 and GW-AP3 is statically set from 10ms to 500ms for each simulation run. The performance of ANHA with different weight settings of α will be discussed for this configuration.
 - In the second group, similar settings are used as in the first group, the only difference is the link delay of GW-AP2 is configured as 450ms.
- **Overlapped Topology:** for this simulation, the simulation topology shown by figure 5.7 is considered. The simulation on this topology is also divided into two simulation groups.
 - In the first group, the link delay of GW-AP1 is statically set to 10 ms, whilst the link delay of GW-AP2 is statically set from 10ms to 500ms for each simulation run. The performance of ANHA with different weight settings of α is discussed for this configuration.
 - In the second group, similar settings are used as in the first group, the only difference is the link delay of GW-AP1 is configured as 450ms

The delay configuration here is to cover the four kinds of scenarios: (i) handover from network with HIGH MOS to HIGH MOS, (ii) handover from network with HIGH MOS to LOW MOS, (iii) handover from LOW MOS to HIGH MOS, and finally (iv) LOW MOS to LOW MOS. In this simulation the main concern is the enhancement of the average MOS experienced by the users throughout the whole simulation run. Thus the comparison of the averaged MOS obtainable by ANHA and ECHO is compared.

- The comparison of the handover behavior of ANHA and ECHO: in this discussion, the graph of instantaneous MOS obtained by the user when using ANHA and ECHO is compared. Take note of the difference between average MOS and instantaneous MOS. Average MOS is as stated above, whilst instantaneous MOS is the MOS level obtained

by the MH at one instance in the simulation. For example, the MOS level of the AP1 obtained by MH at simulation time $t=50s$ is 4.2.

Other important information on the simulation settings is included in tables 5.6 and 5.7 and 5.8. An implementation of ECHO is also included in the simulation as a reference. However, ECHO in this simulation is executed using HIP, instead of SCTP.

Table 5.6 The configuration of the wireless link for the access points in the Sequential Topology.

Sequential Topology	AP1	AP2	AP3
Frequency Band	2.4GHz	2.4GHz	2.4GHz
Link rates	11Mbps	2.6Mbps	11Mbps
Coverage	170 m	170 m	170 m
Receiver Sensitivity	-82dBm	-82dBm	-82dBm
Configurable link delay	10ms – 500ms	10ms or 450ms	10ms – 500ms

Table 5.7 The wireless link configuration for the access points in Overlapped Topology.

Sequential Topology	AP1	AP2
Frequency Band	2.4GHz	2.4GHz
Link rates	2.6Mbps	11Mbps
Coverage	1000m	170 m
Receiver Sensitivity	-82dBm	-82dBm
Configurable link delay	10ms – 500ms	10ms or 450ms

Table 5.8 The variables used for the formulas in ANHA.

$\eta = 0.1$	$W = 50$	$z = 1.7$	α is varied from 0.1 to 1.0, in 0.1 increments
--------------	----------	-----------	---

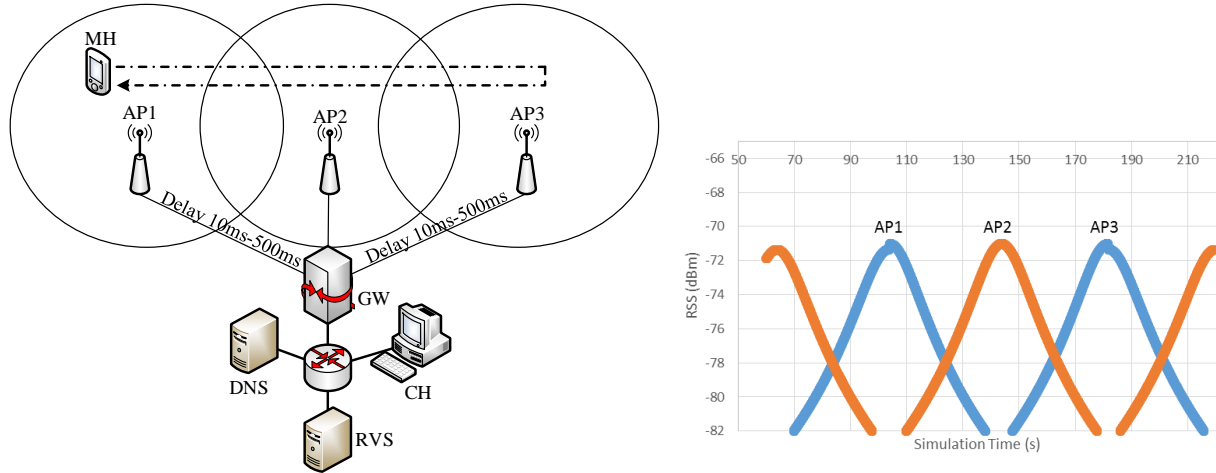


Figure 5.6 The simulation topology considered for the sequential network. The link delay between the Gateway (GW) and are varied from 10ms to 500ms (statically set for each simulation run). Meanwhile the delay between the GW and AP2 is configured as 10ms or 450ms to create two sets of different simulation. On the right is the RSS profile of the topology as seen by the MH.

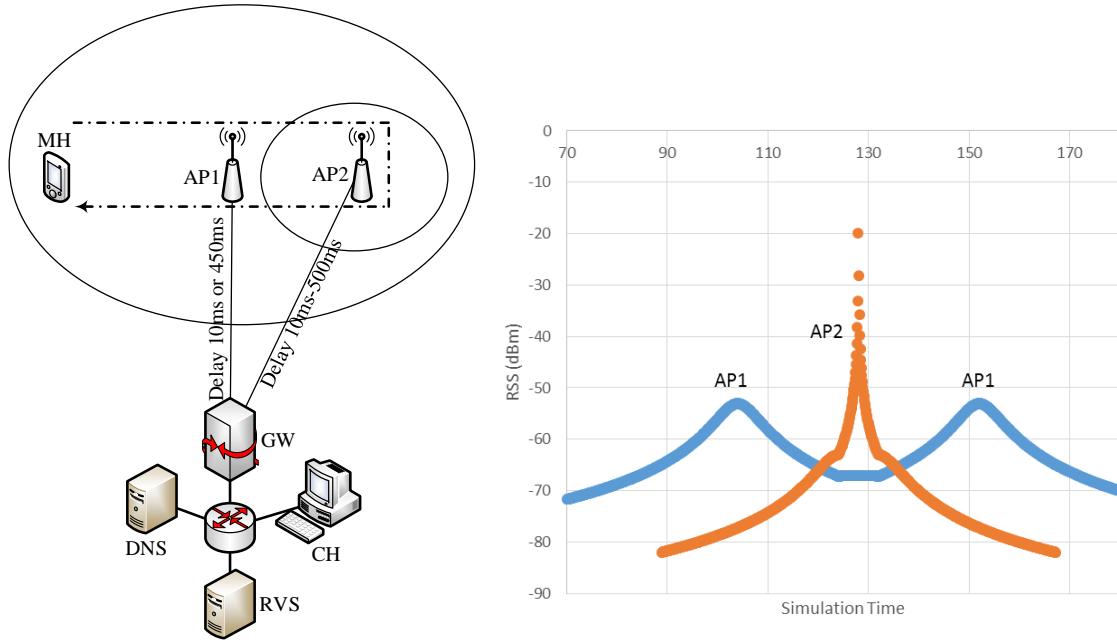


Figure 5.7 The simulation topology considered for overlapped topology. The link delay between the GW and AP2 is varied from 10ms to 500ms (statically set for each simulation). The link delay between GW and AP1 is configured as 10ms or 450ms also to create two different sets of simulation. On the right is the RSS profile of the topology as seen by the MH.

5.3.1. Sequential Topology, AP2 delay = 10ms, AP1 AP3 delay: 10ms-500ms

Figures 5.8 to 5.17 show the average MOS obtained by ECHO and also the average MOS obtained by ANHA with different α configurations. The vertical axis is the average MOS for the whole simulation run. Take note that in this subsection until subsection 5.3.4, the vertical axis of

the graphs is the average MOS for the whole simulation run and not the instantaneous MOS used for the handover triggering. The average MOS is the overall QoE achieved by the users for one whole VoIP session. The horizontal axis the delay settings on AP1 and AP3, which is a static value from 10ms to 500ms, configured correspondingly for each simulation run, whilst the delay for AP2 is set at a static value of 10ms for all simulation runs. Browsing through the graphs in 5.8 until 5.17, it can be seen that generally, when the delay of AP1 and AP3 increases, the average MOS for each configuration drops. ANHA has nearly the same performance as ECHO when the link delay of AP1 and AP3 is not very high (within the range of 10ms to 100ms). Starting for delay of 150 and above, it can be seen that ANHA starts to outperform ECHO, and as the delay increases, the difference becomes clearer. The delay of a network directly affects the MOS level of a network. Thus when the link delay of AP1 and AP3 increases, the MOS level of AP1 and AP3 decreases, making the disparity in the instantaneous MOS level between AP2 and AP1/AP3. In all the figures, it can be concluded that when the disparity of instantaneous MOS level of the networks is small, there are not much difference in the performance of MOS. However, when the disparity becomes larger, it can be clearly seen that ANHA outperforms ECHO. More interestingly, sometimes, ANHA manage to maintain the average MOS at one category higher compared to ECHO, when referring to the MOS classification table in chapter 2 (table 2.4). Thus, it can be concluded that ANHA can maintain higher overall average MOS even when handing over to a network with low MOS levels.

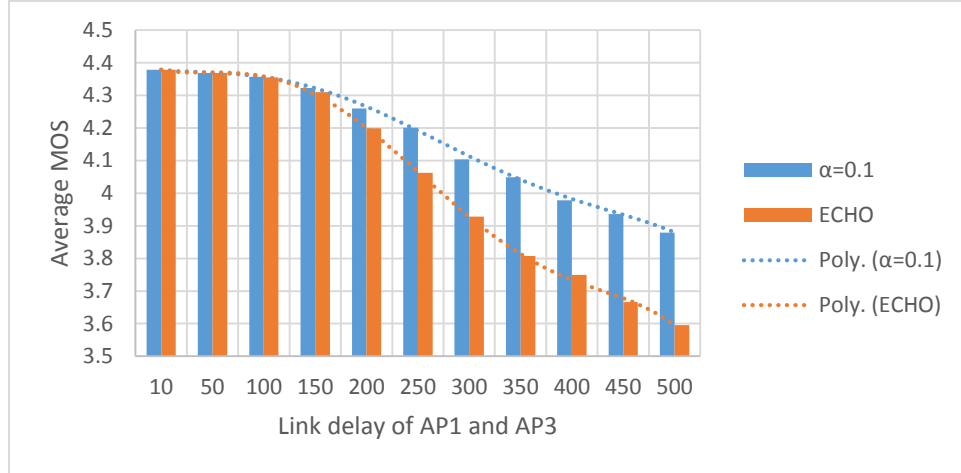


Figure 5.8 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.1$, AP2 = 10ms.

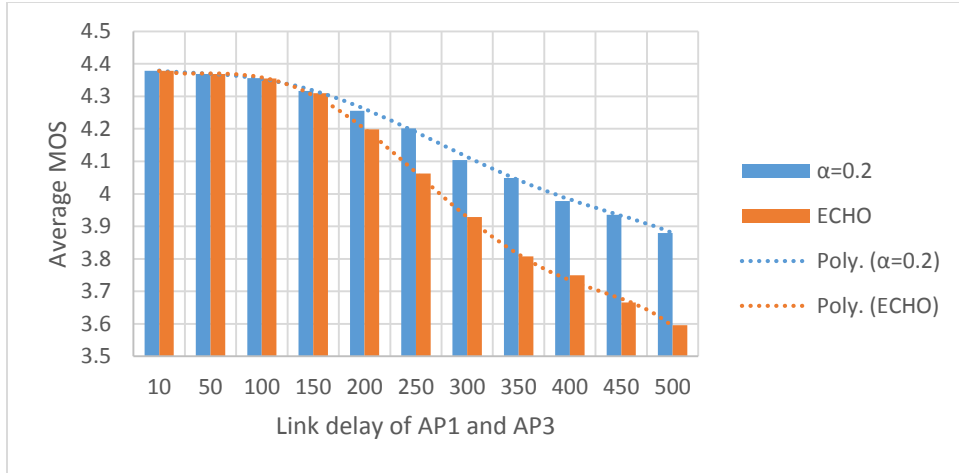


Figure 5.9 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.2$, $AP2 = 10\text{ms}$.

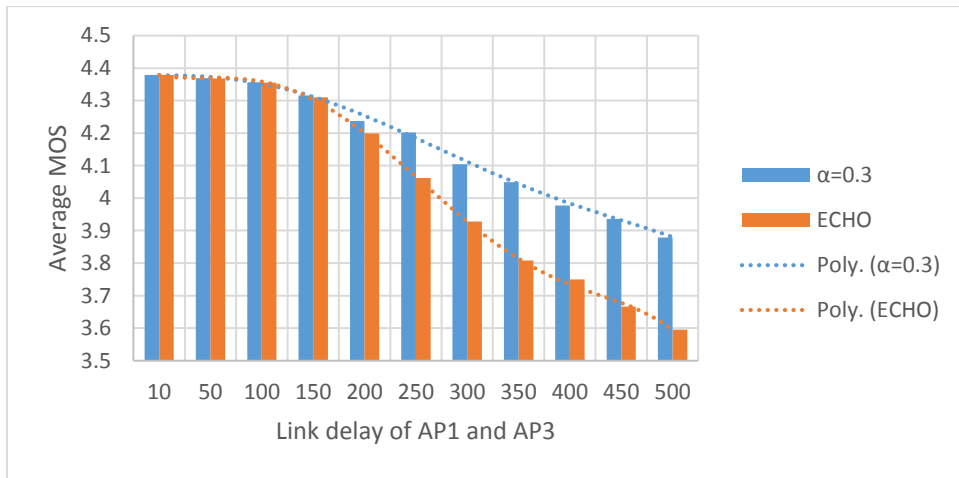


Figure 5.10 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.3$, $AP2 = 10\text{ms}$.

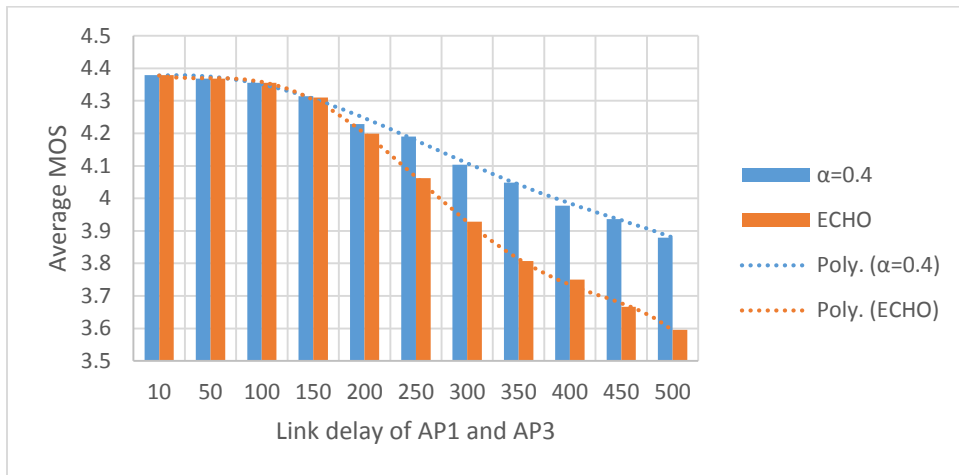


Figure 5.11 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.4$, $AP2 = 10\text{ms}$.

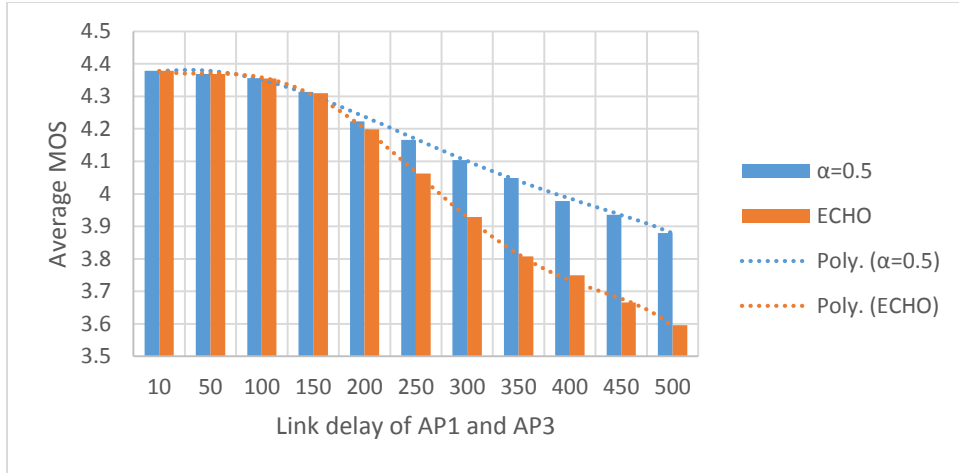


Figure 5.12 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.5$, $AP2 = 10\text{ms}$.

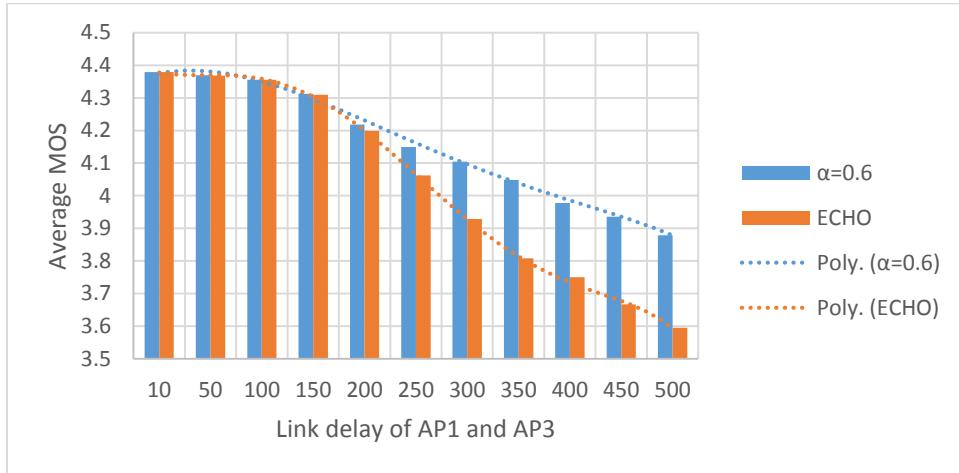


Figure 5.13 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.6$, $AP2 = 10\text{ms}$.

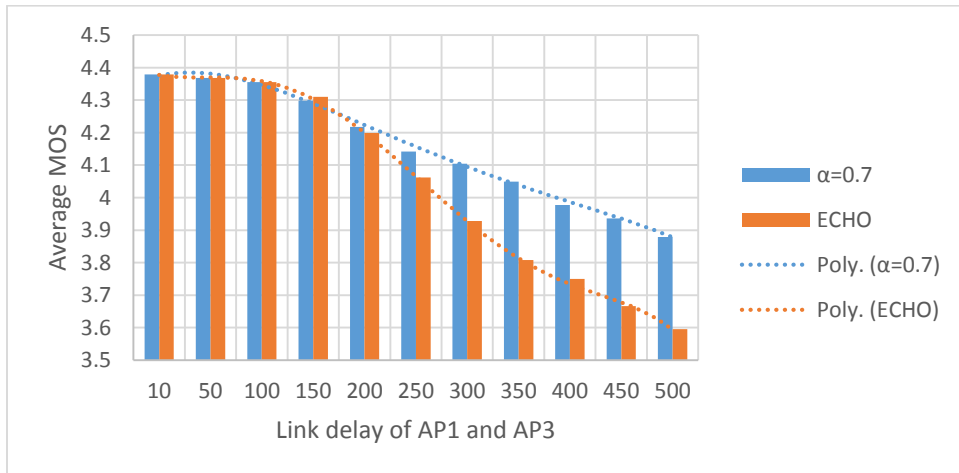


Figure 5.14 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.7$, $AP2 = 10\text{ms}$.

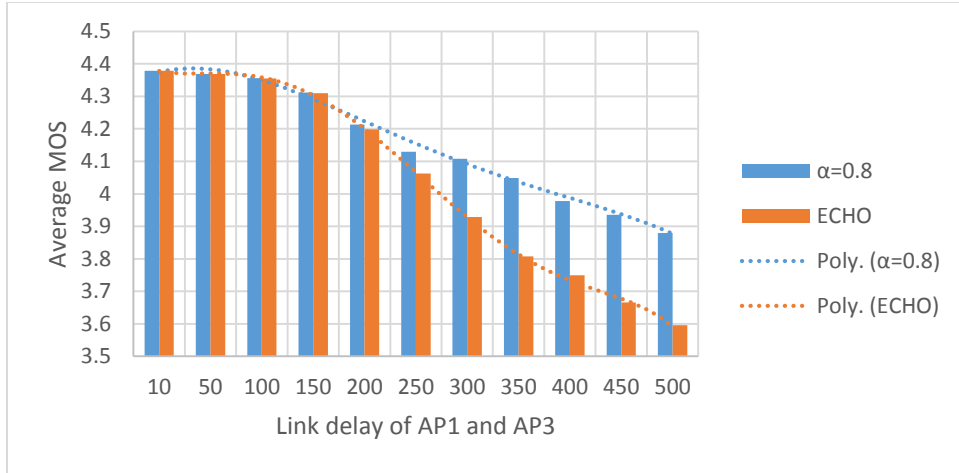


Figure 5.15 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.8$, $AP2 = 10\text{ms}$.

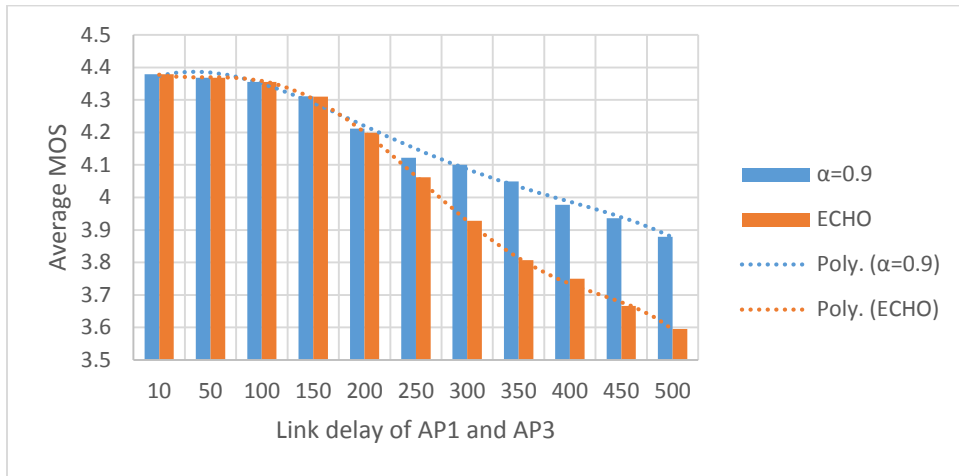


Figure 5.16 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.9$, $AP2 = 10\text{ms}$.

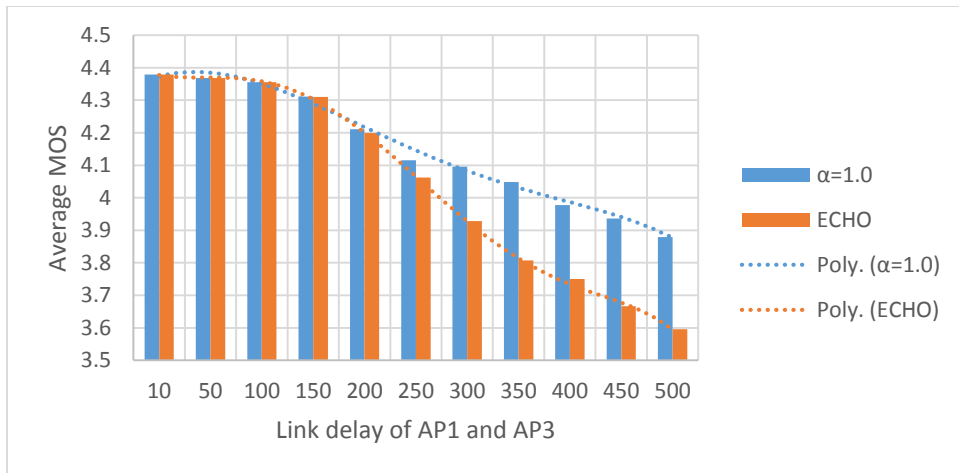


Figure 5.17 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=1.0$, $AP2 = 10\text{ms}$.

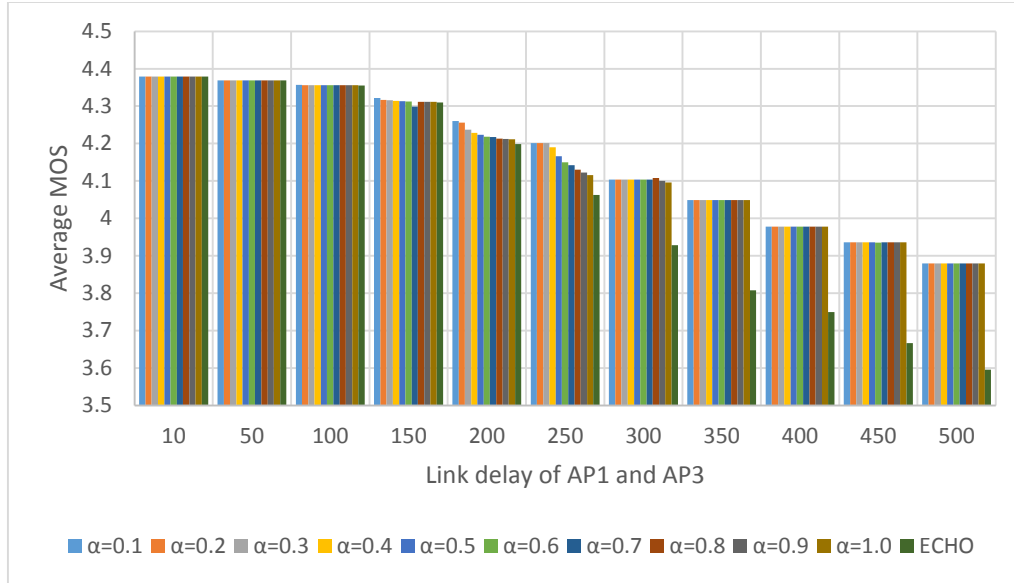


Figure 5.18 Summary of the performance of ANHA with different weights compared to ECHO

Figure 5.18 summarizes all the results in this section. It can be seen that between the delay of 150ms and 300ms, some difference can be seen clearly between the performances of ANHA with different weights. This is mainly due to the emphasis put on the value of RSS. When more weight is put on RSS, ANHA becomes more sensitive to RSS changes, thus changing the handover triggering time. In a nutshell, in this scenario, ANHA outperforms ECHO.

5.3.2. Sequential Topology, AP2 delay = 450ms, AP1 AP3 delay: 10ms-500ms

In this simulation, the link delay between AP2 and the GW is set to 450ms. In this second simulation it can be seen that ANHA shows better performance than ECHO when α is below 0.6. The graphs in figures 5.19 to 5.29 have similar vertical and horizontal axis as in the previous subsection. Similar outcome as the previous subsection could be observed, when the disparity of MOS level between ANHA has shown better performance compared to ECHO. However, starting from figure 5.23 to 5.29, the overall performance of ANHA drops to a similar level with ECHO for most of the results. This may also be due to the weight that makes ANHA more sensitive to RSS changes. For the weight configuration of 0.1 to 0.5, ANHA shows better performance compared to ECHO. In this simulation, the ANHA tries to connect to the network with higher MOS at a longer time, and shorten the time connected to the network with lower MOS.

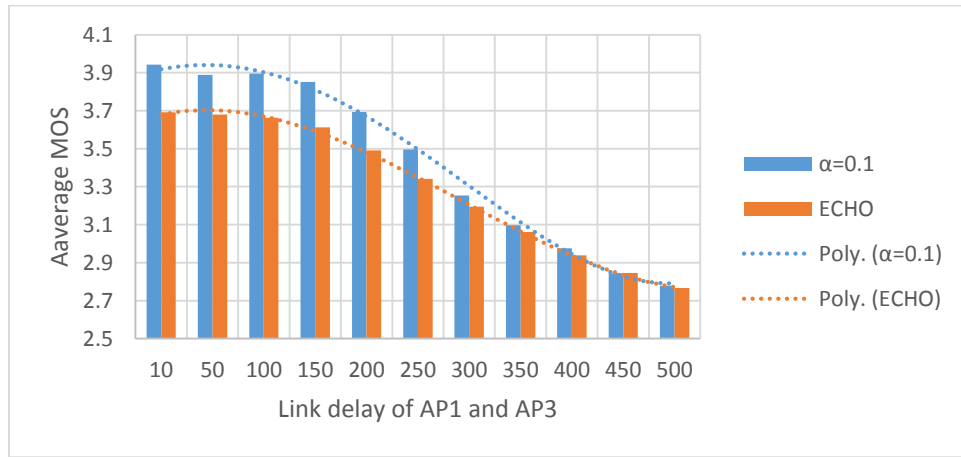


Figure 5.19 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.1$, AP2 = 450ms.

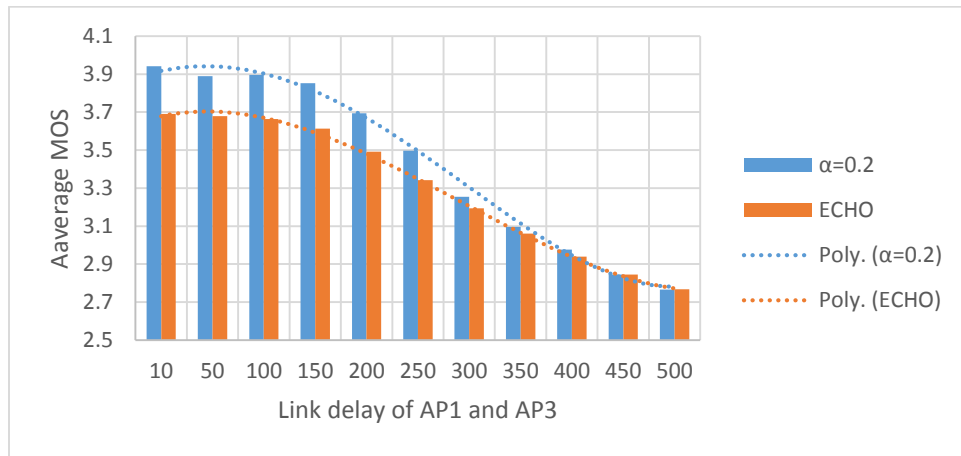


Figure 5.20 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.2$, AP2 = 450ms.

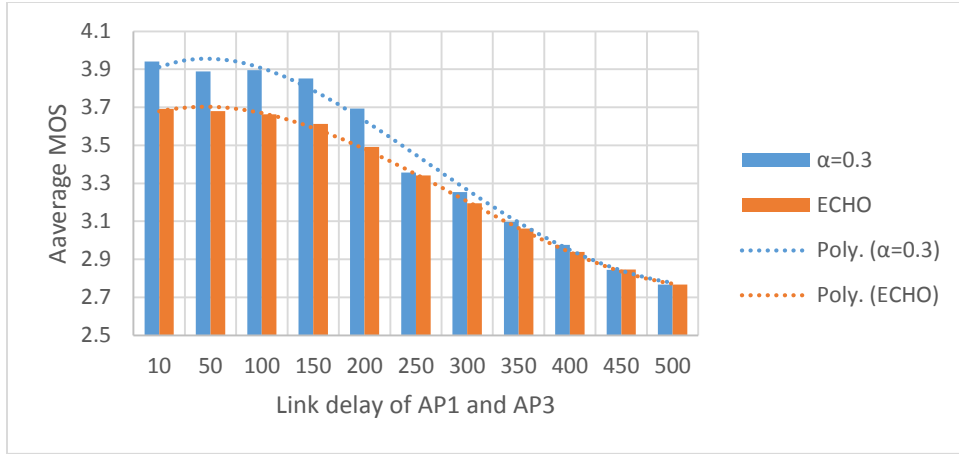


Figure 5.21 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.3$, $AP2 = 450$ ms.

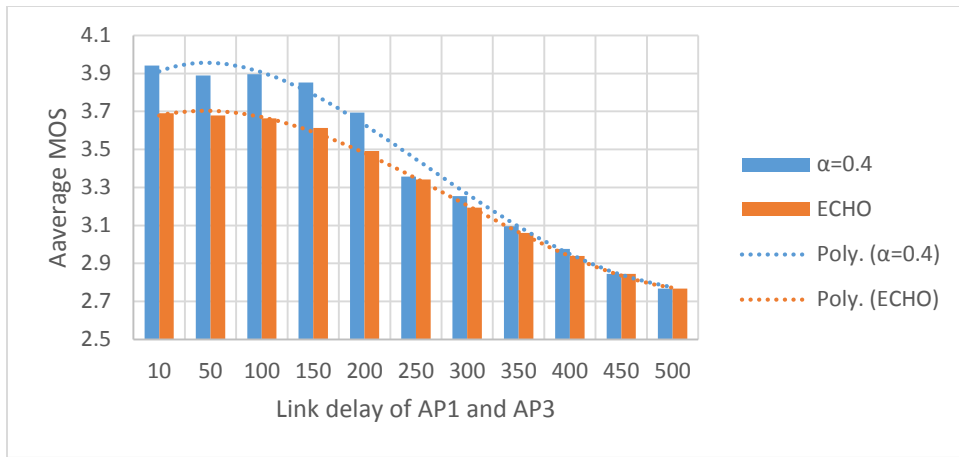


Figure 5.22 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.4$, $AP2 = 450$ ms.

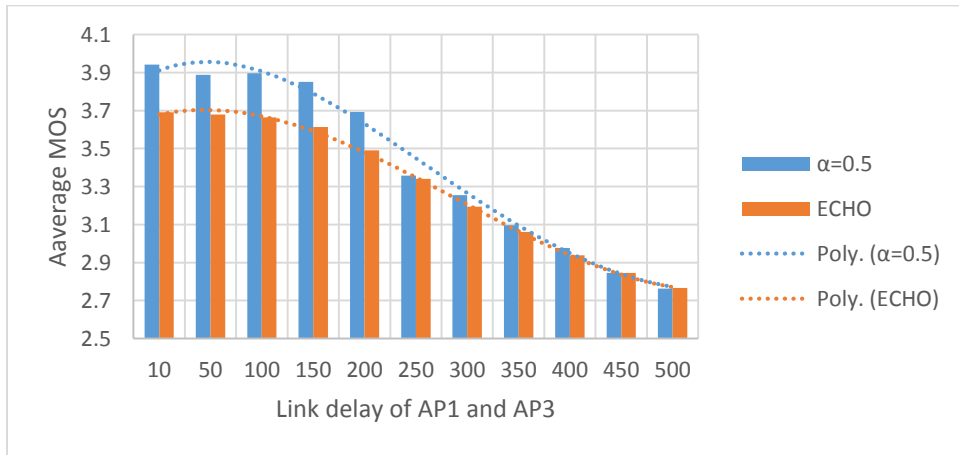


Figure 5.23 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.5$, $AP2 = 450$ ms.

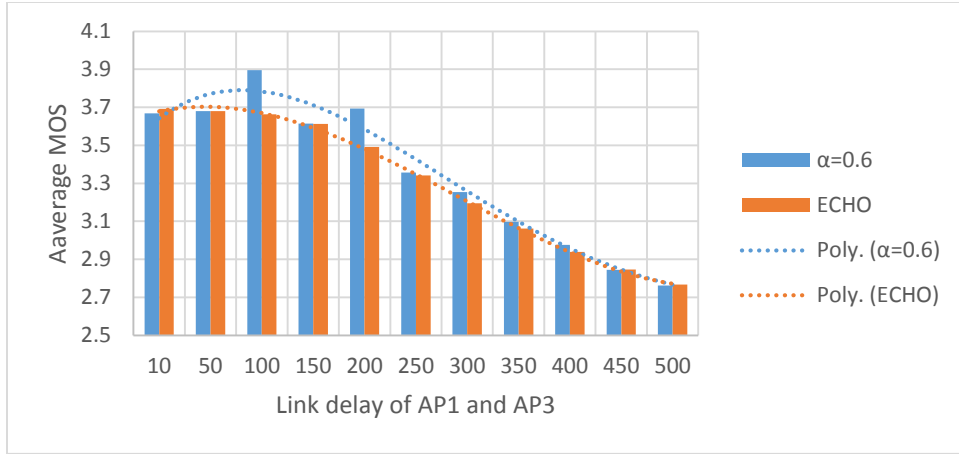


Figure 5.24 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.6$, $AP2 = 450\text{ms}$.

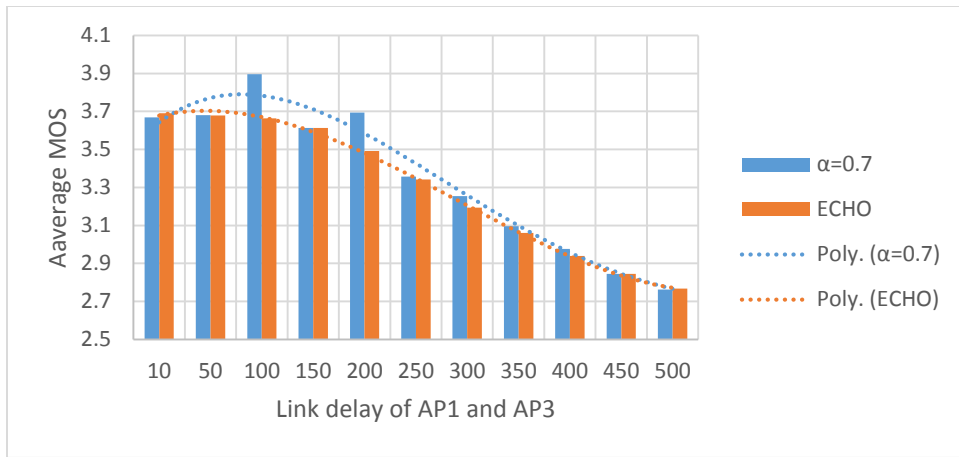


Figure 5.25 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.7$, $AP2 = 450\text{ms}$.

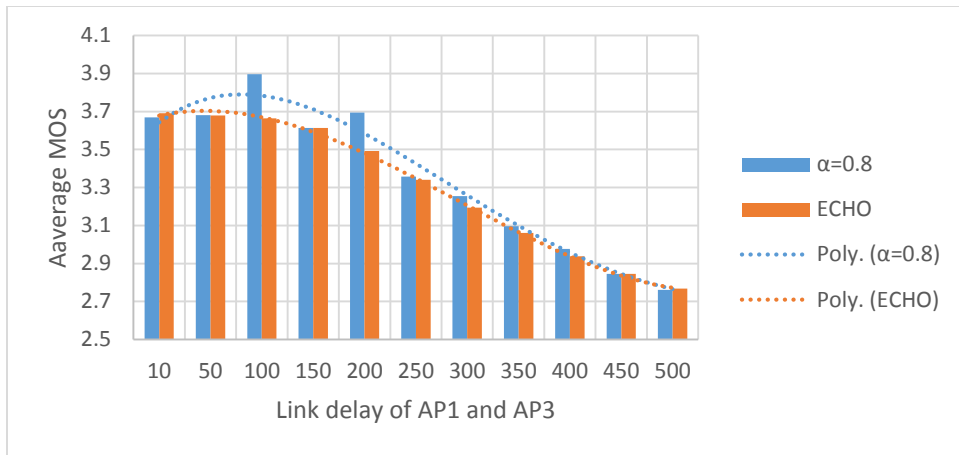


Figure 5.26 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.8$, $AP2 = 450\text{ms}$.

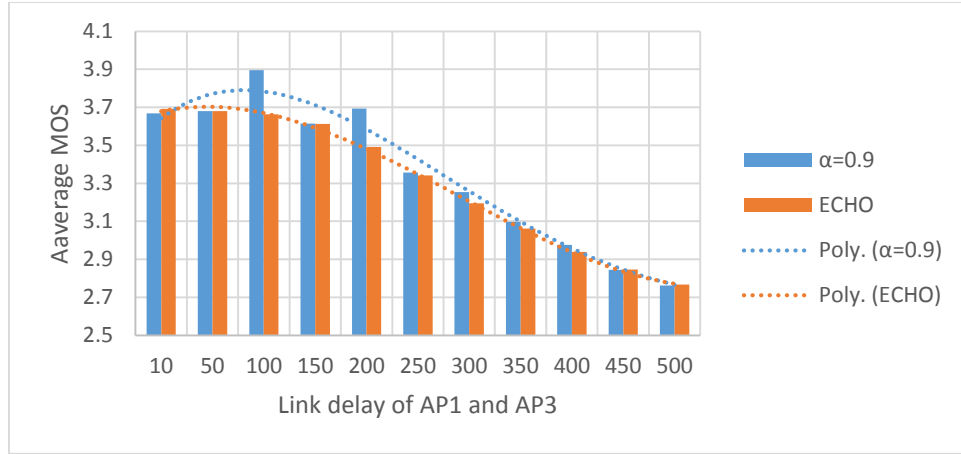


Figure 5.27 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.9$, $AP2 = 450\text{ms}$.

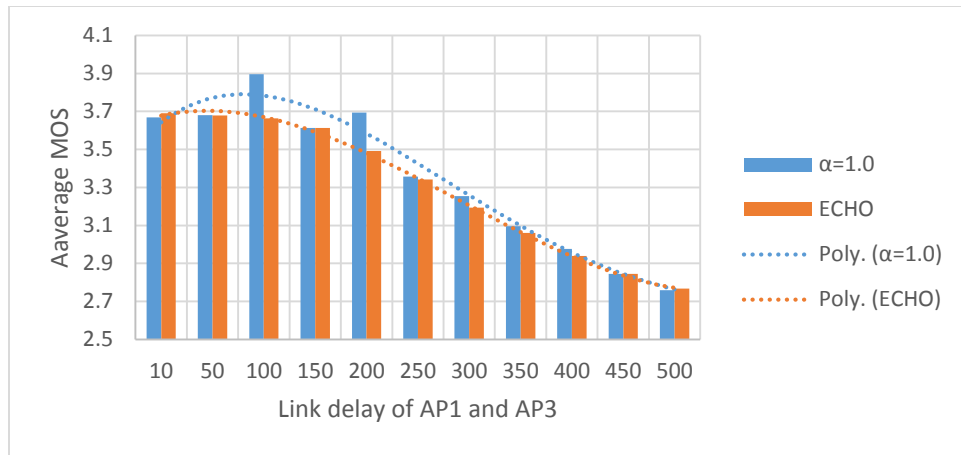


Figure 5.28 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=1.0$, $AP2 = 450\text{ms}$.

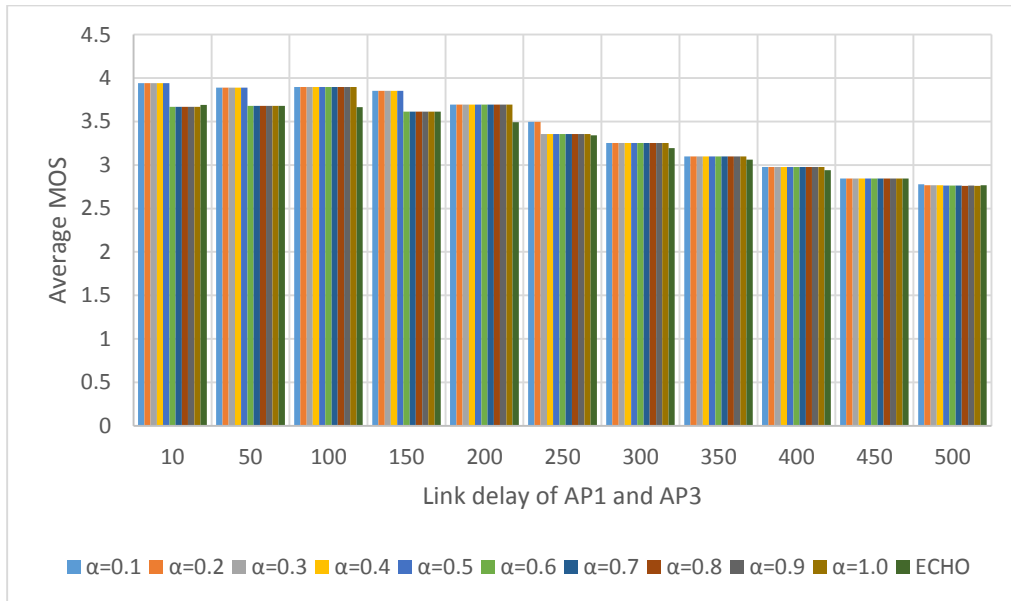


Figure 5.29 Summary of the performance of ANHA with different weights compared to ECHO for the second simulation

Figure 5.29 shows the summary of the handover performance using ANHA compared to ECHO. From this graph, overall performance of ANHA is better than ECHO for lower link delay at AP1/AP3 (10 to 250). However, both approaches have similar performance for higher link delays. This result conforms to the result in the previous subsection. Thus, in a nutshell, for sequential topology, ANHA is able to perform better than ECHO when the difference between the current network and the target network (disparity) is large. When the disparity is low, ANHA has similar performance as ECHO.

5.3.3. Overlapped Topology, AP1 delay = 10ms, AP2 delay: 10ms-500ms

In this section, the simulation the overlapped topology is discussed. This simulation was done using the topology in figure 5.7. As can be seen from the graphs in figure 5.30 to figure 5.32, in most cases ANHA has similar performance as ECHO, with the configuration of $\alpha = 0.2$ as the configuration with the best performance. The results shown in figures 5.33 to 5.38, it can be seen that when the link delay of AP 2 is higher, ANHA has lower performance than ECHO. This is due to the fact that ECHO utilize a MOS threshold for the handover process. When delay is 400 and above, the value of MOS becomes lower than the threshold used (3.1 in ECHO's case), thus no handover occurs. That is why ECHO has better results in that region. ANHA on the other hand, utilizes both MOS and RSS simultaneously. That is why the performance is somewhat lower, due to the handover trigger when the RSS of AP 2 peaks (as shown by the RSS profile in figure 5.7). Furthermore, when the delay becomes higher, the results from ANHA become lower because handover is still executed due to the RSS of AP2. However, the performance achieved by ANHA is still acceptable since the MOS is still above 4 with all ANHA settings. Besides that, ANHA with $\alpha = 0.2$ (figure 5.3) shows that ANHA is capable of getting at least the same performance as ECHO with some points showing that ANHA is better.

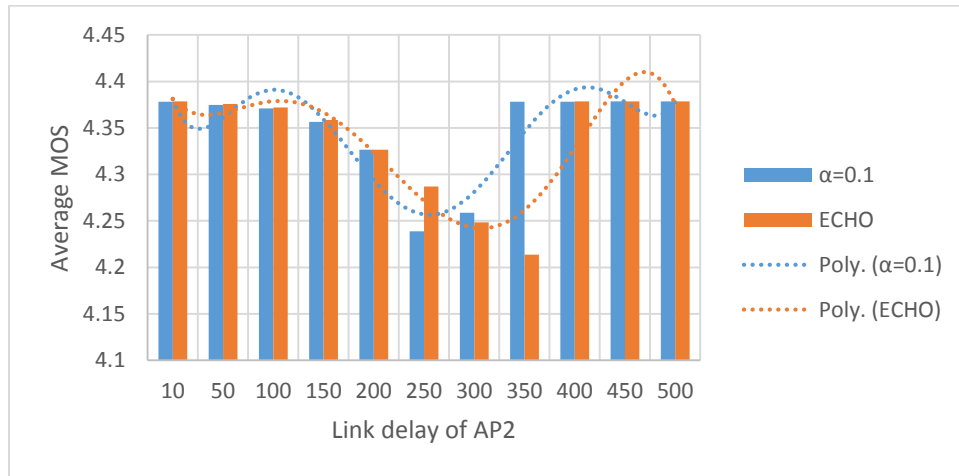


Figure 5.30 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.1$, AP1 = 10ms.

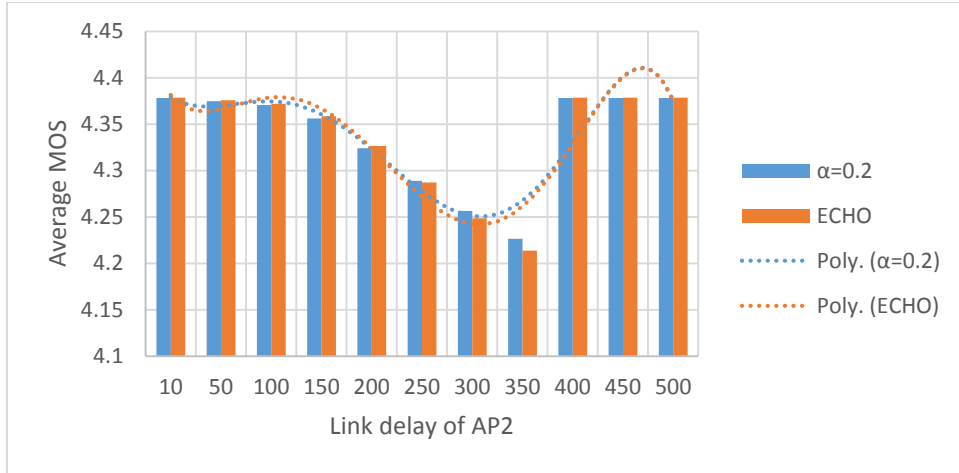


Figure 5.31 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.2$, $AP1 = 10\text{ms}$.

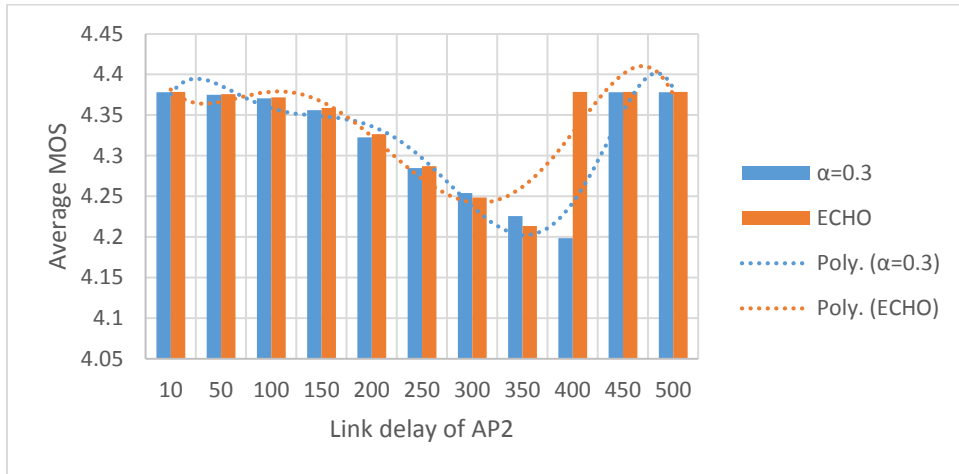


Figure 5.32 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.3$, $AP1 = 10\text{ms}$.

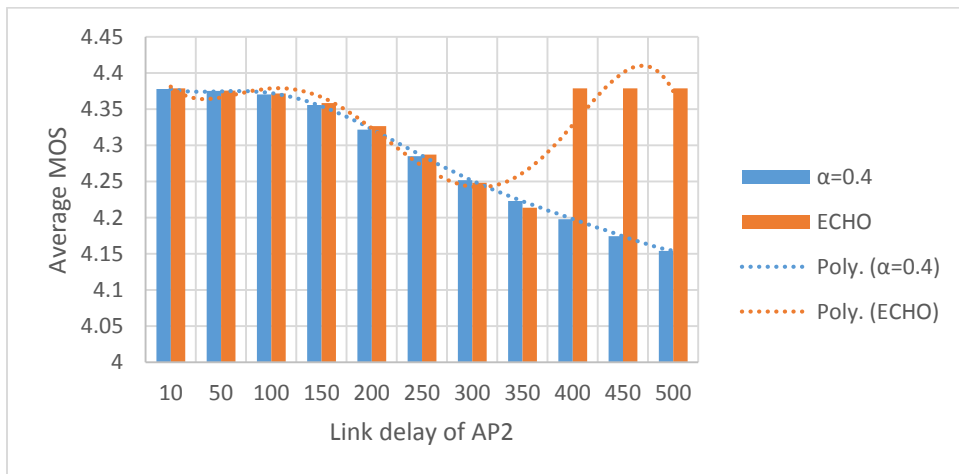


Figure 5.33 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.4$, $AP1 = 10\text{ms}$.

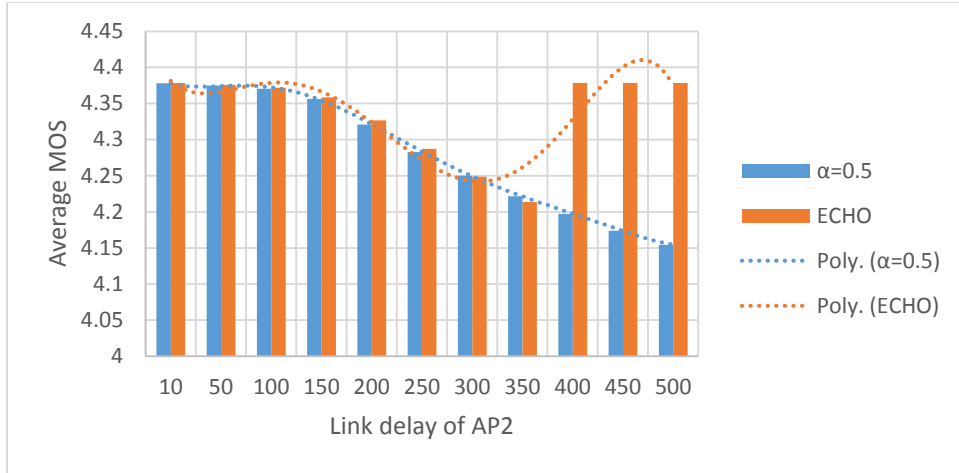


Figure 5.34 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.5$, $AP1 = 10\text{ms}$.

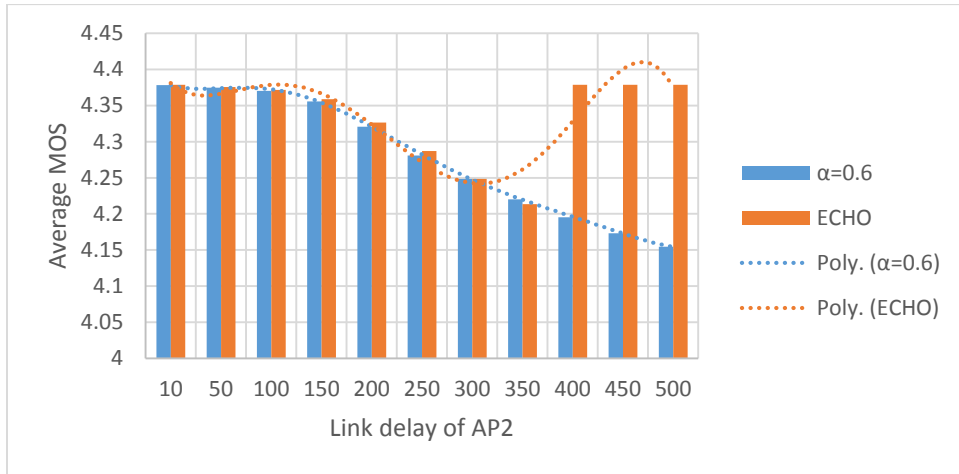


Figure 5.35 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.6$, $AP1 = 10\text{ms}$.

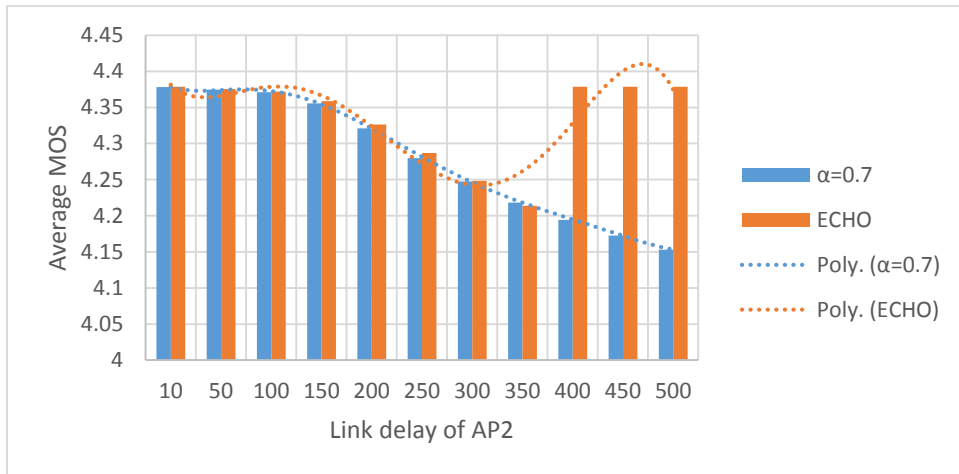


Figure 5.36 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.7$, $AP1 = 10\text{ms}$.

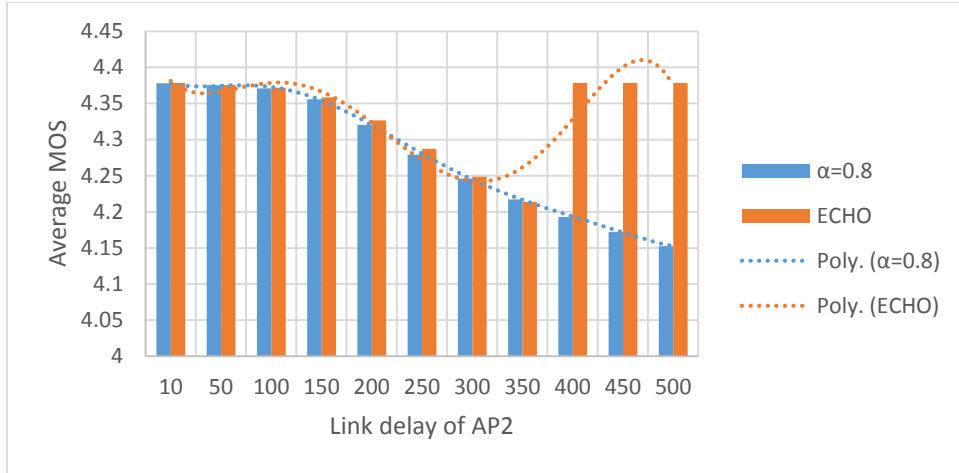


Figure 5.37 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.8$, AP1 = 10ms.

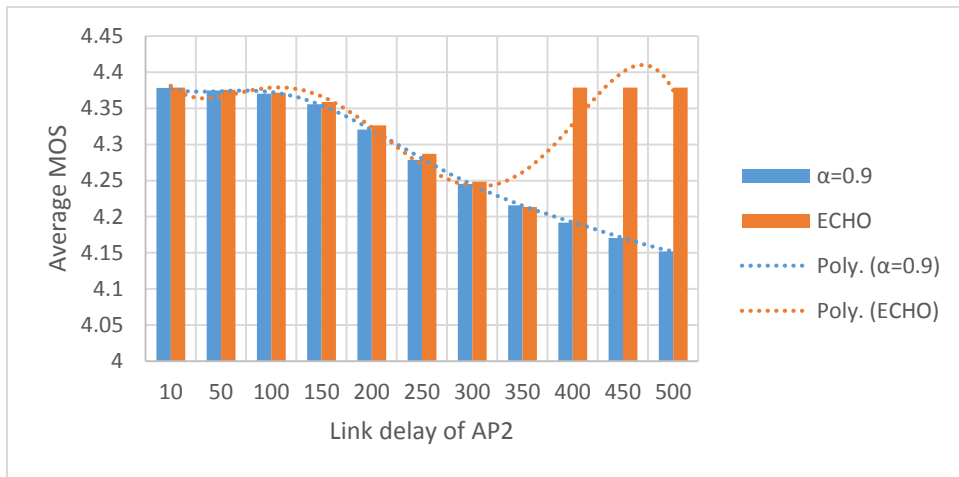


Figure 5.38 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.9$, AP1 = 10ms.

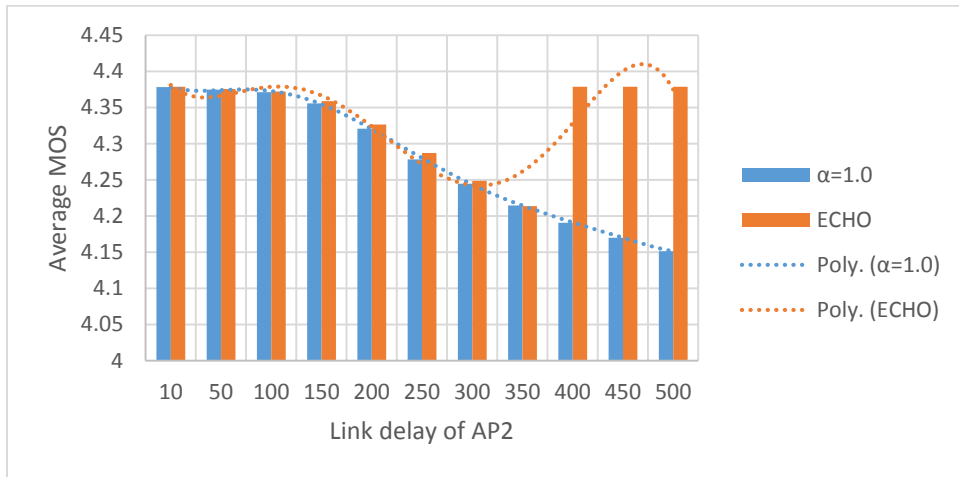


Figure 5.39 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=1.0$, AP1 = 10ms.

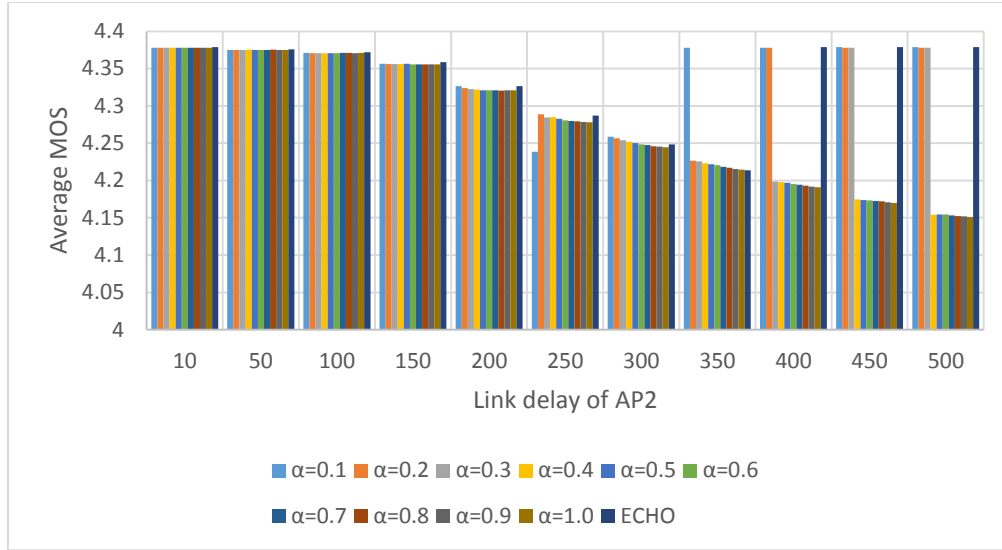


Figure 5.40 Summary of the performance of ANHA with different weights compared to ECHO for the third simulation.

Figure 5.40 shows the summary of all ANHA settings for this simulation. The results shows that ANHA's performance is better than ECHO when $\alpha = 0.2$. The results here confirm that ANHA takes into account both MOS and RSS simultaneously when deciding the handover trigger. Furthermore, the performance of other settings for ANHA is still acceptable, since the average MOS is still above 4 (which is actually very good quality).

5.3.4. Overlapped Topology, AP1 delay = 450ms, AP2 delay: 10ms-500ms

This is the final simulation done in evaluating the performance of ANHA compared to ECHO. Figure 5.40 to 5.49 show the results obtained for this simulation. From observation of these results, it can be seen that all configurations of ANHA outperforms ECHO accept for when the link delay of AP 2 is configured to 500ms. This result is also due to the simultaneous consideration of MOS and RSS in ANHA's decision algorithm. Despite this drawback, overall, ANHA outperforms ECHO in maintaining the average MOS level.

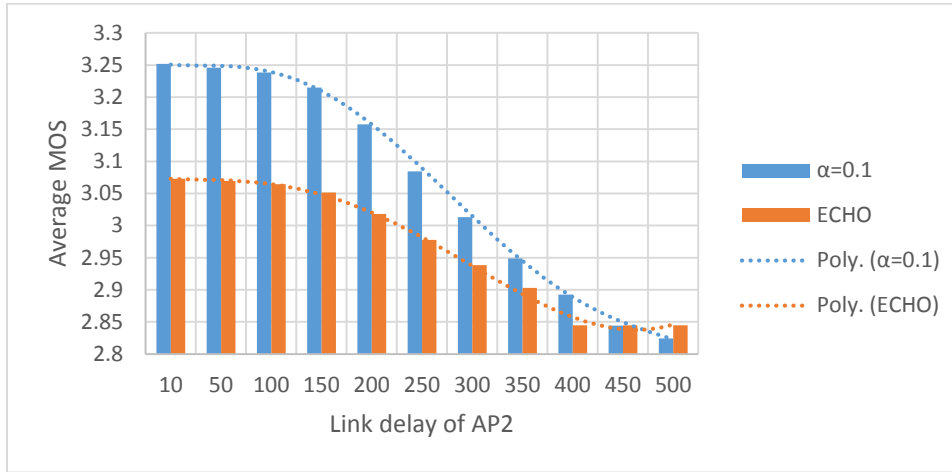


Figure 5.41 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.1$, AP1 = 450ms.

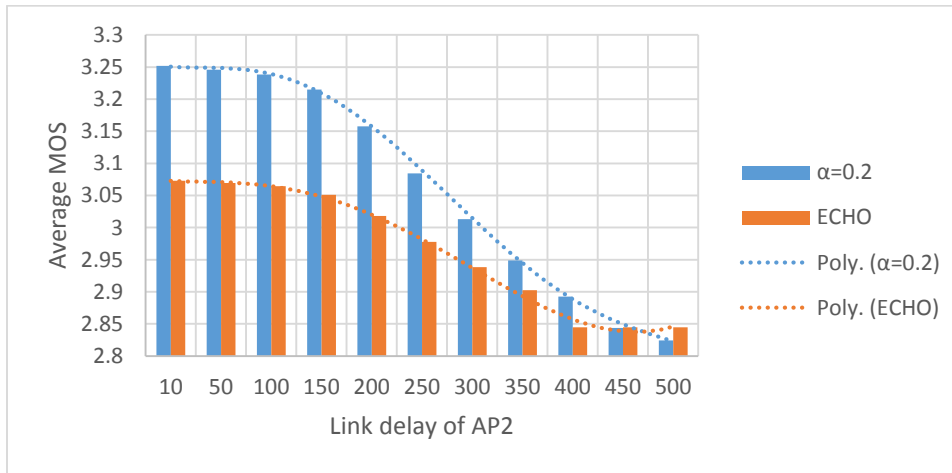


Figure 5.42 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.2$, AP1 = 450ms.

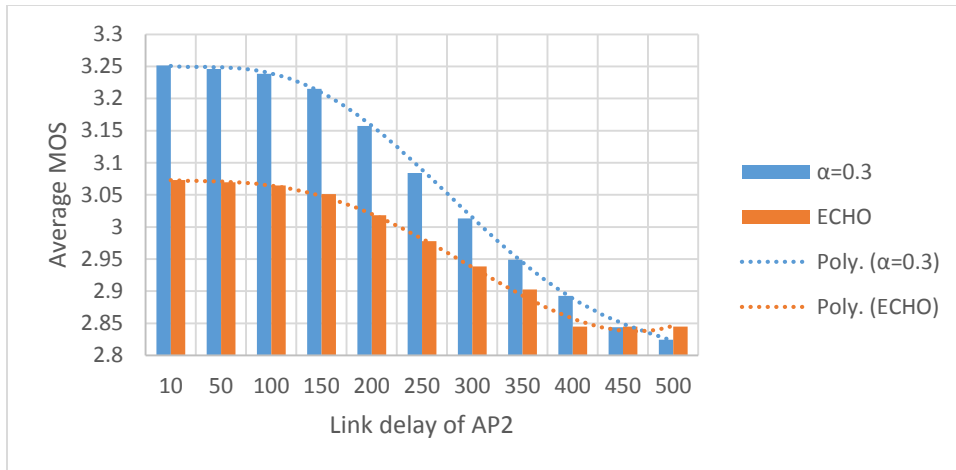


Figure 5.43 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.3$, AP1= 450ms.

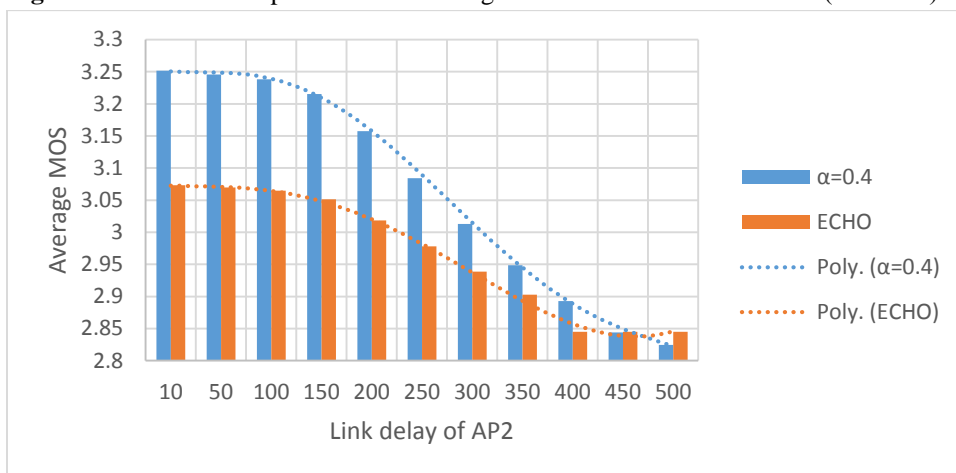


Figure 5.44 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.4$, AP1 = 450ms.

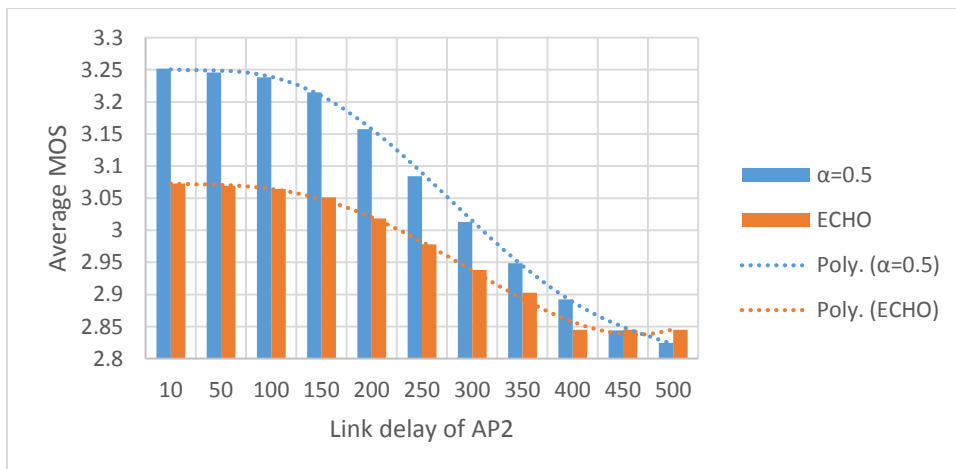


Figure 5.45 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.5$, AP1 = 450ms.

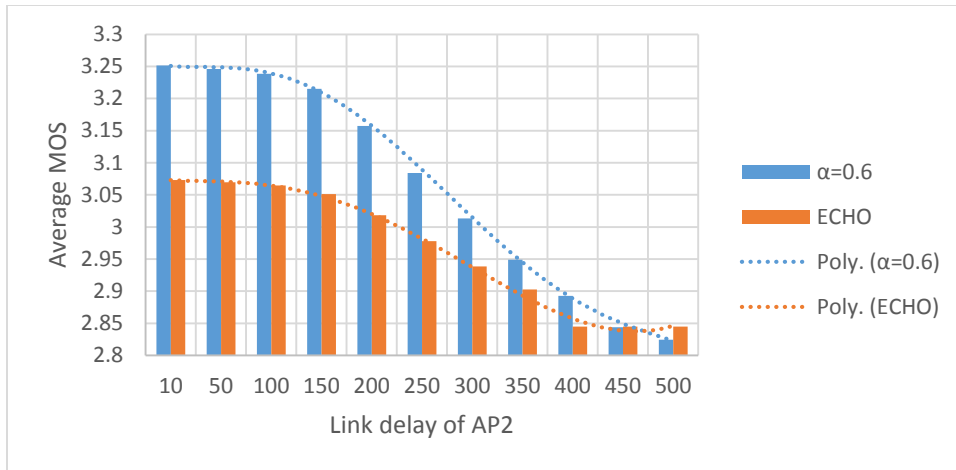


Figure 5.46 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.6$, AP1 = 450ms.

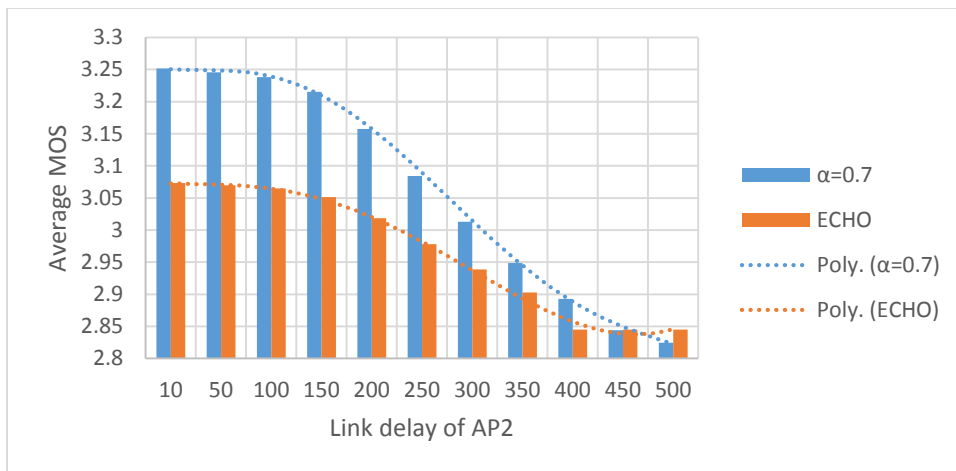


Figure 5.47 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.7$, AP1 = 450ms.

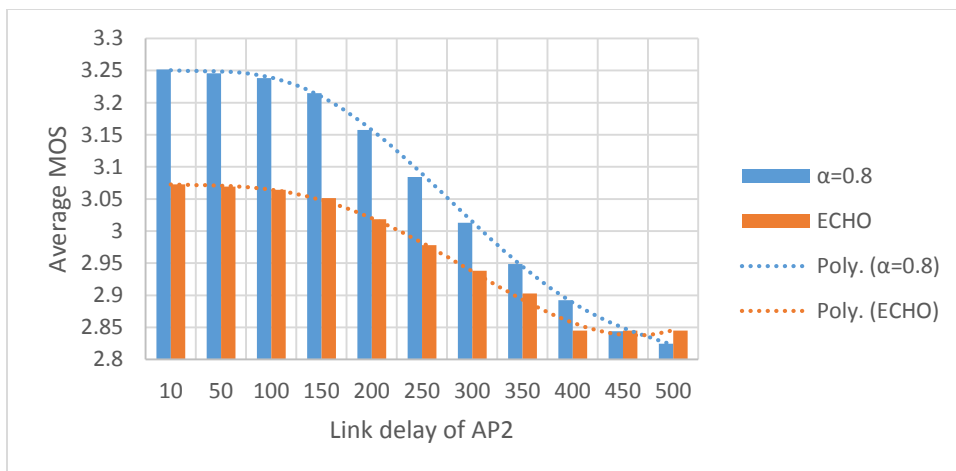


Figure 5.48 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.8$, AP1 = 450ms.

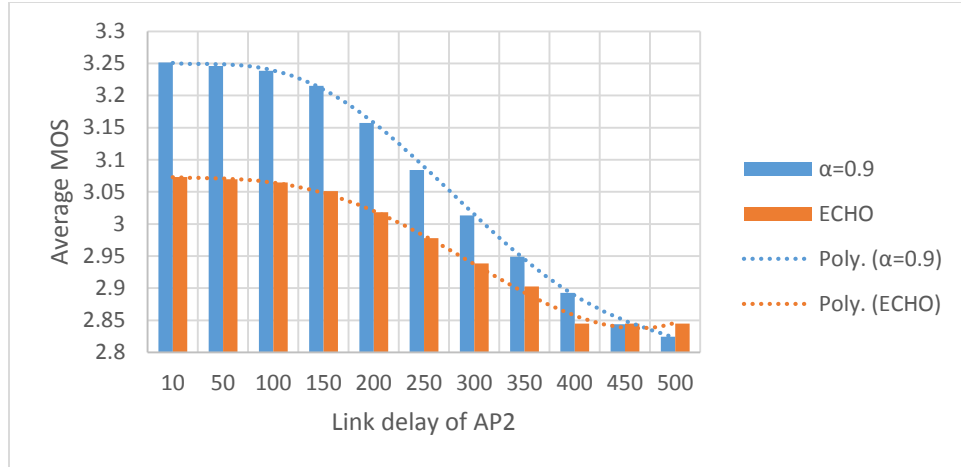


Figure 5.49 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=0.9$, AP1 = 450ms.

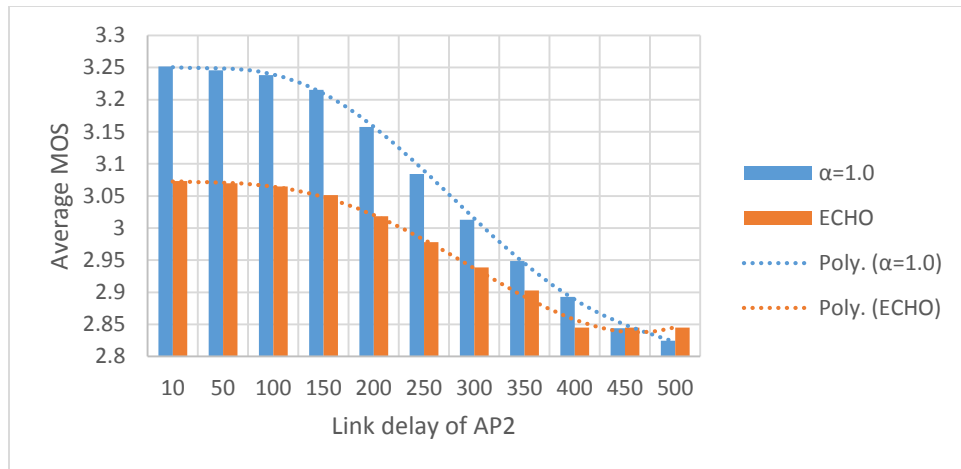


Figure 5.50 Detailed inspection of the average MOS for ECHO and ANHA (blue bars) with $\alpha=1.0$, AP1 = 450ms.

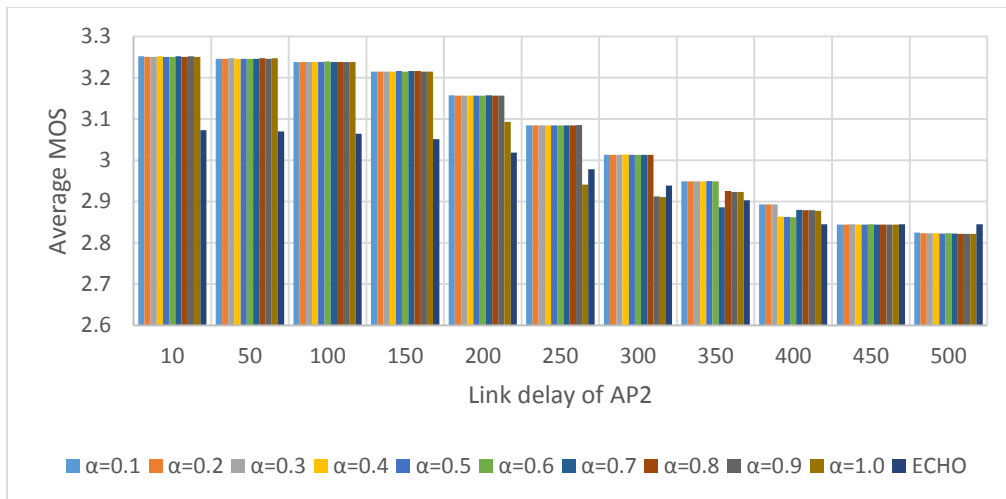


Figure 5.51 Summary of the last simulation.

Figure 5.51 shows the summary of all results by ANHA and ECHO. Here, it can be seen that ANHA outperforms ECHO in overall when the link delay of AP 2 is lower than 500ms. The results here further conform to the previous results, where ANHA performs better when the disparity of instantaneous MOS level between the current network and the target network is apparent. When the disparity is small, ANHA still outperforms ECHO in some situations

5.3.5. Handover triggering timing comparison

Part of the reason why ANHA outperforms ECHO is the triggering characteristic opted by ANHA. Figure 5.51 and 5.53 are samples of the instantaneous MOS obtained by ANHA in Overlapped topology and Sequential Topology respectively. Meanwhile, figure 5.52 and 5.54 are the instantaneous MOS obtained by ECHO in Overlapped and Sequential topologies respectively. The vertical axis is the instantaneous MOS used for the handover trigger, and the horizontal axis is the simulation time in seconds. As can be seen from figures 5.51 and 5.52, ANHA chooses the path with higher instantaneous MOS for a longer duration compared to ECHO in overlapped topology. In the overlapped topology, the delay on AP2 is varied. To understand more clearly, let's take the link delay of AP2 = 450ms, which will give a high instantaneous MOS on AP2. The MH host using ANHA selects AP2 for nearly 64 seconds whilst ECHO only select AP2 for 36 seconds only which is nearly half of ANHA. That is how ANHA achieve high average MOS in the overlapped topology.

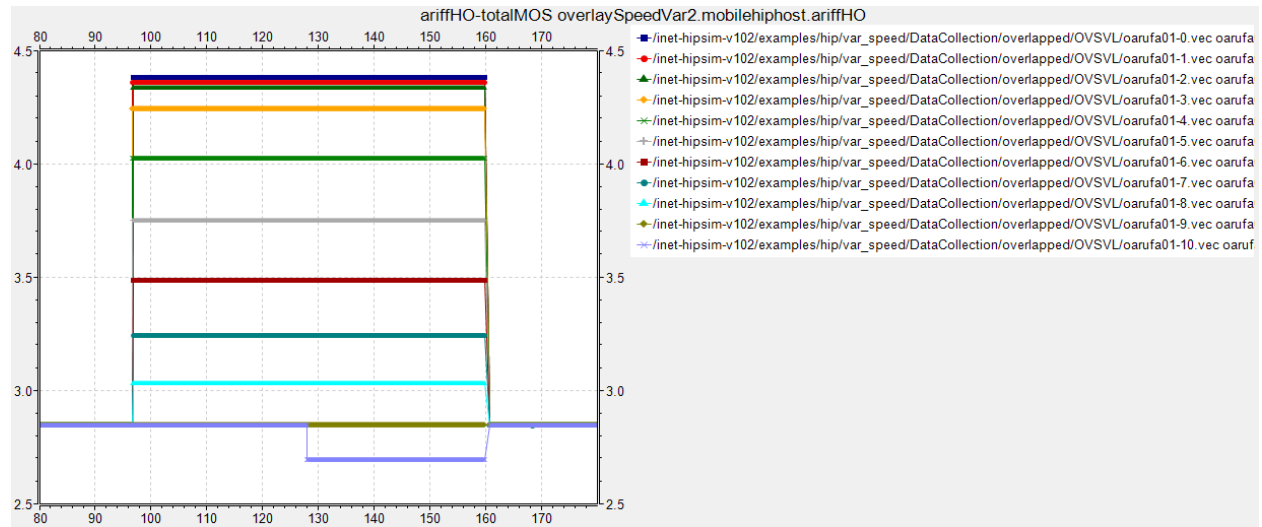


Figure 5.52 ANHA handover behavior in overlapped network.

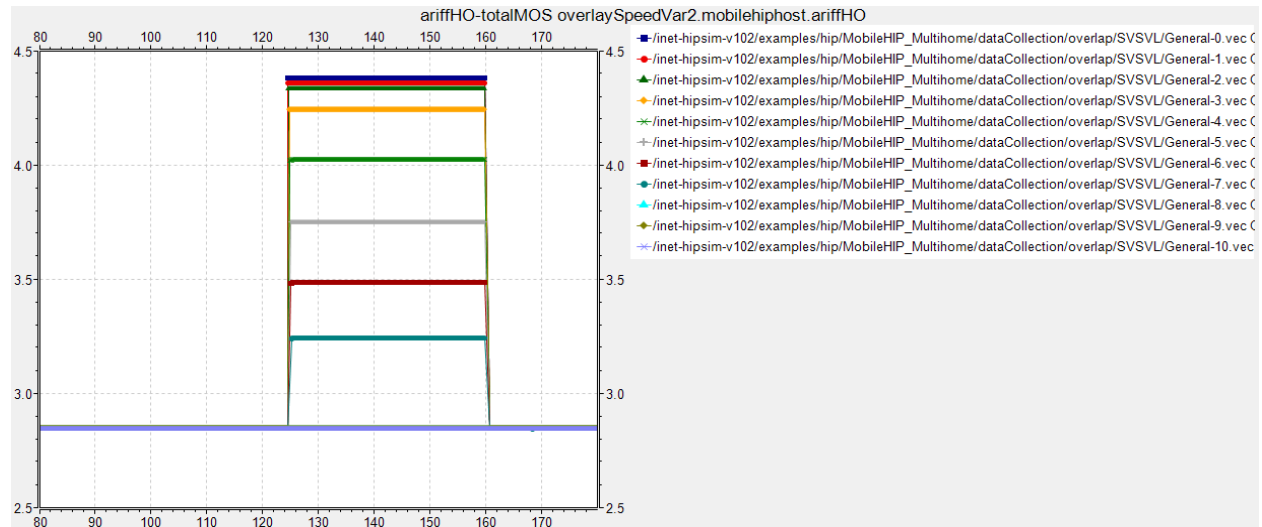


Figure 5.53 ECHO handover behavior in overlapped network.

Meanwhile, for the sequential topology, the MH with ANHA shorten the time it selects AP with low instantaneous MOS to maximize the overall average MOS. For example, in figures 5.53 and 5.54, the ANHA-equipped MH connects to the network with the lowest MOS (the lowermost purple line) for only around 24 seconds. Whilst for ECHO, due to the use of static threshold, chooses the same path for nearly 40 seconds. That's why ANHA can perform well in retaining high average MOS in most of the test scenario in this chapter.

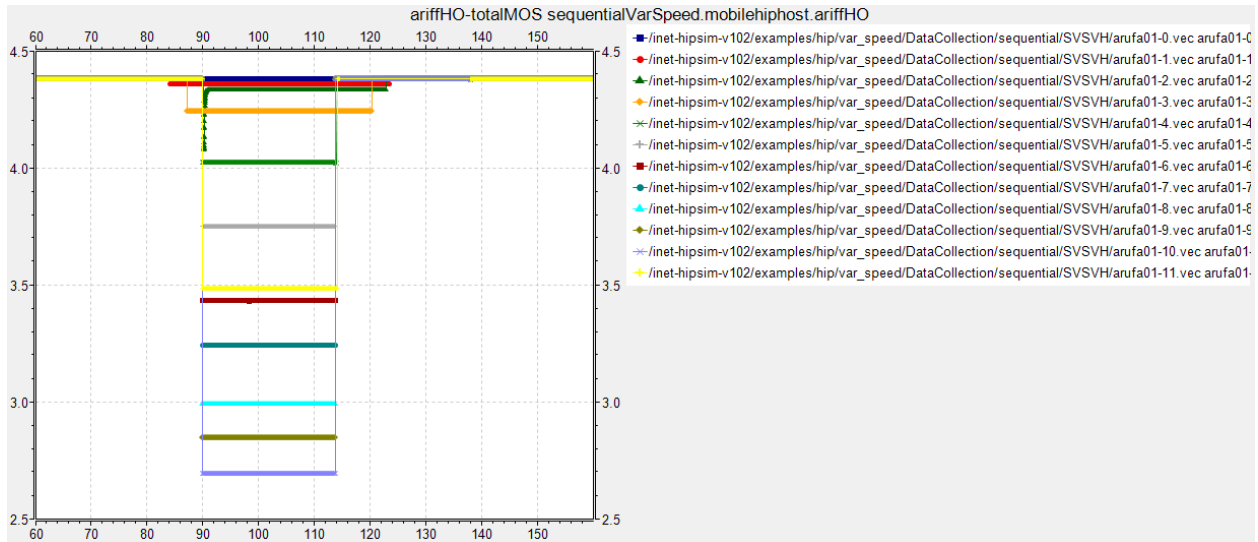


Figure 5.54 ANHA handover behavior in sequential network.

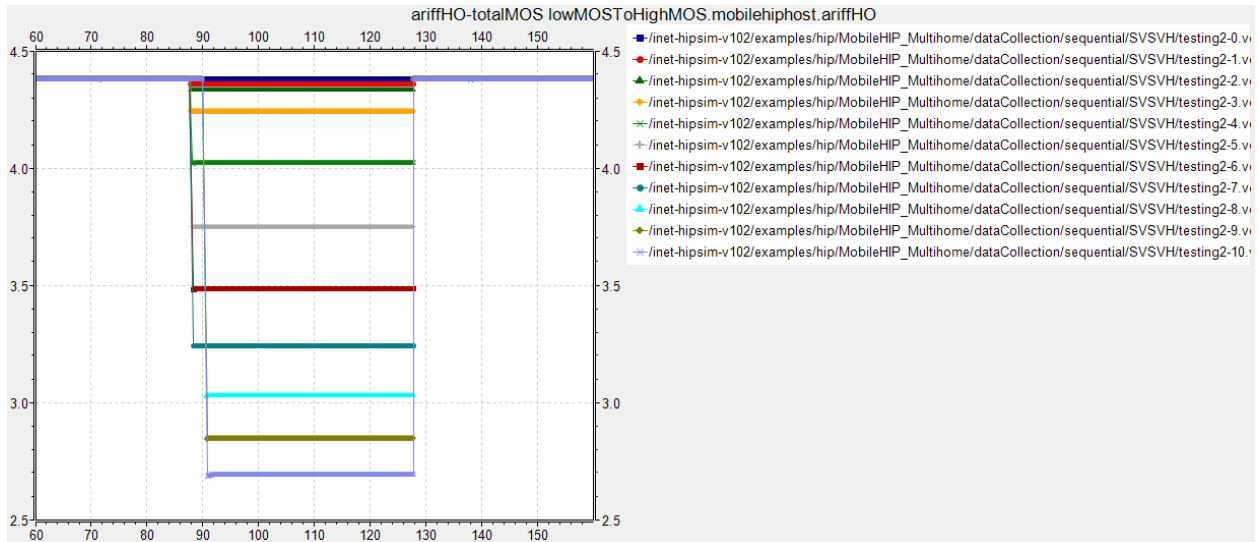


Figure 5.55 ANHA handover behavior in sequential network.

5.4. Discussion on Different Possible Scenario

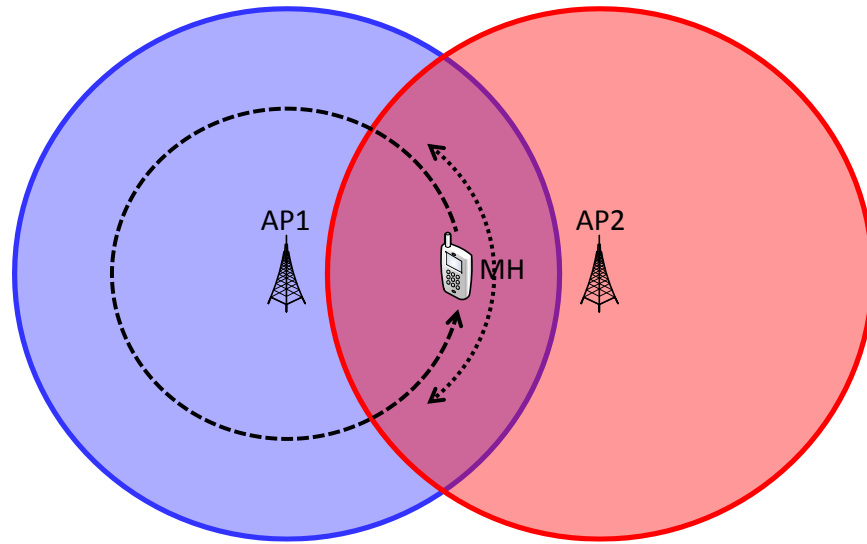


Figure 5.56 Scenario 1: Adjacent cells.

Let the MH moves in a curving/circular motion as shown in figure 5.56. The RSS of AP1 will be theoretically constant, since the distance between AP1 and MH does not change. The RSS of AP2 on the other hand will gradually increase when the MH enters AP2's coverage boundary and gradually decrease when the MH moves towards AP2's coverage boundary.

For ease of discussion, summaries of the decision flow of ECHO and ANHA from figure 5.5 is included.

ECHO: first checks the MOS threshold, if not triggered, it will monitor the MOS and RSS of the primary network. Then if current network MOS falls below the threshold or the RSS falls below the RSS threshold, handover will still be executed.

```
If (MOStarget>MOSthreshold)
    Handover
else if (MOScurrent<MOSthreshold or RSScurrent<RSSthreshold)
    Handover
```

ANHA: first checks the P value of the target network, if it is larger than the P of the current network, handover will be triggered. If not triggered, it will monitor the P of the current network and the RSS. If the P of the current network falls below the target network or the RSS falls below the RSS threshold (take note this is the threshold adapted to the MH's velocity), the handover will be executed.

```
If (Ptarget>Pmain)
    Handover
else if (Ptarget>Pmain or RSScurrent<RSSthreshold)
    Handover
```

Since the RSS values of AP1 does not change, the handover trigger mainly depends on the MOS and the RSS of AP2. This scenario can be divided into three types. Type 1: MOS AP1 is better than AP2, Type 2: MOS AP1 similar to AP2 and Type 3: MOS AP2 better than AP1.

ANHA: for Type1 and Type3, the results will be similar to results obtained from the simulation since there are clear distinctions between the P_N for AP1 and AP2. Meanwhile, for type 2, the handover will be determined by the RSS value. However, the end result is also the same as from the simulation, since the MOS value does not change.

ECHO: bear in mind that two static thresholds are used (RSS and MOS); for Type 1 the result will be worse than ANHA if the handover is triggered by the thresholds. If the thresholds were not triggered, then the results will still be similar to ANHA, since ANHA maintains the connection to AP with better MOS as long as possible. For Type 2, the results will be the same as the simulation, since the MOS level will not change even if handover is triggered. For Type 3, ECHO's performance will be lower than ANHA, since the handover to AP2 will be triggered at a later time, compared to ANHA.

In a nutshell, the outcome will still be similar to the simulation results, thus for this scenario, it can be concluded that ANHA works better than ECHO.

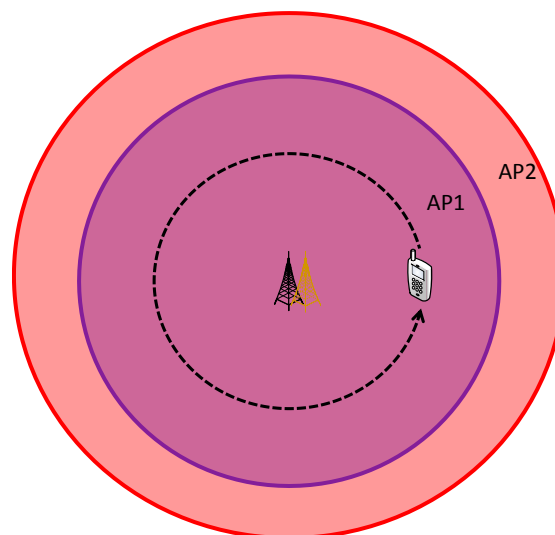


Figure 5.57 Scenario 2: Overlapped cells.

Another scenario that can be considered is scenario 2 as shown in Figure 2. In this scenario, the RSS of both networks theoretically does not change. This scenario can be divided into two types; Type 1: AP1 has better MOS than AP2 and Type 2: AP2 has better MOS than AP1.

ANHA: For Type1, ANHA will always select AP1 since AP1 has better MOS, while for Type 2, AP2 will always be selected.

ECHO: For Type 1, if ECHO connects to AP1 first, the MH will have same performance as ANHA. However, if MH connects to AP2 first, the handover consideration process will not start, since the RSS of the connected AP will not deteriorate, thus the MH will continuously be connected to AP2 and maintain low overall MOS. It is also the same for type 2.

5.5. Concluding Remarks

ANHA combines the received signal strength and the Mean Opinion Score of the networks available to the MH for triggering a handover, has been proposed. A simulation on OMNeT++ was conducted using the proposed approach and an existing method known as the Endpoint Centric Handover (ECHO). The results showed that the proposed method can achieve better average MOS performance compared to ECHO when the disparity of MOS between the current serving network and the target network is large. Furthermore, it can be concluded that the proposed approach maximizes its connection time to a network with high MOS level and minimizes its connection time with a network with low MOS level, thus showing better results than ECHO. Furthermore, from the results in this chapter, it can be seen that the weight for the heuristic component, α , is best set to 0.2. This due to the good and stable results obtained by this configuration.

CHAPTER 6

Mobile Host Velocity Awareness

6.1. Introduction

When discussing mobile communication environments, one of the most important processes that have to be considered is the handover process. This is the process where the Mobile Host (MH) transfers its wireless connection and application sessions from the current serving Access Point (AP) to another adjacent AP. The handover process is especially crucial not only in the current scenario but also in the future since the number of mobile users has drastically increased during the recent decade. Nowadays, almost everyone owns a mobile device. Sometimes, two or more devices are carried at the same time. This fast growth has incited the rapid evolution of RAT (Radio Access Technology); for example, the RAT has evolved from the first generation of wireless access to the current fourth generation, which includes LTE/LTE-A [2] and WiMAX [1] just within the recent decade [99], [100]. This evolution has created a mix of different wireless accesses that overlaps with one another structuring a heterogeneous network [4]. Furthermore, the computational capabilities of the current mobile devices are on a par with laptops or desktop computers. They are also equipped with additional components such as GPS and accelerometers, as well as multiple antennas to support the different available Radio Access Technologies (RAT). With these features, the MHs are capable of more complex computation, capable of obtaining their velocity and location information, and capable of connecting to different RATs. These capabilities open new possibilities to enhance the performance during the handover process.

The users' expectation grows with the advancement of the MHs features and the applications that are available on mobile devices. Since most of the applications require the Internet connection, the users expect to get constant and continuous service from the service providers. The MH's connection quality is particularly critical for real time applications that are generally delay and packet loss sensitive. The most common handover technique currently used is the hard handover known as break-before-make approach. This method breaks the MH's connection to the current serving AP first, and then establishes a new connection to a selected neighboring AP. For a homogeneous network, this method is quite feasible because the target AP has already obtained the information on the MH (network based handover, i.e. UMTS and LTE) when handover occurs.

However, for a heterogeneous network, where the networks have different RATs and different core networks, the MH information is not always available in the target AP specifically in the networks operated by different network carriers. Thus when handing over from one RAT to another (this process is also known as vertical handover), the disruption time (the time duration when the MH is not connected to any AP during the handover process) will be very long; it can range from several hundred milliseconds up to several seconds. This kind of disruption visibly degrades the service quality of real time applications, such as VoIP and video conference in particular.

Moreover, it should be noted that an application session is broken during a vertical handover process, unless some mobility management protocol is applied. The most popular mobility management protocol is the Mobile IP (MIP) [26]. This approach manages the mobility-related information, but it suffers from high handover latency due to the binding update process required to maintain the connection between the MH and its CH (Corresponding Host). Some other variants of MIP, such as the HMIP (Hierarchical MIP), FMIP (Fast MIP), FHMIP (Fast Hierarchical MIP), CIP (Cellular IP) and HAWAII (Handoff-Aware Wireless Access Internet Infrastructure) [101], [102], have been developed to alleviate the issues faced by MIP. Nevertheless, these methods are very dependent on the hierarchical structure of the core network. According to Handley [7], not much deployment of these technologies can be seen in the real networks due to the extensive modifications on the existing system. These modifications will incur a huge cost which is not very attractive for mobile network carriers.

Consequently, researchers are starting to study alternative approaches to avoid this issue. One of the available alternatives is to use endpoint centric approaches to circumvent or minimize the necessary changes. Since the specifications of existing mobile devices have been improved, it is possible to push the complexity of the handover process to the endpoints. There are several existing methods in the literature; some of them are: HaMAT (Host-based autonomous Mobile Address Translation) [103], SIGMA (Seamless IP diversity based Generalized Mobility Architecture) [46], and ECHO (Endpoint Centric Handover) [14]. These methods follow very similar approaches; each method uses multi-home (MH can connect to two or more networks at the same time via different Network Interface Card (NIC) or NIC virtualization) that enables the MH to establish a connection to a target network before disconnecting or losing the connection to the current network (this is also known as soft handover or make-before-break approach, which is the opposite approach of hard handover). The differences among these approaches are the protocols to realize the multi-home environment and the parameters considered to trigger the handover. HaMAT uses the network layer Mobile Address Translation (MAT) protocol whilst SIGMA and ECHO use Stream Control Transmission Protocol (SCTP). As for the handover triggering parameters, SIGMA only considers the Received Signal Strength (RSS); ECHO is an enhancement of SIGMA, thus it also uses RSS but with the addition of ITU-T E-Model's Mean Opinion Score (MOS). Similarly, HaMAT also uses the RSS with the addition of MH velocity information to trigger the handover [104].

The RSS and MH moving speed information are important to maintain the MH connection to the AP, whilst the MOS value is important to determine the level of user satisfaction that the user

will experience from the service. These information is useful to timely achieve handover triggering [105], [106] and to optimize the handover execution algorithm [107]–[109]. However, for mobile users moving at undetermined speed (they may move at a slow pace like pedestrians, or they may move at higher speeds, i.e. moving on a train), the velocity will also affect the handover process itself. For example, if the total time needed to handover into and out of a network is larger than the MH's connection time (dwell time) within a target network, the MH will not be able to utilize that network since it has to start the handover out of that network immediately. This is considered as an unnecessary handover. Moreover, in the worst case, if the MH moves out of that network before completing the handover process, a handover failure will occur and this will cause significant disruptions to the application session. Avoiding such unnecessary handovers and handover failures is also important in ensuring the quality experienced by the users. Thus, the velocity information should also be well-considered in enhancing the handover process [97].

In the literature, the velocity of the MH has been used for: (1) Estimation of the MH dwell time (the duration time staying within a network, this is related to the travel distance estimation) which corresponds to the traveled distance within a network, (2) the Handover Necessity Estimation (HNE), and (3) the Handover Failure Estimation (HFE) (which is directly related to Handover Failure Avoidance (HFA)) [110]. The initial framework for HNE and HFE was established by Yan et al. [110] in the situation where MH enters a WLAN cell from a 3G network. The model is based on the geometrical angle of arrival and angle of departure; assuming that the MH moves in a straight line within the network, the MH's travel distance and the distance threshold to trigger the handover are obtained. A similar study was reported by Abrar et al. [104] using the same concept with a different HNE and HFE approach. Both of these methods were evaluated by Monte Carlo simulation based on the probabilistic model. The evaluation resulted in some significant improvement compared to a classical method by Mohanty [111]. The evaluation metrics were the number of handover reduction and the probability of handover failure. These types of evaluations are useful to show the effectiveness of their proposed methods, but do not show the direct effects on the Quality of Service (QoS) and Quality of Experience (QoE) received by users (i.e. packet loss rate). Furthermore, in [110], they showed that their proposed travel distance estimation (TDE) method did not work well for a MH with velocity lower than 10 meter per second (m/s). To the best of my knowledge, little research has considered velocity for the handover process in existing literature.

Thus, the study in this chapter aims to glean on the issues of high velocity estimation error for velocities higher than 100km/h in [97], high TDE error for velocities lower than 10m/s in [110], as well as providing insights on the effects of degradation by handover failures in terms of packet loss. Moreover, new methods to estimate the MH's velocity, MH's travel distance within a network, and HFA (namely, the Adaptive RSS Threshold approach) are proposed and discussed thoroughly. Ultimately, the contribution of this study can be summarized as follows:

1. Propose a new approach to estimate the MH's velocity, namely Velocity Estimation Method (VEM), using Artificial Neural Network (ANN) as the estimation engine, with the RSS and

the rate of RSS change (ΔRSS) as the input. It is generally assumed that the velocity information can be obtained from the embedded GPS or accelerometer. However, sometimes this information is not available due to the power saving schemes of the MH or other factors such as no embedded GPS/accelerometer, no GPS coverage area or high GPS error. Therefore, the proposed method can provide velocity information with improved accuracy without using GPS/accelerometer.

2. Discuss the reason for the accuracy issues of the TDE method proposed in [110] and propose a time interval described as the warm-up time, in this chapter, that can be used to increase the accuracy of the TDE technique. This approach may be used to improve the efficiency of existing Unnecessary Handover Avoidance (UHA) method.
3. Propose a new Adaptive RSS Threshold (ART) method based on the TDE. This method can significantly reduce the number of packet loss during the handover process by avoiding handover failures.

The rest of this chapter is organized into five sections with their respective sub-sections. Section 6.2 describes the existing travel distance estimation which is incorporated in the proposed work as well as the concept and notations generally used throughout this chapter. Next, section 6.3 discusses the proposed VEM method based on ANN; this section is divided into three subsections. In the first subsection, the underlying concept of the proposed ANN-based VEM is disclosed followed by the proposed model in the second subsection and the results in the third subsection. Then, section 6.4 focuses on TDE; here the warm-up time concept is introduced to increase the accuracy of TDE, which will indirectly increase the effectiveness of the unnecessary handover avoidance and adaptive RSS threshold approach. This section is divided into two subsections; section 6.4.1 discloses the method to get the suitable warm-up time whilst section 6.4.2 shows the effectiveness of using the proposed warm-up time. After that, section 6.5 describes the proposed Adaptive RSS Threshold. Finally, this chapter is concluded in section 6.6.

6.2. Main Concepts and Notations

To make the discussions in the following sections easier, some concepts and notations used throughout this chapter are provided in this section.

6.2.1 Main Concepts related to and used in the proposed methods

Velocity Estimation Method (VEM): The method to estimate the velocity of the MH from the contexts such as RSS. In this chapter, an ANN-based VEM is proposed. The ANN utilizes the RSS and the rate of RSS change, namely ΔRSS , as the inputs to estimate the velocity of the MH. An overview of the ANN method is included in section 6.2.3.

Travel Distance Estimation (TDE): This is a method to estimate the distance that will be traveled by the MH within a wireless network cell. The base concept used in this chapter is extracted from

the work by Yan *et al.* [110]. This base concept is discussed in section 6.2.2.

Warm-up Time: This is a new concept proposed in this chapter. The warm-up time is a duration time needed for estimating an accurate value of TDE.

Unnecessary Handover Avoidance (UHA): From the estimated velocity (via VEM) and the estimated travel distance (via TDE), the MH's dwell time within a wireless network can be predicted. In the UHA, if the processing time to enter and leave a network is larger than the dwell time, a handover to that network is deemed unnecessary. Therefore, the MH should avoid handing over its connection to that network. The accuracy of this method highly relies on the credibility of the MH's estimated velocity and travel distance. This feature is not proposed in this study. However, UHA can benefit from the proposed VEM and warm-up time approaches.

Adaptive RSS Threshold (ART): RSS threshold is a value used to trigger a handover in traditional handover approach. The standard approach uses a static RSS threshold; when the RSS of the current network becomes lower than this threshold, a handover to the target network will be initiated. Even though there exists other methods – for example, handover initiation based on the comparison between the RSS of the current network and the target network, or handover trigger based on other variants of this approach – an RSS threshold is still useful to avoid total connection loss due to situations that does not fall in the handover triggering rule (assuming that there are other existing networks within the MH's vicinity). In this chapter, an adaptive RSS threshold approach is proposed. The term adaptive here is relating to how the threshold is adjusted to the MH's velocity. Because it takes time to complete the handover process, depending on the network delay and the network type, a handover failure will occur if the MH moves out of the current network before completing this process. An RSS configuration that is suitable for low velocity might not be suitable for higher velocity since more distance is traversed within shorter time duration. Thus, adapting the RSS threshold to the MH's velocity can improve the handover process.

Handover Failure Estimation (HFE)/ Handover Failure Avoidance (HFA): Handover failure estimation or handover failure Avoidance is a technique to avoid handover failure due to the MH exiting the current serving network before the handover process completes. One of the existing methods is the approach proposed by Abrar *et al.* [104] that estimates the handover failure probability to trigger the handover process, and avoid handover failures. In this chapter ART, which will be discussed in Sect. 5, is proposed to avoid handover failures.

6.2.2 Travel Distance Estimation Method

The TDE concept used in this paper is based on the work by Yan *et al.* [110], which uses the topology illustrated in figure 6.1 to formulate the model. It is assumed that the cell has a circular shape with the AP at the center of the circle labeled as o . The list of notations based on figure 6.1 is as follows:

AP Access Point.

P_i The point where MH enters the cell.

P_o The point where MH leaves the cell.

s A point within the cell after MH has moved for a certain duration.

m A point at the center of the MH's travel distance (half of P_i-P_o distance).

a The critical point based on the RSS threshold where MH should start the handover process. The position of A on the P_i-P_o line is determined by the value of d_{crit} which is calculated based on the proposed ART.

l_{os} Distance between point o and point s . This value changes with the RSS since its calculation is based on the current RSS.

l_{om} Distance between point o and point m . This value is the perpendicular distance from point o to the P_i-P_o line.

l_{oa} Distance between point o and point a . This value is based on d_{crit} , which is calculated based on the proposed adaptive RSS threshold (ART).

d_{crit} The critical distance, which is the minimum distance between the cell edge and the point where MH should start the handover process.

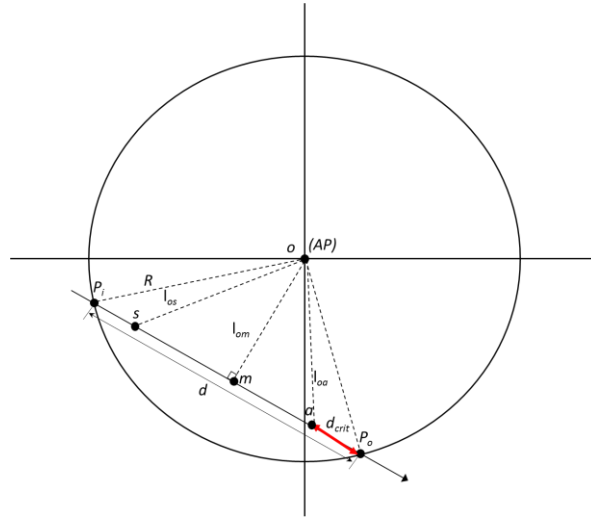


Figure 6.1 Geometrical model for TDE concept based on [110]; assuming that the cell has a circular shape and MH is moving in a straight line with constant speed, the distance P_i-P_o can be estimated.

Referring to figure 6.1, assuming that the MH moves in a straight line with a constant velocity from its entry point at P_i to its exit point at P_o , the travel distance of the MH, d , can be estimated. This distance is useful in determining the necessity of handing over to a target network, and thus the RSS threshold level to avoid handover failure. When the MH enters AP's cell, the radius of the cell, R , can be estimated using a propagation path loss model. In this case, the Free Space Path Loss (FSPL) model is used for simplicity of its calculation. Take note that this method is not limited to just FSPL, other propagation path loss models are also applicable. Thus the value of R can be

obtained from the following equation:

$$R \cong \sqrt{\frac{P_T \lambda^2}{P_{P_i} 16\pi^2}} \quad (6.1)$$

R Radius of AP's cell.

P_T Transmission power of the AP.

P_{P_i} Power received by MH at point P_i .

λ Wavelength (speed of light/carrier frequency)

Then after some interval t , another sample is taken at point s , where the distance between AP and s , l_{os} , is estimated as:

$$l_{os} \cong \sqrt{\frac{P_T \lambda^2}{P_s 16\pi^2}} \quad (6.2)$$

l_{os} Distance between AP and point s .

P_s Power received by MH at point s .

Next, let M be a point at the center of d and perpendicular to O . From the geometric configuration in figure 6.1, applying the Pythagoras Theorem, the values of l_{os} and l_{om} can be calculated as:

$$l_{os}^2 = l_{om}^2 + \left(\frac{d}{2} - vt^2 \right) \quad (6.3)$$

$$l_{om}^2 = R^2 + \left(\frac{d}{2} \right)^2 \quad (6.4)$$

After that, substituting Eq. (4) into Eq. (3), the value of d can be estimated as:

$$d = \frac{R^2 - l_{os}^2 + v^2 t^2}{vt} \quad (6.5)$$

Notice that in Equation (6.3) to Equation (6.5), TDE depends on the MH's velocity information. Thus, it is very important to always have accurate velocity information.

6.2.3 Overview of the ANN technique

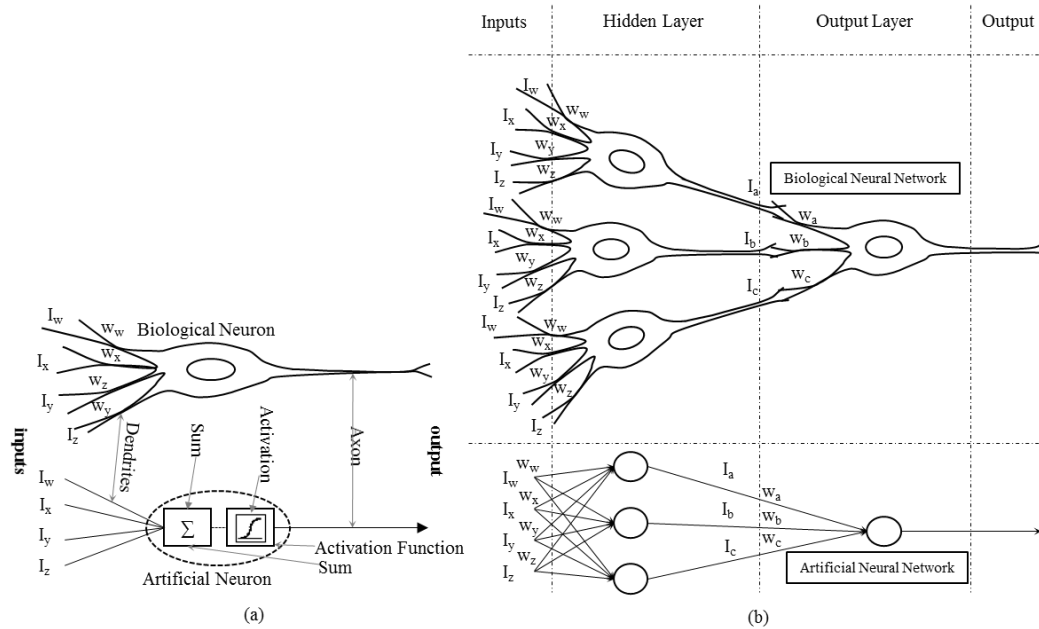


Figure 6.2 The biological neural network adaptation to computational model. (a) The component comparison between biological neural cell and the artificial neural model. Both of them have the same components such as dendrites, neurons and axons. (b) The comparison between a biological neural network and artificial neural network. The neural network consists of the neural cells illustrated in (a) which are connected with each other to create a network of neural cells. The connections considered here are the connections from inputs (similar to biological sensory cells) to the neural cell in the hidden layer and from the axons of the neural cells in the hidden layer to the dendrites of the neural cell in the output layer. The connections in the artificial neural network are weighted by the synaptic weight coefficient to reflect the characteristics similar to its biological counterparts.

ANN (Artificial Neural Network) is a computational model inspired by the biological neural network [114], [115]. As illustrated in figure 6.2(a), the ANN model literally imitates its biological counterpart in terms of its structure and its functions. The four main components in the neural network are: (1) dendrites, the entries which receives and passes the information to the neuron; (2) neuron, the main cell body which processes the information; (3) axon, the branch that sends the processed signals from the neuron to another neuron via the axon-dendrite connection, also known as synapse (as illustrated in figure 6.2 (b)); (4) synapse, specialized connection between cells where information are passed.

The artificial neural network is modeled based on the following four components: (1) the inputs which imitate the dendrites where the input signals enters into the neuron; (2) the artificial neuron, the core computation part in a neural network which sums up the signals collected. All the signals that enter the neuron are summed via the summing function and the amplitude of the output is controlled by the activation function. For example, the range of output may be between 0 and 1,

or it could be between -1 and 1, depending on the selected activation function; (3) the output that imitates the biological axon, where it could be connected to another neuron, via dendrites, to create a neural network, or it could be used as the output decision of the model; (4) the synaptic weights, a coefficient to determine the impact of each input, which is modeled after the biological synapses. A negative value reflects an inhibitory connection whilst a positive value replicates excitatory connection.

There are several types of existing ANN designs [115]. In this work, a multilayer feed-forward back-propagation method is employed. A feed-forward neural network is an ANN where data flow is strictly fed forward without any feedback to the cells in the current layer or the layers before. The term “back-propagation” means that the output of the network is compared to a desired output for the set of input values. This network is illustrated by the artificial neural network shown in figure 6.2 (b). The proposed ANN model will be discussed in detail in the following section.

6.3. Velocity Estimation Method (VEM)

The work in [97] is one of the most recent studies on the velocity estimation method. In their work, they assumed that the velocity of the MH can be divided into two parts which are the radial velocity, v_r [m/sec] and radian velocity, v_θ [theta/sec]. The v_r is only affected by the change in the distance of the MH from the AP, namely d_{MH-AP} . They stated that the distance of the MH and AP is only affected by the v_r . Hence, they opted to use only v_r as well as the RSS as the handover trigger in their proposed technique. The current distance, $d_{MH-AP}(t)$, and the previous distance, $d_{MH-AP}(t-1)$, are obtained via path loss model, then using the difference between these two distances per time, the v_r can be estimated. Their proposed method achieved the velocity (radial velocity, v_r , in particular) estimation with high accuracy up till 100 km/h. However, the estimation error increases drastically, as the velocity becomes higher. For example, from their results, when the velocity is 350 km/h (approximately 97 m/s), the average error of the estimated velocity compared to the actual velocity is as high as approximately 24 km/h (6.67 m/s), which is a quite significant error. Thus, to improve the performance of the velocity estimation approach, a new ANN based VEM is developed in this study.

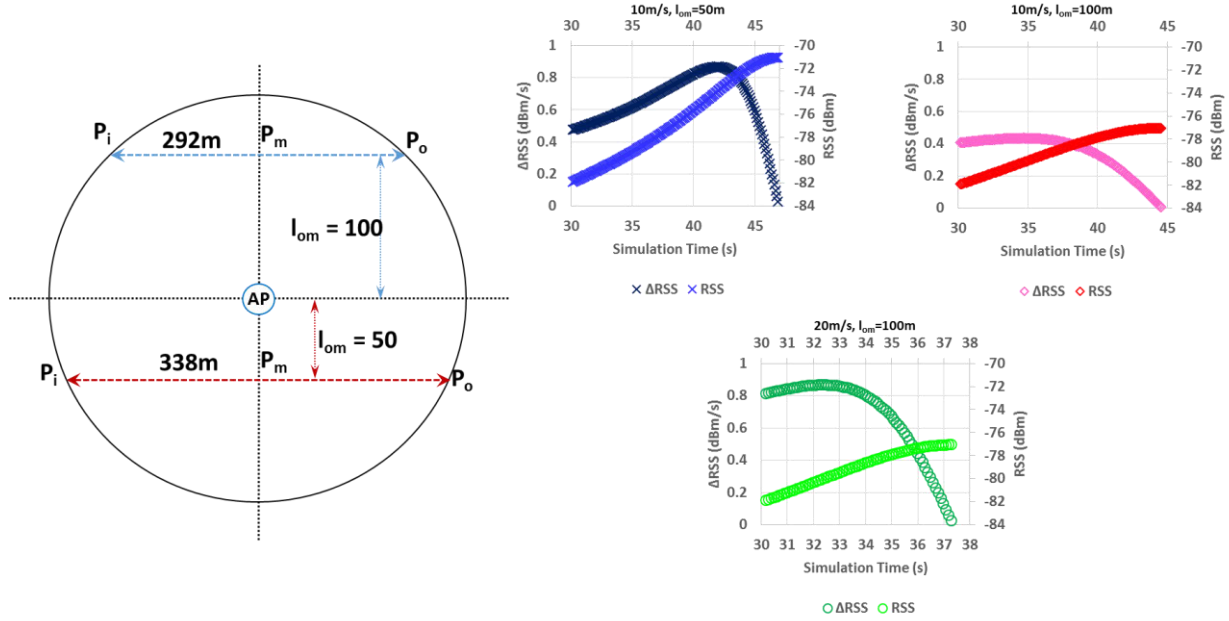


Figure 6.3 The effects of different speed and l_{om} on the RSS and Δ RSS values obtained by MH. In the network topology, two different paths with different l_{om} are shown where the blue path (the path with blue arrows) have a $l_{om} = 100$ meters distance whilst the red path (the path with red arrows) have a l_{om} value of 50 meters. Due to the different l_{om} values, the red path has a longer distance compared to the blue path, and the maximum value of RSS achievable is higher as shown by the blue plots in the graph, since the MH will be closer to the AP. The graph shows the different RSS and Δ RSS patterns for different l_{om} and MH velocity configurations.

In this section, the proposed ANN-based VEM and its effectiveness will be discussed. This section starts with the concepts used for the proposed VEM; since the target of this method is to estimate the MH velocity with high accuracy, some meaningful context information related to the velocity is needed. Here, the RSS and the rate of RSS change, namely Δ RSS is utilized to estimate the velocity. Section 6.3.1 discusses the ideas on how RSS and Δ RSS can be used to estimate the velocity. In section 6.3.2, the proposed ANN circuit will be discussed followed by the performance evaluation in section 6.3.3.

6.3.1 Utilization of RSS and Δ RSS in estimating MH's velocity

The proposed ANN circuit takes two inputs; the value of RSS and the rate of RSS change, namely Δ RSS. The Δ RSS here is the temporal differentiation of the RSS; specifically the value of Δ RSS is obtained by subtracting the previous RSS from the current RSS and dividing it by the time interval between the current RSS reading and the previous RSS reading (Δ RSS (t) = (RSS (t) - RSS (t₁)) / (t - (t₁))). To clarify the significance of using RSS and Δ RSS as the input for the proposed ANN, consider the network topology and graph shown in figure 6.3. Let AP be the access point with a circular coverage area, the blue arrows be the MH's travel path with a l_{om} of 100 meters, namely the blue path and the red arrows be the MH's travel path with a l_{om} of 50 meters, namely

the red path. The blue path has a theoretical distance of 292 meters and the red path has a theoretical distance of 338 meters. The MH enters the network from P_i and exits the network at P_o whilst P_m is the center of the P_i - P_o line.

The graphs in figure 6.3 shows three different patterns of RSS- Δ RSS pairs collected from P_i to P_o with the assumption that the RSS and Δ RSS patterns for P_i - P_m and for P_m - P_o are symmetric, and denoted by three different markers. The pair represented by the 'x' marker (blue for RSS and dark blue for Δ RSS) is obtained when $l_{om} = 50$ meters and MH speed = 10 m/s. Next, the pair symbolized by the ' \diamond ' marker (red for RSS and pink for Δ RSS) is achieved when $l_{om} = 100$ meters and MH speed = 10 m/s. Finally, the pair signified by the 'o' marker (light green for RSS and dark green for Δ RSS) is attained when $l_{om} = 100$ meters and MH speed = 20 m/s.

There are two main points that should be noted from the graphs in figure 6.3. **The first point** is the pattern of the RSS. Two MHs have similar RSS range and pattern when they have the same l_{om} albeit the temporal difference; in contrast, two MHs have different RSS range and pattern when they have different l_{om} . This is shown by the red ' \diamond ' plots and light green 'o' plots, where they both have the same range of RSS which is approximately -77dBm to -82dBm. Hence, RSS information alone is not sufficient to estimate the velocity. **The second point** is the Δ RSS pattern. Two MHs with different velocity have different Δ RSS pattern and range when they have the same l_{om} ; however, two MHs with different velocity may have the same pattern when the l_{om} is different. For example, observe the dark blue 'x' plots and the dark green 'o' plots. Part of the dark blue plots has similarity with the dark green plots. If only Δ RSS is considered for the velocity estimation, this point might pose some difficulties, since two different velocity has similar Δ RSS pattern.

These two points are complementary. RSS pattern is the same within the same l_{om} and differs with different values of l_{om} ; the Δ RSS pattern is different with different velocities, but may be the same for different l_{om} values. Hence, combining these two information, unique pairs for different l_{om} and different MH velocity throughout the wireless network coverage boundaries can be obtained and can be used to train the ANN-based VEM.

6.3.2 The proposed ANN model for velocity estimation improvement

It has been established in the previous section that the RSS- Δ RSS pair are useful information to estimate the velocity of the MH. Thus these information is collected and used to train the proposed ANN VEM. Figure 6.4 shows the sample data used to train the proposed ANN. The graph with '+' marker depicts the RSS patterns for MH velocities from 10m/s to 100m/s. Meanwhile, the graph with 'x' marker depicts the Δ RSS patterns for MH speeds from 10m/s to 100m/s.

With the same assumption as in the previous section, the RSS- Δ RSS pairs are obtained from P_i to P_m as in figure 6.3 and they are also assumed to be symmetric with the pattern from P_m to P_o . Take note that the value of RSS is converted to positive value; that means, smaller number will have higher value; for example, 70dBm is larger than 80dBm. As can be seen from figure 6.4, even though the RSS pattern is nearly similar albeit the temporal difference, the Δ RSS pattern is distinctly different. This is the data collected from a network topology similar to figure 6.3, with the MH's travel path at $l_{om} = 100$ m (see the blue path in figure 6.3). Using this data as input, the

proposed ANN model is trained using MATLAB, a software for numerical calculation, visualization and programming. A Feed-forward Back-propagation [112] ANN model is developed with the following configurations (see figure 6.5):

- **The inputs used are:** a) $x_1 = \text{RSS}$; b) $x_2 = \Delta\text{RSS}$
- **The configuration of the ANN circuit in MATLAB nntool:**
 - Network Type: Feed-forward back-propagation
 - Training Function: TRAINLM
 - Adaptation learning function: LEARN_GDM
 - Performance Function: MSE
 - Number of layers: 2
 - Layer 1 (hidden layer): number of neurons - 20; Transfer Function – LOGSIG(log-sigmoid)
 - Layer 2 (output layer): Transfer Function – PURELIN(linear)

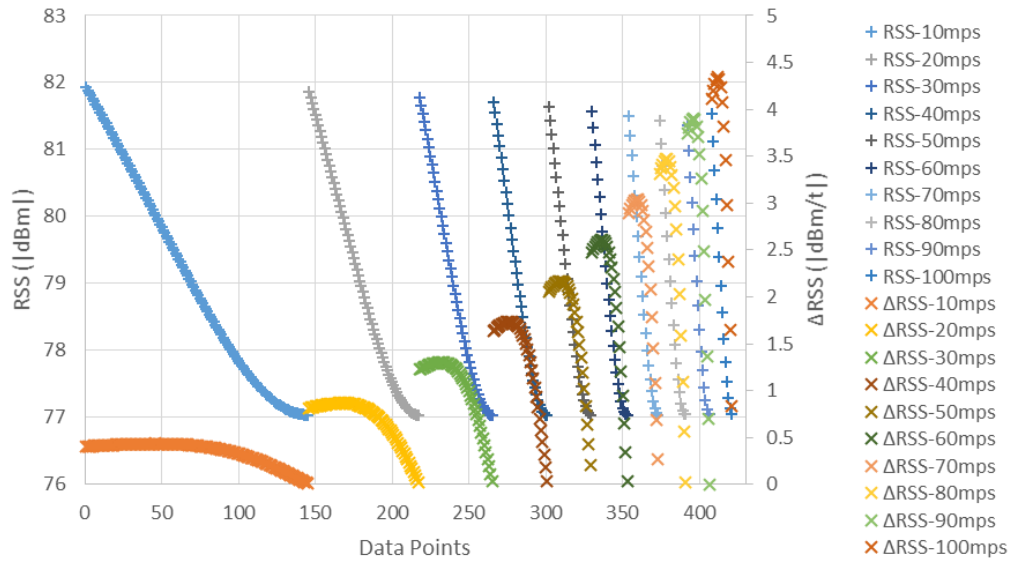


Figure 6.4 The RSS- ΔRSS pair used to train the ANN circuit for the ANN-based VEM. The '+' plots signifies the RSS whilst the 'x' plots signifies the ΔRSS . The RSS- ΔRSS are collected for MH velocity from 10m/s to 100m/s.

Figure 6.5 shows the structure of the proposed ANN model. This circuit has two inputs, x_1 (RSS) and x_2 (ΔRSS). For all variables in figure 6.5, m denotes the number of input, where $m = (1, 2)$; n denotes the number of nodes in the hidden terminal, where $n = (1 \dots 20 \text{ with a step of } 1)$; i indicates that the parameter is from the input layer and processed by the hidden layer; o indicates that the parameter is processed by the output layer nodes. Each input is given a weighting coefficient, $W_{m,n}^i$. The weighted inputs are then summed up and added with a bias value b_n^i to obtain the s_n^i , which

in turn will be normalized with a log-sigmoid function f^i to obtain t_n^i . The values of t_n^i are then weighted with w_n^o , summed and added with a bias value b^o to obtain s^o . Then s^o is processed by the linear function f^o to finally obtain the output, t^o , in this case, the estimated velocity. This ANN circuit is trained using the RSS and Δ RSS data. In this study, the samples for 10m/s to 100m/s with a step of 10m/s as shown in figure 6.4 are used to train the ANN model. The sample data were obtained from simulations using OMNeT++ with the INET module.

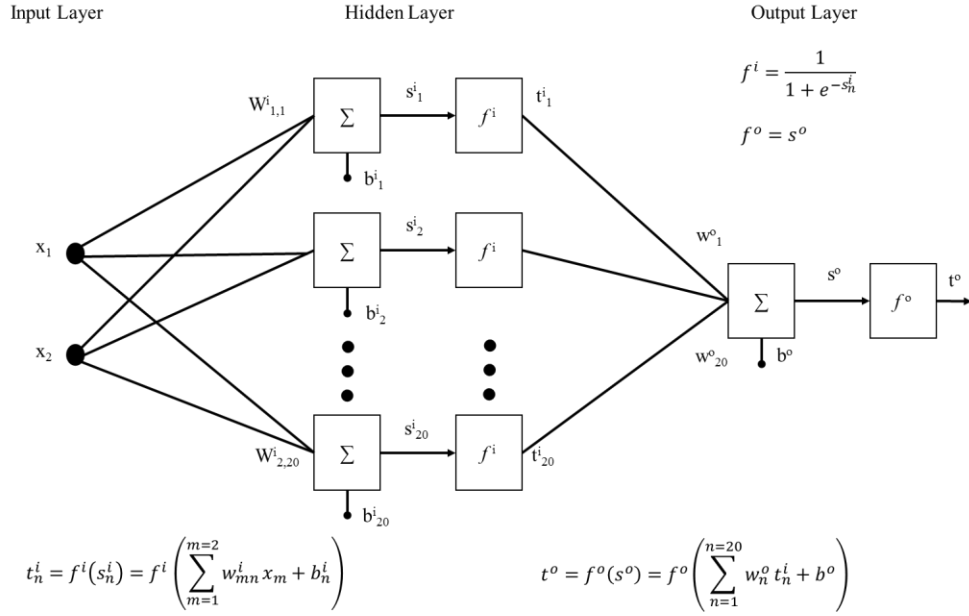


Figure 6.5 The ANN configuration used in MATLAB to obtain the velocity estimation. This model uses the Feed-forward Back-propagation method. In the hidden layer, the activation function used is the log-sigmoid function, whilst in the output layer, pure linear function is used as the activation function.

6.3.3 Evaluation of the proposed ANN-based VEM

To evaluate the effectiveness of the proposed ANN-based VEM, the error of the estimated velocity is measured. Figure 6.6 shows the MH's velocity average estimation error with 95% confidence interval. From the graph, it can be seen that the average estimation error is 0.3m/s or below with a maximum error of less than 0.5m/s. Figure 6.7 compares the performance of the proposed ANN VEM with the existing method done in [97]. From the graph, it can be seen that the existing method has smaller error for the velocity less than 20m/s.

However, as the velocity increases, the average estimation error of the existing method increases whilst the proposed method maintains its average error below 0.5m/s. Thus, from figure 6.7, it is clear that the proposed method has improved velocity estimation capabilities compared to the existing method. Meanwhile, figure 6.8 shows an example of the estimated velocity compared to the corresponding real velocity. It can be seen that the estimated velocity using the proposed ANN VEM is very close to the real velocity value. The proposed VEM with high

accuracy is useful to get accurate results for the warm-up time and the adaptive RSS techniques that will be discussed in section 6.4 and section 6.5 respectively.

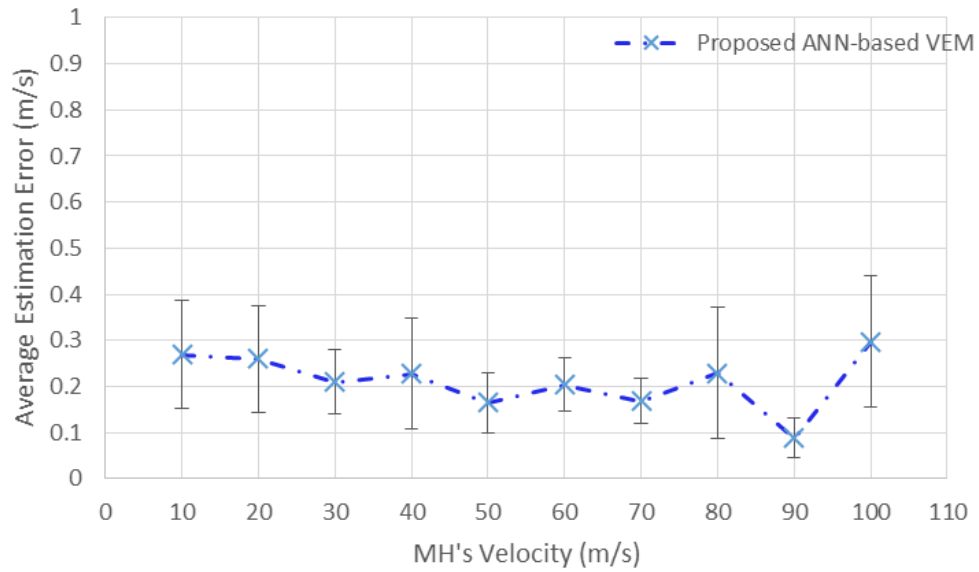


Figure 6.6 The estimation error of the proposed ANN-based VEM. The estimation error is defined as how far the estimated value is from the real value. From the graph, it can be seen that the proposed approach has an estimation error of 0.3 m/s or less, with a maximum of error less than 0.5 m/s. This graph is focused on the proposed method to show the 95% confidence interval.

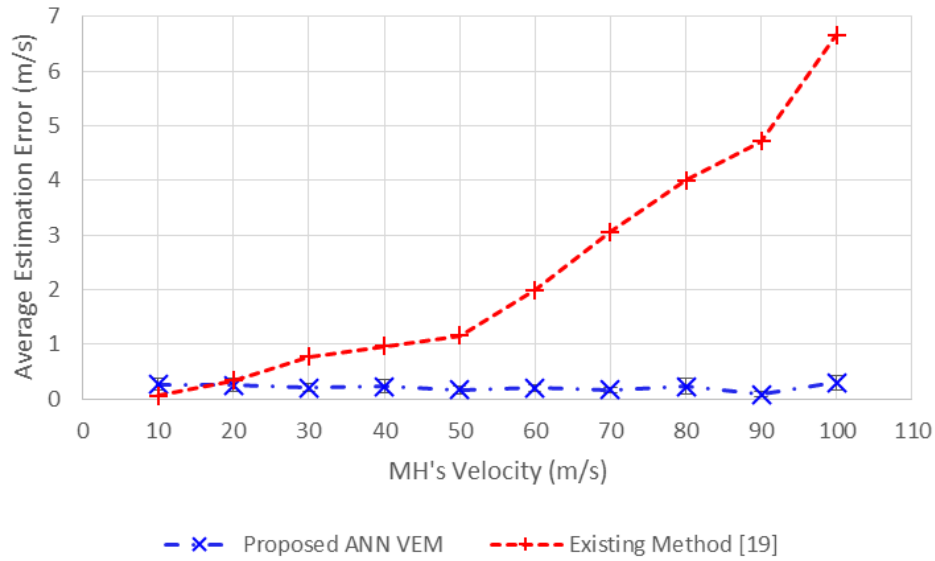


Figure 6.7 Average estimation error comparison between the proposed ANN-based VEM and the existing VEM method by [19]. From the graph, even though the method from [97] has lower error when the MH velocity is less than 20m/s, it is obvious that the proposed method can maintain low estimation error for the velocity more than 20 m/s.

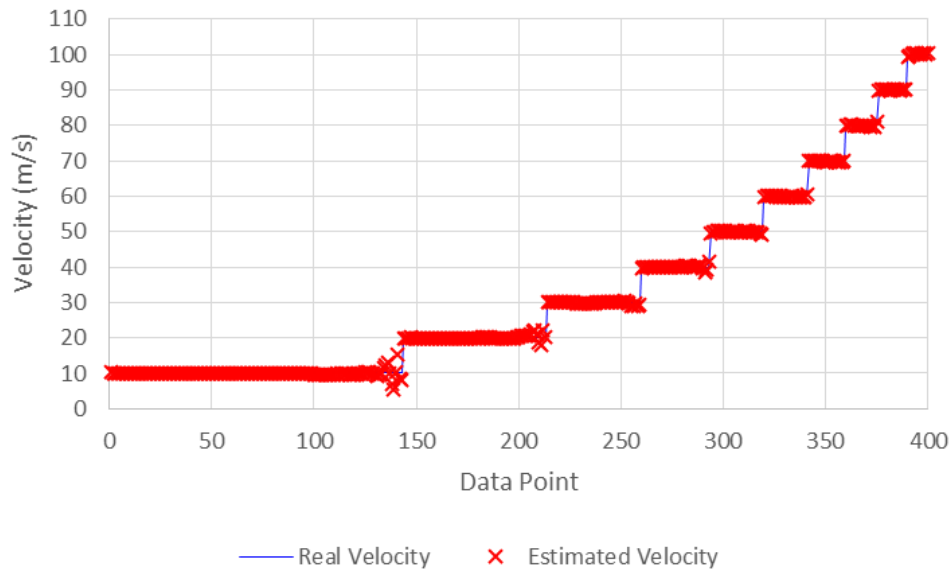


Figure 6.8 A sample of the estimated velocity results using the proposed ANN-based VEM compared to the real velocity of the MH. From this graph, it can be seen that the estimated value of the proposed ANN-based VEM is capable of estimating the velocity of MH accurately.

6.4. Proposed warm-up time approach

Using the estimated velocity of the MH, the MH's travel distance within a network cell can be estimated using the Travel Distance Estimation method discussed in section 6.2. In this section, a new concept called the warm-up time is introduced to increase the accuracy of the TDE. The travel distance is a useful information that can be used to avoid unnecessary handover and to adapt the RSS threshold to trigger a handover, if a handover is deemed necessary.

6.4.1 Obtaining the warm-up time

As discussed in the introduction, from the study in [110] the estimation error is less than 10% for the MH's velocity 15m/s to 40m/s. However, the error increases more than 10% for the velocity less than 10m/s. To clarify the reason for this issue, the same method proposed in [110] (the travel distance estimation using equation (6.5)) is used. A topology as shown in figure 6.9 is considered, where the MH is configured to move according to the blue dotted lines. The l_{om} is varied from 10m to 100m as shown in figure 6.9 by the topology on the left. This is to cater for the possibilities of different point of entry, P_i , as shown by the rotated topology on the right in figure 6.9. Then, the travel distance of MH within the network is continuously estimated from the point where the MH enters the network, point P_i , to the point where the MH exits the network, point P_o . From this analysis, it was found that the estimation error is very high during the initial estimation of the travel distance, and the error drops sharply and arrives at the actual distance after a certain interval depending on the MH's velocity

Figure 6.10 shows a sample of the MH's estimated travel distance when the $l_{om} = 10m$ and the MH moves at 10 m/s. Theoretically, the travel distance should be 353.435 meters. Notice that during the early part of the estimation, the estimated distance has a large error. Then after some time (approximately 35 seconds) the value settles down to the values close to the theoretical value. This analysis is the basis of the proposed warm-up time technique. In the previous method [110], only 20 RSS samples collected within the early 100ms interval, which is insufficient to get an accurate estimation of the travel distance, was used to estimate the travel distance. This is the reason for the high estimation error for velocity less than 10 m/s. From figure 6.10, it is obvious that the early estimation values have large errors, whereas after a certain interval, the estimated distance becomes more accurate. Thus, it is concluded that to improve the accuracy of the estimated distance, the early estimation values should be skipped for a suitable duration to avoid or reduce the errors. This duration is called the warm-up time.

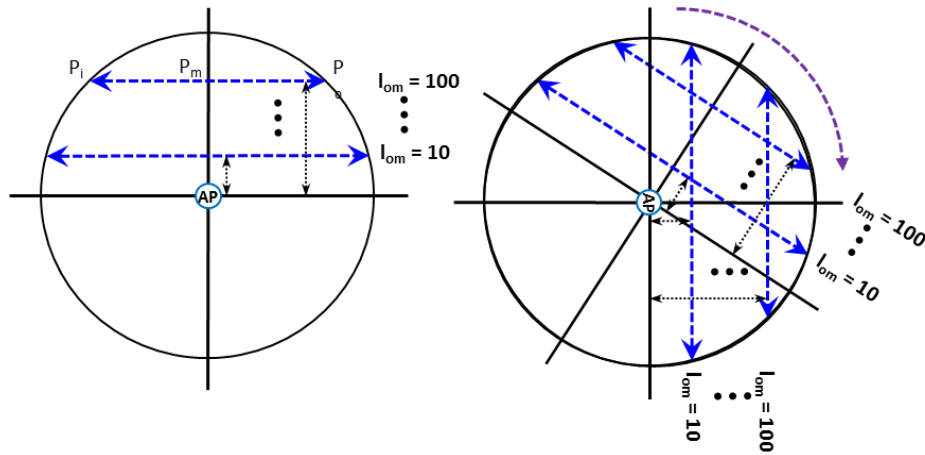


Figure 6.9 The network topology considered for the proposed warm-up time approach. The topology on the left shows the travel path of the MH, which is varied from 10 m to 100 m with 10 m steps. The topology on the right shows the changed MH path when the topology is rotated. From this topology, it can be seen that, when considering that the MH moves in a straight line within the wireless coverage area, travel paths similar to the topology on the left can be seen, even though the point of entrance is different. Thus it is assumed that considering the topology on the left is sufficient to estimate the outcomes for different point of entries.

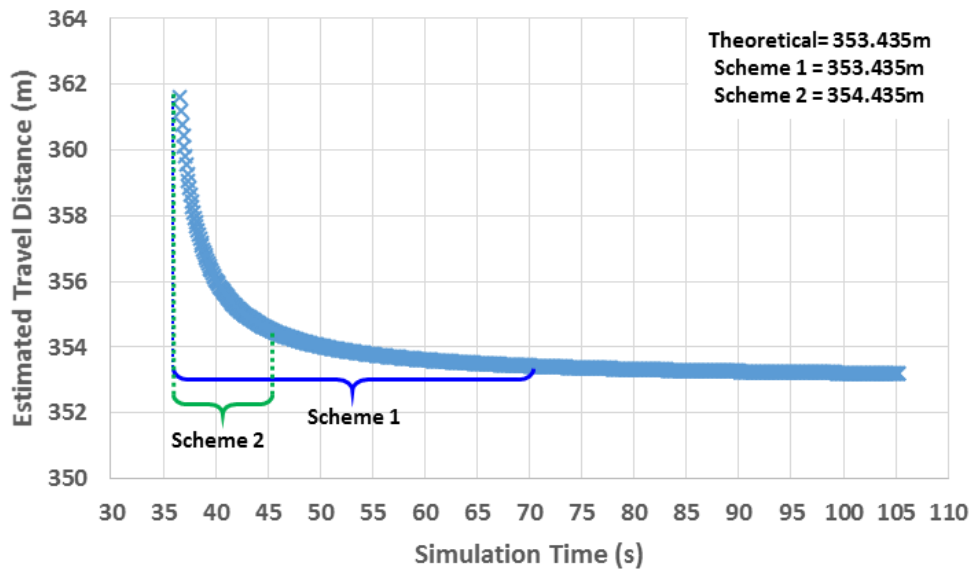


Figure 6.10 Sample data of the MH's estimated travel distance. The blue bracket illustrates the first scheme whilst the green bracket shows the second scheme. The theoretical distance is 353.435 m. The first scheme estimates the distance as 353.435 m whilst the second scheme estimates the distance as 354.435 since it allows a 1m error to reduce the required estimation time.

In this chapter, two schemes have been used to investigate the tendencies of the warm-up time for different l_{om} and different MH velocities. In the first scheme, the duration is measured from the first estimation value to the value the same or closest to the theoretical value as shown by the blue bracket in figure 6.10. On the other hand, the second scheme tolerates a 1m error as shown by the green bracket in figure 6.10; this means that the duration is measured from the first estimation value, similar to the first scheme, to a value of 1m away from the theoretical value. Figure 6.11 and figure 6.12 show the required warm-up time for the first scheme and the second scheme respectively. The horizontal axis is the MH's velocity, whilst the vertical axis is the required warm-up time. These graphs show the value of the required warm-up time corresponding to the velocity of the MH. As stated previously, l_{om} is varied from 10m to 100m and the average warm-up time with 95% confidence interval is also included. This average plot is used to derive the fitting function, which can be used to adapt the warm-up time configuration of the MH to the current MH velocity.

Using the first scheme, high accuracy can be achieved since this method skips to the estimation value closest or the same as the theoretical value. In figure 6.11, it can be seen that there are deviations in the value of the required warm-up time due to the different l_{om} configuration. Nevertheless, from the average plot with its 95% confidence interval, it can be seen that the deviations are very small. Thus, the fitting equation in Figure 6.11 ($y = 238.91x - 1.195$) can be used to accurately adapt the warm-up time configuration according to the MH's velocity. However, for the MH velocity less than 40 m/s, the required warm-up time is considerably long, which reduces the effectiveness and usability of this approach with networks with small coverage area. For example, in Figure 6.11, the required warm-up time when the MH moves at a velocity of 5 m/s is approximately more than 31 s, which is very large. From this example, the distance covered during this warm-up interval is already more than 150m which is already longer than the maximum distance within a normal WLAN with a 50m radius. This might force the MH to avoid handing over to such networks, even though the MH is moving at a slow pace (5m/s).

The second scheme, where a 1m error tolerance is considered, is introduced to reduce the required warm-up time for the velocity less than 40m/s. Figure 6.12 shows the required warm-up time for the second scheme. Using this scheme, the deviations are nearly similar to the first scheme in Figure 6.11. The fitting equation of the average required warm-up time shown in Figure 6.12 can be used to adapt the warm-up time configuration of the MH according to the second scheme. However, it is obvious that the required warm-up time is significantly reduced compared to the first scheme. For example, using the second scheme, the average warm-up time for 5m/s is approximately 4s, compared to 30s in the first scheme. However, from the 95% interval of the average plot, it can be seen that the deviations for velocity higher than 40m/s is slightly larger than the first scheme. Nevertheless, the average required warm-up time for the first scheme and the second scheme is very similar, despite the slight difference in the deviations.

Thus, by combining these two schemes the required warm-up time for MH moving at the velocity less than 40m/s can be reduced from approximately 31s to just 4s by incorporating the second scheme. Furthermore, the required warm-up time deviations for the MH velocity more than

40m/s can be maintained by using the first scheme. Figure 6.13 shows the required warm-up time for the first scheme (blue plots), the second scheme (green plots), and the combination of the first and the second schemes (gray plots). From the gray plots, a fitting function is derived as $y = 0.0002x^2 - 0.0608x + 4.7119$. This fitting function can be used by the MH to adapt the warm-up time configuration used by the MH according to the velocity of the MH.

In a nutshell, if the accuracy is the priority and the considered network has large coverage areas, with sufficient overlapping between the networks, then the first scheme is the most suitable. On the other hand, if a network with small coverage area is to be considered, the combination approach is more suitable since the required warm-up time for the velocity less than 40m/s is shorter and the warm-up time estimation is more stable (less deviation) for the velocity more than 40m/s. In the next sub-section, the effectiveness of the proposed warm-up time approach will be discussed.

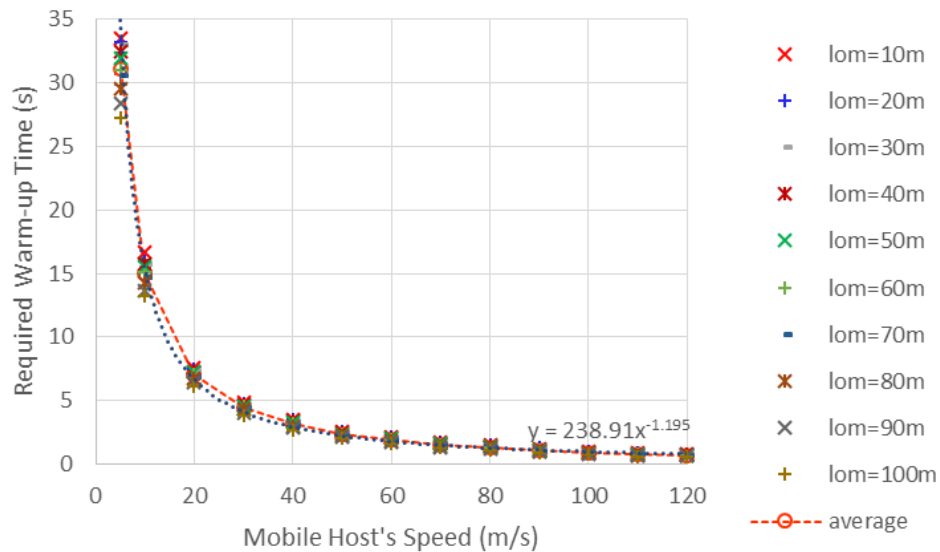


Figure 6.11 The required warm-up time for the first scheme. This graph shows the required warm-up time for MH with different velocity and different l_{om} values. It can be seen that when the MH speed is low, the deviation between required warm-up times is different for different l_{om} . However, as the velocity of the MH increases, the required warm-up time for different l_{om} values becomes similar. The average value of the required warm-up time is also shown and the fitting equation can be used to estimate the required warm-up time according to the MH's velocity.

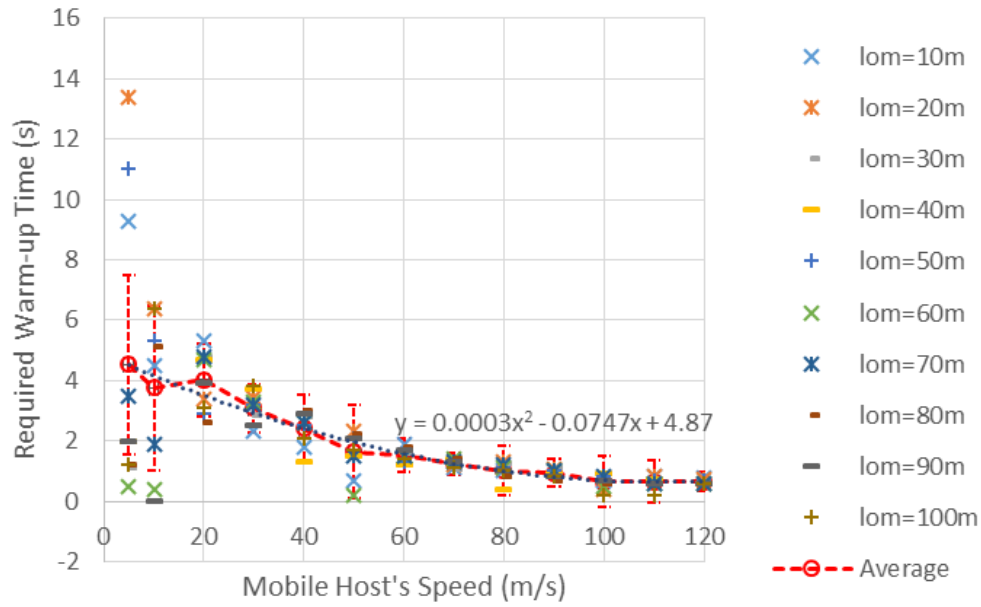


Figure 6.12 The required warm-up time for the second scheme. The l_{om} is varied from 10m to 100m with 10m steps. From this graph, it can be seen that the required warm-up time for velocity less than 10 m/s is significantly shorter than the first scheme. As for the velocity more than 10m/s, the deviations are slightly higher compared to the first scheme. The average value of the required warm-up time is depicted by the red 'o' plots with 95% confidence interval. The fitting equation shown on the graph is derived from the average value plots. This equation can be used to estimate the required warm-up time for each velocity.

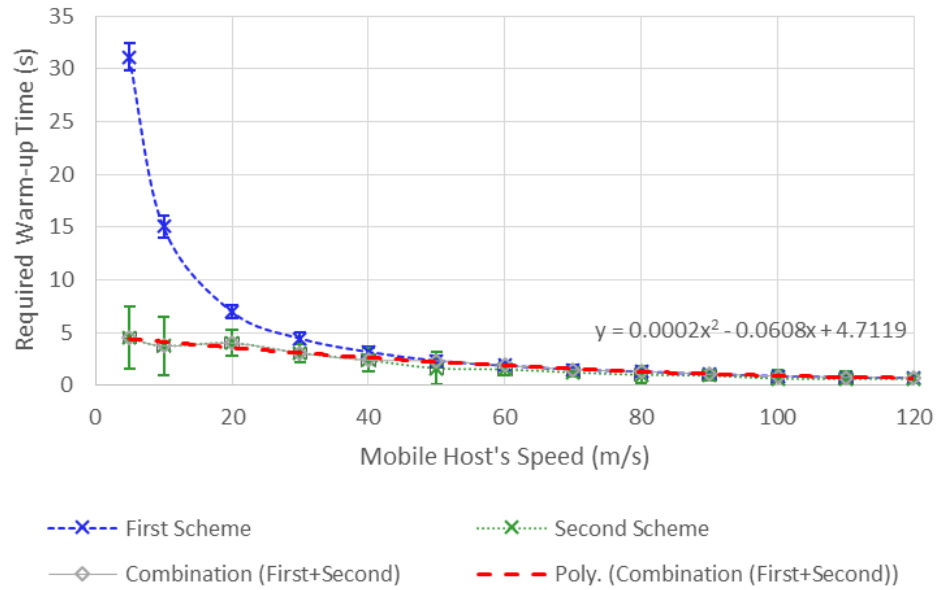


Figure 6.13 The comparison between the first scheme and second scheme. It is clear that for the velocity less than 40m/s, the required warm-up time is very high (more than 30 seconds). On the other hand, the second scheme has significantly reduced warm-up time compared to the first scheme, but with larger deviation shown by the error bar. To maintain a low required warm-up time throughout the whole spectrum of velocity, the first scheme and the second scheme can be synergized to construct a combined approach. The gray '◇' plot shows the combination of both first and second scheme. The equation shown on the graph is the fitting equation for the gray plot (combination scheme). This equation can be used to estimate the required warm-up time for each velocity.

6.4.2 The effectiveness of using the proposed warm-up time approach

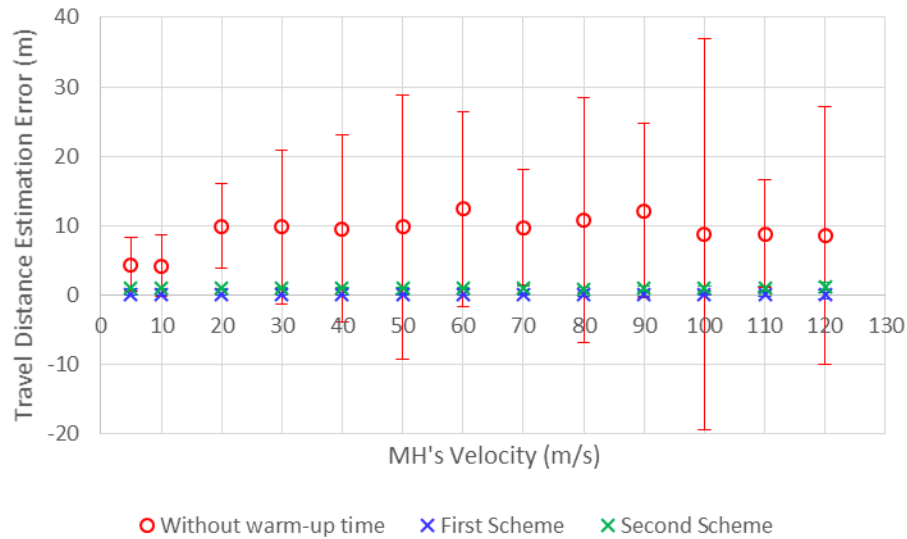


Fig 6.14 Travel distance estimation error performance comparison among the first scheme, second scheme and without using warm-up time.

To measure the effectiveness of the proposed approach, the accuracy of the TDE with the proposed warm-up time and without the warm-up time will be discussed. Figure 6.14 shows the travel distance estimation error for the corresponding MH's velocity with 95% confidence interval for the first scheme, the second scheme and without warm-up time. The vertical axis is the travel distance estimation error in meters; this value is obtained by subtracting the estimated value from the theoretical travel distance value. From Figure 6.14, it is apparent that without implementing the warm-up time, the estimation error is larger and the deviations from the average value is also large. In contrast, the proposed method has significantly better accuracy compared to without using warm-up time.

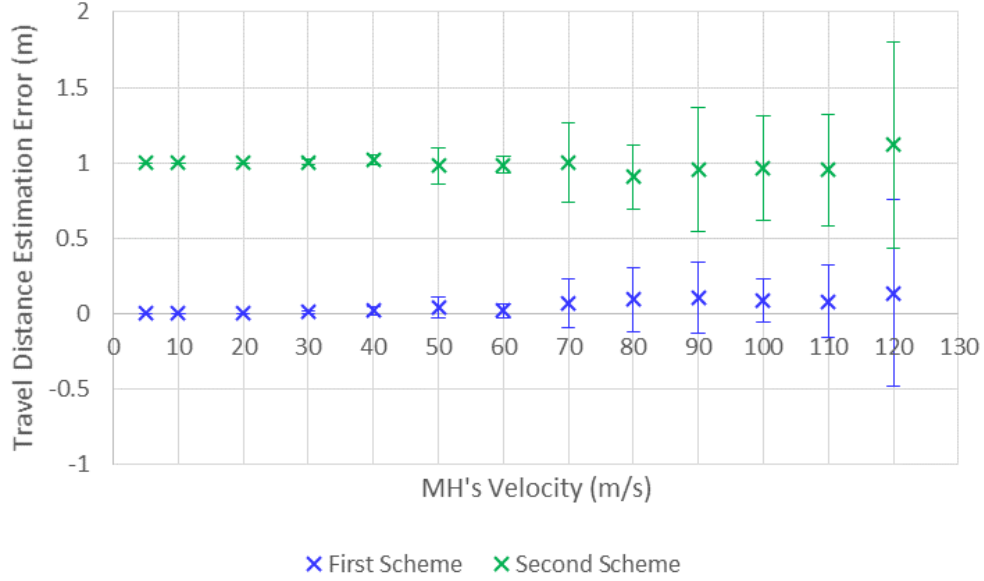


Figure 6.15 Direct comparison between the first scheme and the second scheme.

Figure 15 shows a closer comparison between the first scheme and the second scheme. As expected, the second scheme has an average error of 1m since a tolerance of 1m error is given in the second scheme. Furthermore, the deviation of the first scheme is slightly smaller compared to the deviations of the second scheme for MH's velocity of more than 40m/s. Thus, it can be concluded that, by utilizing the proposed warm-up time approaches, the accuracy of the TDE can be significantly improved.

Finally Figure 6.16 illustrates the estimated travel distance with respect to the varying l_{om} . The horizontal axis is the l_{om} whilst the vertical axis is the estimated travel distance. The graph plots the values of the estimated distances obtained using the first and second schemes, without using the proposed warm-up time scheme, and the theoretical travel distance values. Noticeably, the estimated value, when the proposed warm-up time is not implemented, is far from the theoretical value (at an average of 10m error). Moreover, from the error bar, it is clear there are more variability in the samples obtained without the proposed method. In comparison, the proposed schemes is able to give a stable and accurate estimation of the travel distance. In this section, it can be concluded that the proposed warm-up time approach can deliver stable and accurate estimation of the MH's travel distance within a wireless network. This is useful to ensure the accuracy of the adaptive RSS threshold that will be discussed in the following section.

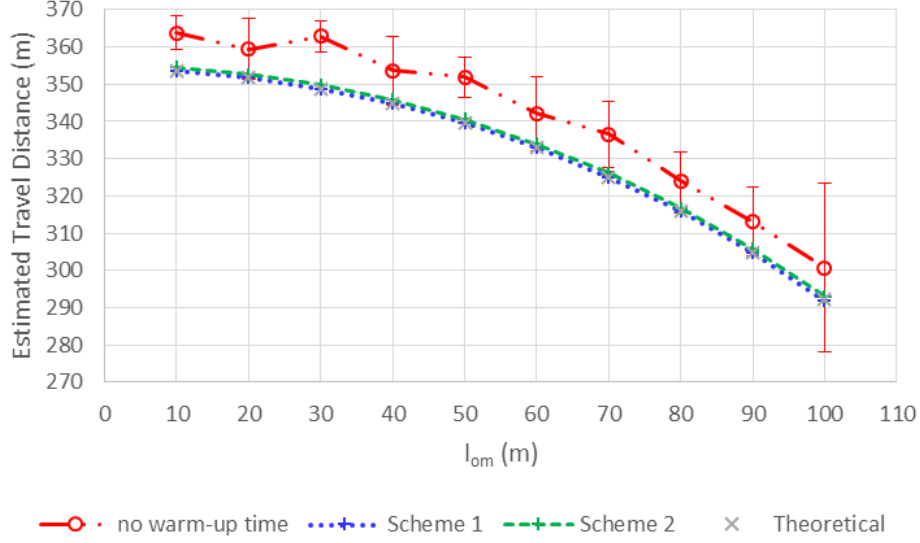


Figure 6.16 The estimated travel distance for different l_{om} configurations.

6.5. Adaptive RSS Threshold

In current existing mobile network, the RSS is the most common context used to trigger a handover. The RSS information is either directly used as the trigger or used with other contexts to implement more sophisticated methods. Sometimes, the RSS can also be used as a failsafe to avoid losing connection to the network. Hence, it is obvious that the RSS is a useful information to maintain the MH's connection to the network. One of the most typical failsafe approach is using a RSS threshold; for example, when the RSS of the current network falls below such threshold, the handover will be initiated. Furthermore, in most of the existing methods, only static threshold is used to determine the point where the MH should start the handover process [110]. However, when MH moves at a higher velocity, a threshold suitable for lower velocity (for example -80dBm) will not be suitable and may cause handover failure due to the latency in initiating the handover process. One way to avoid this is to use a high value of RSS (for example -70dBm) as the threshold value to satisfy all possible velocity. However this will reduce the effectiveness of the handover process, because the MH has to handover very early, even though the current network can still provide good quality of service to the MH (in the case of low velocity). Hence, adapting the RSS threshold to the velocity of the MH is desirable.

Consider a topology as shown in figure 6.3, when the travel distance of the MH within the network can be estimated, the distance between P_o and P_m can be estimated via the Pythagoras theorem. Then the critical distance (the distance from P_o , when MH should initiate a handover) can be calculated as:

$$d_{crit} = v(rtt) \quad (6.6)$$

This critical distance is the distance that will be covered while the MH is completing the handover process. The rtt, which is the round trip time is considered, since it is assumed that at least one round trip time is needed for the MH to successfully complete the handover process. This is due to the endpoint-centric handover approach considered. In an endpoint-centric handover approach, the handover message is communicated directly to the corresponding node to notify the MH's migration to a different network. Let point a in Figure 6.1 be the critical point where the MH has to start the handover process. The distance from the AP to point a , l_{oa} can be estimated as:

$$l_{oa} = \sqrt{l_{om}^2 + \left(\frac{d}{2} - d_{crit}\right)^2} \quad (6.7)$$

Finally, the critical RSS threshold, assuming the free space path loss, RSS_{crit} can be calculated as:

$$RSS_{crit} = \frac{P_T \lambda^2}{16\pi^2 l_{oa}^2} \quad (6.8)$$

Using this approach, the RSS threshold can be adapted to the velocity of the MH. For example, considering the same topology as Figure 6.3, let l_{om} be 20 meters, the radius of the wireless cell be 177 meters, the power transmitted by the AP be 2mW and the end to end delay be 450ms, the adaptive RSS threshold values is as shown in Figure 6.17. The blue line denotes the estimated critical RSS whilst the green dotted line is the fitted line and the equation shown on the graph is the fitted equation of the adaptive RSS threshold. As can be seen in Figure 6.17, when the velocity increases, the RSS threshold becomes higher. This means that when the MH moves at a higher velocity, it will start the handover process earlier, compared to when it is moving at lower velocity. This graph proves that the MH will be able to start the handover process earlier, in order to ensure a successful handover process.

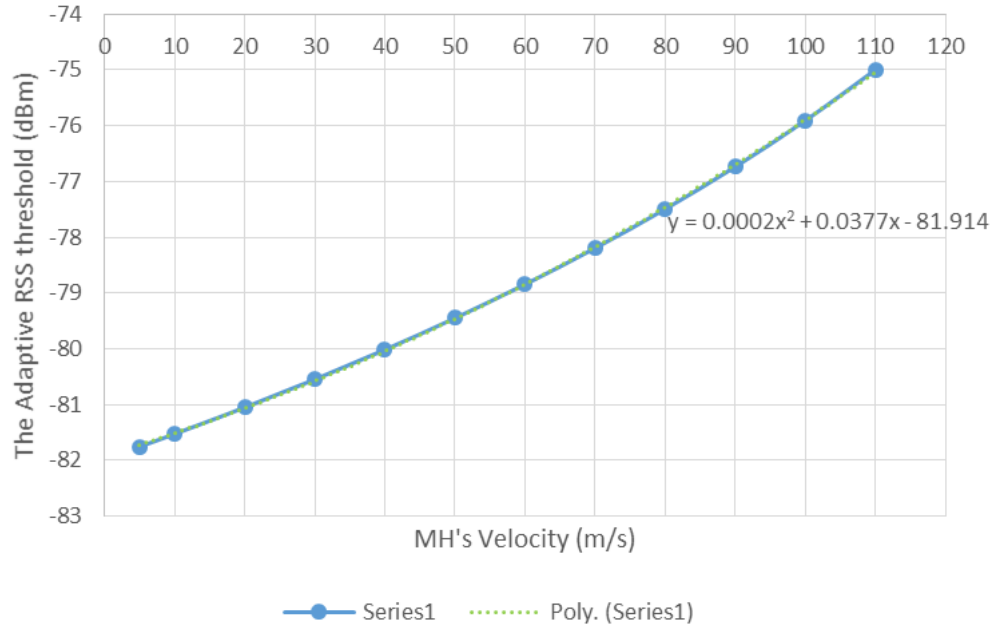


Figure 6.17 An example of the adaptive RSS threshold calculated using the proposed scheme. The configuration for this example is $R = 177\text{m}$, $l_{om} = 20\text{m}$, transmission power by the AP = 2mW and the end to end delay = 450ms .

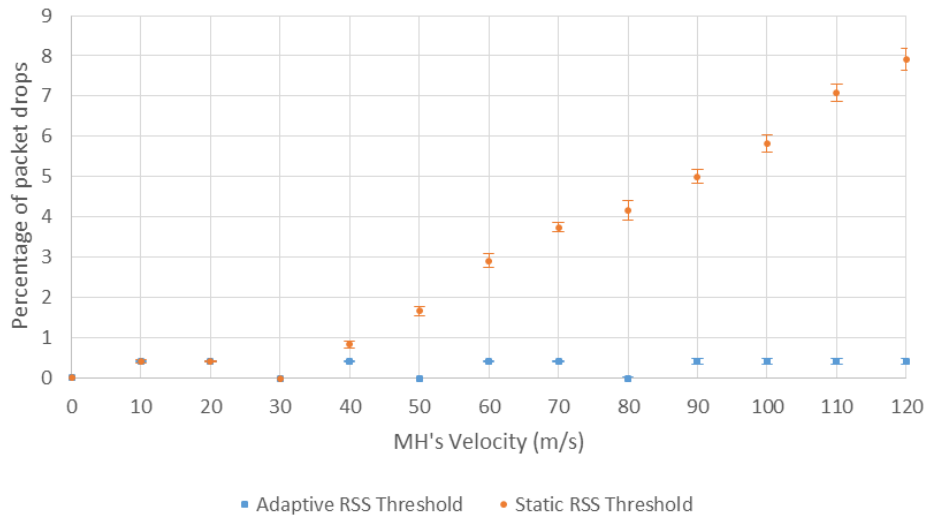


Figure 6.18 The comparison between the proposed adaptive threshold approach and the static threshold approach (static at -81dBm).

To test the effectiveness of the proposed scheme, the same configuration used to obtain figure 6.17 is considered. Figure 6.18 shows the comparison of the simulation results when using the proposed adaptive RSS threshold, and when using static threshold configured at -81 dBm . Figure 6.18 clearly shows the advantage of using the adaptive threshold scheme. Take note that the end

to end delay is configured at 450ms. This means, the minimum time required to finish a handover process will be 900ms, and this value can be more depending on the network traffic. Thus, packet drops are inevitable even though multi-home environment is used. From the graph in figure 6.18, it can be seen that when using static RSS threshold, more packet drops can be observed. This is due to the handover failure, where the MH has already moved out of the overlapped area between network 1 and network 2 before the handover process is completed. Meanwhile, using the proposed adaptive threshold scheme, the percentage of packet drops can be maintained below 1%. Thus this graph has clarified the effectiveness of the proposed adaptive RSS threshold approach.

In a nutshell, static RSS threshold is not enough to cope with the velocity of the MH when the MH moves at faster speeds. Thus, the adaptive RSS threshold scheme is proposed in order to alleviate the issues faced when using static RSS threshold. The effectiveness of the proposed approach has been clarified, and the result shows that the proposed adaptive RSS threshold outperforms the static RSS threshold.

6.6. Concluding Remarks

In this paper, several components to maintain the Quality of Service during the handover was proposed, namely: 1) an ANN-based velocity estimation method; 2) the warm-up time technique to improve the travel distance estimation; and 3) the adaptive RSS threshold approach to adapt the RSS threshold to the MH's velocity. The information on the mobile host's velocity is imperatively important. Thus, a simple and accurate velocity estimation using Artificial Neural Network based on the Received Signal Strength (RSS) and the rate of RSS change as the input has been proposed. The concept of warm-up time is useful to increase the accuracy of the estimated travel distance. Then, the travel distance is useful to avoid unnecessary handover and adapt the RSS threshold to avoid handover failure due to late handover initiation. The results showed that the proposed velocity estimation method can estimate the velocity of the mobile host accurately. Furthermore, the warm-up time can ensure the accuracy of the travel distance estimation, which in turn is used to adapt the RSS threshold to the velocity of the mobile host to maintain minimum packet loss during the handover process. Finally, it can be concluded that with the combination of the proposed methods, the overall service quality experienced by the users can be maintained even though mobile hosts are moving at high velocity.

CHAPTER 7

DISCUSSIONS, CONCLUSIONS AND FUTURE WORKS

This chapter concludes the dissertation and figures out open research directions of great interest for the future work.

7.1. Discussion and summary

In the current scenario, powerful mobile devices has become one of the norms in human life. Most people these days possess at least one mobile device; usually a person would have multiple mobile devices that he/she carries in daily life. Furthermore, the mobile devices have become more sophisticated and is estimated to have the computational power that rivals a normal desktop PC. These devices are capable of using a variety of computationally challenging applications. Some such application is the real-time application, namely the VoIP.

VoIP is a delay-sensitive application. For normal static users, the quality of the call can be guaranteed via stable wired networks. However, for mobile users, to maintain the quality of the call while moving is quite a challenge, especially when the users are capable of achieving high mobility by commuting with trains or busses or other form of transportation. In coping with this issue, several mobility management protocols have been proposed; most popularly known are Mobile IP and its variants. However these methods have faced deployment issues due to problems such as needing major modifications before the protocol can be used and other problems such as the handover latency and the scalability issues.

Thus, researchers are becoming more interested in ways that can avoid the stated issues. One way of doing it is by pushing the complexity of the network based protocols to the endpoints, which are the mobile devices and the service servers. Several approaches have proposed and two of the methods most significant to this study are SIGMA and ECHO. These two approaches which are based on the endpoint centric and multi-home concept has shown interestingly good results. However, there is still room for improvement. SIGMA which only implements RSS values to trigger the handover is not able to sustain the quality of the network needed for real time applications like VoIP. Thus ECHO strives to alleviate this issue by considering the ITU-T E-Model Mean Opinion Score which can be used as the indicator of the call quality received by users. However, ECHO was restricted static threshold based triggering approach.

Therefore, the proposed work in this dissertation strives to enhance these to mechanism to further improve the handover process for mobile users. Towards this end, several issues needs to be clarified and resolved:

- I. Which protocol is suitable for the implementation of the handover process?
- II. Single-home or Multi-home?
- III. Endpoint centric or Network centric?
- IV. What kind of decision metrics should be used?
- V. What kind of algorithm to utilize these metric?
- VI. How effective compared to existing methods?

Therefore, the work disclosed in this dissertation aims to clarify these questions.

For the first three questions, the answers are, (i) the Host Identity Protocol (HIP), (ii) Multi-home, (iii) endpoint centric. As discussed, to alleviate the deployment problems, endpoint centric approach is one of the best choices, because the method is implemented on the endpoints which are the MH and CH. No network modifications are required. Then the question that arises as mention by the first and second questions, which mobility management protocol should be used, and what kind of performance that endpoint centric and multi-home approach can obtain. These questions have been answered the feasibility study in chapter 3. From the study in chapter 3, it is proven that multi-home is important and useful to avoid packet losses. This is especially true for VoIP applications. The study in chapter 3 has also justified that HIP is the best protocol to use for mobility management, compared to SCTP and MIP-based approaches. Then in chapter 4, the effectiveness study of using end-point centric in terms of signaling cost has also been presented. From the study in chapter 4, the endpoint centric approach, HIP particularly, has better signaling cost compared to network centric approach (HMIPv6). With this study it has been clarified that the endpoint centric, multi-home approach is the best option to implement the handover approach.

For real-time applications, such as VoIP, the most important thing is the timeliness of the packet arrival. The main culprit in deteriorating the quality of a call is the end-to-end delay, jitter and the packet loss. One way to measure the quality of the call is using the ITU-T recommended model, the E-Model to estimate the Mean Opinion Score, which is a method to score the Quality of Experience (QoE) of the users. This model is very useful to monitor the expected QoE of a wireless access. However, as it is, the E-Model cannot be used in real-time, because of some parameters that cannot be monitored by the MH. Thus a simplified version of the E-Model is used, where the original equations are simplified using some standard values so that the only component left is the delay metrics and the packet loss component of the original equation. Using the simplified E-Model, the MOS value can be obtained from the end-to-end delay, jitter and packet loss information. Thus these three parameters is useful for the handover decision. The QoE factor, namely the MOS value here, is important in order to maintain the quality of a VoIP call. However, not much work has considered this metric as the handover decision metrics. Thus, the study in this dissertation proposed an algorithm to utilize this information efficiently. Besides the QoE, the

physical condition of the wireless network is also important, because it is the deciding factor determining the connection of a MH to the network. Thus, this information was also incorporated into the proposed handover algorithm.

To process the considered handover metrics (end-to-end delay, jitter, packet loss, and RSS) a simple yet powerful algorithm is required. In the literature, the bio-inspired approaches have become more popular, due to the characteristics that came from their biological counterpart which makes them robust and tolerant to the dynamic changes in the environment. One such approach that has been used widely, and has been adapted into many research fields is the Ant Colony Optimization (ACO). In the field of network communication, one of the most successful variant of the ACO is the AntNet algorithm, which is an algorithm developed to tackle the route optimization problem in the communication network. From the literature, this method has shown exceptional performance in finding the best path for packet forwarding in a distributed way. This approach can be related to the network selection algorithm for the vertical handover process.

Thus, the proposed method in this dissertation (chapter 5), namely, the AntNet-Based Handover Algorithm (ANHA), adapted the concept from AntNet into its algorithm in determining the best network for the MH. This algorithm incorporates the bio-inspired features of the ants, where the goodness of the network connected to the MH is updated by ants, namely via the pheromone, and the value of the pheromone increases or decreases according to the information brought by the ants. The key point of the ant approach is the evaluation of a network goodness: (i) implicit way – through the pheromone update process (which incorporates a memory like mechanism), and (ii) explicit way – through the immediate network information (RSS in this case).

Finally, ANHA as whole is compared to the existing method, the ECHO, in order to validate the effectiveness of the proposed handover approach. From the comparison study, it can be concluded that ANHA has improved some of the drawbacks of the existing approach, (ECHO specifically.)

However, velocity is also one of the important factors in the mobile network. The velocity of the MH can affect the seamlessness in the handover process, since it affects the MH connection duration in a network cell, and can also cause handover failures, due to insufficient time to complete the handover process. Thus this issue was also considered in chapter 6a. Before velocity can be dealt with, the MH needs to be able to detect its velocity. It is generally assumed that the velocity is obtainable using the accelerometer or the GPS built in a mobile device. However, the velocity information is sometimes unavailable due power saving features of the mobile device, or the fact that the mobile device itself is not equipped with a GPS. Thus, an ANN-based velocity estimation method that utilizes the RSS and the rate of RSS change is proposed in this dissertation. The results from the study have shown that the proposed method is superior to an existing method [97]. Then the velocity information is used to estimate the travelling distance of the MH in a network and also to estimate the safety margin to avoid handover failure. In this dissertation, a new concept, namely the warm-up time, to increase the accuracy of the travelling distance estimation is also introduced. In an existing method [110] the accuracy of TDE becomes worse when the MH is moving at 10m/s and below. Using the proposed warm-up time, the accuracy of

the TDE can be improved to nearly 100%. Finally, a simple adaptive RSS threshold approach is proposed to avoid handover failure due to losing the current connection during a handover process. The improvement of the MH velocity estimation and the TDE can directly influence the reliability of this method. The synergy of these three proposed approach in chapter 6 can ultimately improve the service quality experienced by mobile users.

7.2. A use case scenario

To provide a better understanding of the functions of the proposed methods, an example is given to demonstrate a scenario of the chronological order in which a MH utilizes ANHA, the existing TDE enhanced with the proposed ANN-based VEM and warm-up time as well as the proposed ART. Figure 7.1 and 7.2 shows the scenario when a MH is moving at low velocity and high velocity respectively.

Low velocity example (e.g. 10m/s):

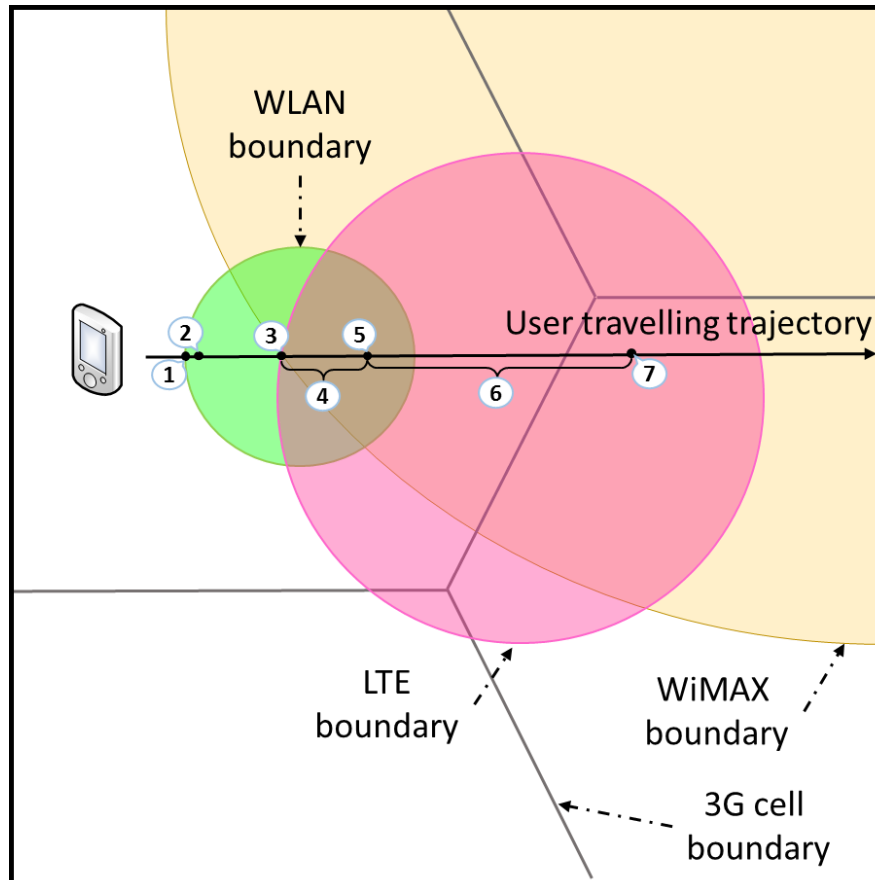


Figure 7.1 The use case scenario for MH with low velocity.

1. The MH is connected to a 3G network and senses the availability of a WLAN, so it incites the UHA to decide whether it should avoid or consider WLAN.

2. MH decides that WLAN is suitable as a candidate and starts ANHA. Then ANHA determines that WLAN has higher $p_i(t)$ compared to the 3G network and MH performs handover to the WLAN.
3. The MH detects the availability of candidate networks including an LTE and a WiMAX network. MH then invokes ANHA.
4. ANHA monitors the current network (WLAN) and the alternative networks (3G, LTE and WiMAX).
5. At this point ANHA decides that LTE is the best network and the MH performs handover to the LTE network.
6. ANHA continues to monitor the current network (LTE) and the alternative networks (3G and WiMAX).
7. At this point ANHA determines that the WiMAX network is the best network and the MH performs handover to WiMAX network.

High velocity example (e.g. 120m/s):

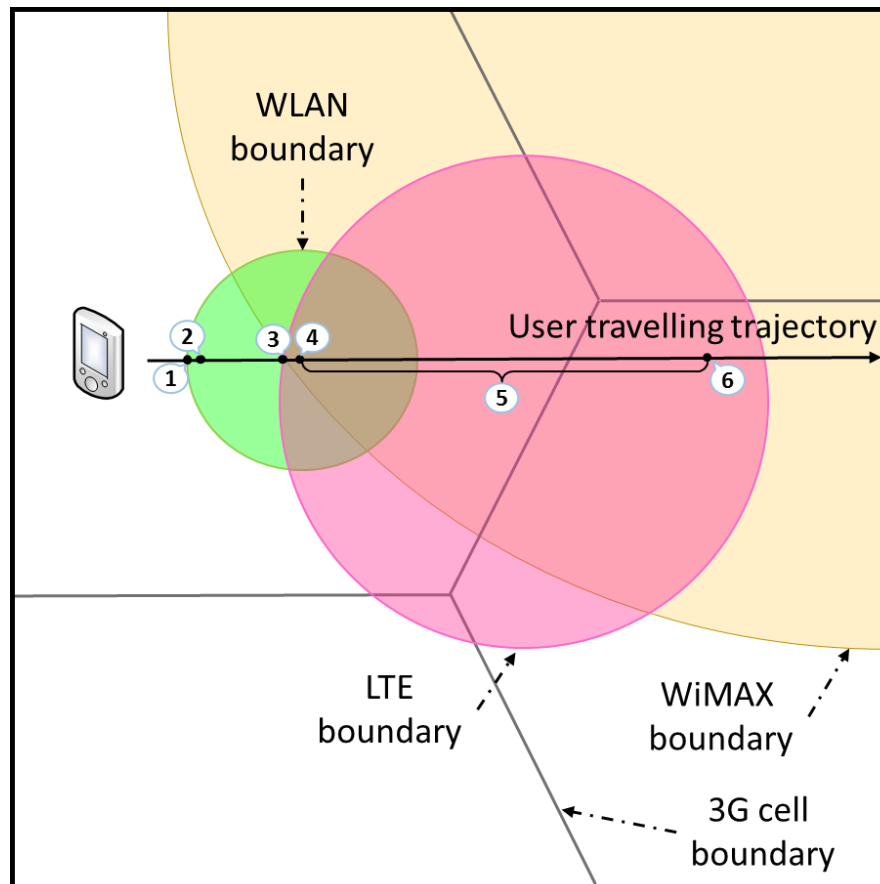


Figure 7.2 The use case scenario for MH with high velocity.

1. The MH is connected to a 3G network and senses the availability of a WLAN, so it incites the UHA to decide whether it should avoid or consider WLAN.

2. MH decides that WLAN is unnecessary due to the short MH travelling distance estimated within the WLAN.
3. The MH detects an LTE network and a WiMAX network; it activates UHA to decide whether the LTE network and WiMAX network is suitable or not, and activates ANHA.
4. ANHA determines that the LTE network is the best network at the moment. MH performs a handover to the LTE network.
5. ANHA continues to monitor the current network (LTE) and the alternative networks (3G and WiMAX). ANHA will trigger a handover if it deems that other network is better than the current network.
6. The threshold configured by ART. If the current primary network is the LTE network, at this point a handover will still be performed to the best target candidate, regardless of the ANHA score of the current network. This is to avoid handover failure due to the lateness of starting the handover process.

7.3.Conclusions

This dissertation proposed a notable handover approach which is based on several components: (i) Endpoint centric, (ii) Multi-home, and (iii) Handover triggering algorithm that is biologically-inspired by the Ant Colony Optimization Concept.

Firstly a feasibility study on the effectiveness of using the endpoint centric, multi-home environment is presented. In this feasibility study, HIP (representative of multi-home-enabled, endpoint centric method) is compared to existing single-homed, network based method such as MIP and other MIP-based methods such as CIP and HAWAII, in terms of the packet loss rate and the handover latency. From this study, it can be concluded that HIP outperforms the other protocols in term of the packet loss rate, since it has the multi-home capability. From this study, it is determined that multi-home is an important factor in achieving a seamless loss less handover. However, in term handover latency, hip has a higher latency due to the end-to-end communication needed during the handover process. Thus making the latency for HIP at least as long as one round trip time from MH to CH. Conversely, the multi-home feature compensates for this drawback by allowing the MH to connect to multiple networks at once. Furthermore, an existing study is also presented with this feasibility study to justify its effectiveness compared to another endpoint centric multi-home protocol, which is the SCTP. That study clearly shows that HIP outperforms SCTP in terms of the multi-home feature, thus solidifying the choice of using HIP for the proposed method.

Then another feasibility study on the effectiveness of endpoint based approach and network based approach, is presented. This study looks into the signaling cost incurred by HIP and HMIPv6. From this study, it can be concluded that HIP outperforms HMIPv6 in terms of signaling cost. This study further fortifies the decision on using HIP as the mobility management and for the execution of the proposed handover method.

Thirdly, the Bio-Inspired Device Oriented Handover approach (ANHA) is proposed. The bio-inspired features have been discussed, as well as the decision metrics collection, processing

and the handover decision strategy has been discussed thoroughly. Then the evaluation of the proposed method is done by comparing the performance of the proposed method to ECHO, an existing endpoint centric multi-home handover approach which also incorporates the same decision metrics as ANHA. From the comparison study, the effectiveness of ANHA compared to ECHO has been clarified.

Finally, the issues on the velocity is considered. In order to cater for the effects of the velocity to the handover process, a simple yet useful velocity estimation tool utilizing the artificial neural network is proposed. Then a simple and effective strategy on using the velocity information to develop an adaptive RSS threshold is presented and the effectiveness clarified.

As concretely presented in this dissertation, not only the enhancement on the handover process was proposed, but the effectiveness of the proposed solution was thoroughly evaluated via simulation studies. The achievements in this work confirm the effectiveness of ANHA in maintaining and enhancing the quality of VoIP application for mobile users as well as the effectiveness of the proposed ANN-based VEM, the warm-up time and the ART in improving the velocity estimation accuracy, the TDE accuracy and avoiding the handover failures respectively.

7.4.Future Works

There are several additional issues that needs to be considered before the proposed ANHA can be fully implemented in the real system.

- **The NAT traversal issues** faced by any new protocols such as SCTP and HIP needs to be resolved. NAT usually filters the packets for security reasons. They are also capable of doing deep searching of packets, and packets with a content using unknown format will be dropped, thus stopping protocols such as HIP and SCTP to be used in the real network.
- **Cross layer communication** is currently not possible in real devices, thus impeding the use of the handover more sophisticated handover metrics that needs to be shared via cross-layer communication. One possible option currently is the implementation of the Mobile Independent Handover, IEEE802.21.
- **Virtual NIC.** Currently mobile devices are only equipped with only one antenna for each network interface card (NIC), making it impossible to connect to multi-home to the same radio access technology (RAT). Recently, Microsoft have introduced the virtual NIC, where one interface card can support several connections to different access points. Such research is on interesting path to go.
- **Real test bed.** Currently, the proposed work have only been evaluated in the simulation. An experiment on the real test bed is needed to further study the applicability of the proposed method into the real world.

References

1. Conti, J.: The long road to WiMAX [wireless MAN standard]. IEE Review, vol. 51, no. 10, pp. 38-42, Oct (2005)
2. Abeta, S.: Toward LTE commercial launch and future plan for LTE enhancements (LTE-Advanced). 2010 IEEE International Conference on Communication Systems (ICCS), vol. 146, no. 150, pp. 17-19, November (2010).
3. Wang, J.; McCann, P. J.; Gorrepati, P. B.; Liu, C.-Z.: Wireless voice-over-IP and implications for third-generation network design. Bell Labs Technical Journal, vol. 3, no. 3, pp. 79-97, July (1998)
4. Marquez-Barja, J.; Calafate, C. T.; Cano, J.-C.; Manzoni, P.: An overview of vertical handover techniques: Algorithms, protocols and tools. Computer Communications , vol. 34, no. 8, pp. 985-997 (2011)
5. Yan, X.; Sekercioglu, Y. A.; Narayanan, S.: A survey of vertical handover decision algorithms in Fourth Generation heterogeneous wireless networks. Computer Networks , vol. 54, no. 11, pp. 1848-1863 (2010)
6. McNair, J.; Zhu, F.: Vertical handoffs in fourth-generation multinet environments. Wireless Communications, IEEE, vol. 11, no. 3, pp. 8-15, June (2004)
7. Handley, M.: Why the Internet only just works. BT Technology Journal, vol. 24, no. 3, pp. 119-129 (2006)
8. Saltzer, J. H.; Reed, D. P.; Clark, D. D.: End-to-end Arguments in System Design. ACM Trans. Comput. Syst., vol. 2, no. 4, pp. 277-288, Nov (1984)
9. Hussain, R.; Malik, S. A.; Khan, S. A.; Abrar, S.: Design, implementation and experimental evaluation of an end-to-end vertical handover scheme on NCTUns simulator. Simulation Modelling Practice and Theory , vol. 26, no. 0, pp. 151-167 (2012)
10. Fu, S.; Ma, L.; Atiquzzaman, M.; Lee, Y.-J.: Architecture and performance of SIGMA: a seamless mobility architecture for data networks. IEEE International Conference on Communications (2005)
11. Fu, S.; Atiquzzaman, M.: Handover latency comparison of SIGMA, FMIPv6, HMIPv6, FHMIPv6. Global Telecommunications Conference, 2005. GLOBECOM '05. IEEE, (2005)
12. Fu, S.; Ivancic, W.; Atiquzzaman, M.: Signaling cost evaluation of SIGMA. IEEE 62nd Vehicular Technology Conference (2005)
13. Fitzpatrick, J.; Murphy, S.; Atiquzzaman, M.; Murphy, J.: Using Cross-layer Metrics to Improve the Performance of End-to-end Handover Mechanisms. Comput. Commun., vol. 32, no. 15, pp. 1600-1612, Sep (2009)
14. Fitzpatrick, J.; Murphy, S.; Atiquzzaman, M.; Murphy, J.: ECHO – A Quality of Service Based Endpoint Centric Handover Scheme for VoIP. Wireless Communications and Networking Conference, 2008. WCNC 2008. IEEE, (2008)

15. Fitzpatrick, J.; Murphy, S.; Atiquzzaman M.; Murphy, J.: Evaluation of VoIP in a Mobile Environment using an end-to-end Handoff Mechanism. Mobile and Wireless Communications Summit, 2007 16th IST (2007)
16. ITU-T: The E-Model, A computational Model for Use in Transmission Planning. G.107, March (2005)
17. Stewart, R.: Stream Control Transmission Protocol. IETF Internet draft, proposed standard, September (2007)
18. Dorigo, M.; Birattari, M.; Stutzle, T.: Ant colony optimization. Computational Intelligence Magazine, IEEE, vol. 1, no. 4, pp. 28-39, Nov (2006)
19. Di Caro, G.; Dorigo, M.: AntNet – Distributed Stigmergic Control for Communications Networks. J. Artif. Int. Res. 9, pp. 317-365, December (1998).
20. IETF: IP Mobility Support. RFC 2002, October (1996).
21. Perkins, C. E.: Mobile IP. Communications Magazine, IEEE, pp. 84-99 (1997).
22. Perkins, C. E.: Mobile networking through Mobile IP. Internet Computing, IEEE, pp. 58-69 (1998).
23. Perkins, C. E.; Kuang-Yeh, W.: Optimized smooth handoffs in Mobile IP. IEEE International Symposium on Computers and Communications, pp. 340-346 (1999).
24. Perkins, C. E.: Mobile IP. Communications Magazine, IEEE, pp.66-82, May (2002).
25. IETF: ICMP Router Discovery Messages. RFC 1256, September (1991).
26. IETF: Mobility Support in IPv6. RFC 3775, (Proposed Standard), June (2004).
27. Shmitz, P.: MIPv6 – New Capabilities for Seamless Roaming Among Wired, Wireless, and Cellular Networks. Intel Developer Update Magazine, September (2002).
28. Dimopoulou, L.; Leoleis, G.; Venieris, I. O.: Fast handover support in a WLAN environment – challenges and perspectives. Network, IEEE, pp. 14-10, May-June (2005).
29. Koodli, R.: Fast Handovers for Mobile IPv6. RFC 4068, (Experimental), July (2005).
30. Castelluccia, C.: HMIPv6 – A hierarchical mobile IPv6 proposal. SIGMOBILE Mob. Comput. Commun. Rev., pp. 48-59, January (2000).
31. Perez-Costa, X.; Torrent-Moreno, M.; Hartenstein, H.: A performance comparison of Mobile IPv6, Hierarchical Mobile IPv6, fast handovers for Mobile IPv6 and their

- combination. SIGMOBILE Mob. Comput. Commun. Rev., pp.5-19, (2003).
32. Koodli, R.; Perkins, C. E.: Fast handovers and context transfers in mobile networks. SIGCOMM Comput. Commun. Rev., pp. 37-47, (2001).
 33. Fogelstroem, E.; Jonsson, A.; Perkins, C. E.: Mobile IPv4 Regional Registration. RFC 4857, (Experimental), June (2007).
 34. Valko, A. G.: Cellular IP – a new approach to Internet host mobility. SIGCOMM Comput. Commun. Rev., pp. 50-65, (1999).
 35. Campbell, A. T.; Gomez, J.; Valko, A. G.: An overview of cellular IP. Wireless Communications and Networking Conference 1999, pp. 606-610, (1999).
 36. Campbell, A. T.; Gomez, J.; Kim, S.; Valko, A. G.; Chieh-Yih, W.; Turanyi, Z. R.: Design, implementation, and evaluation of cellular IP. IEEE Personal Communications, pp. 42-49, August (2000).
 37. Ramjee, R.; Varadhan, K.; Salgarelli, L.; Thuel, S. R.; Shie-Yuan, W.; La Porta, T.: HAWAII – a domain-based approach for supporting mobility in wide-area wireless networks. IEEE/ACM Transactions on Networking, pp. 396-410, Jun (2002).
 38. Campbell, A.; Gomez-Castellanos, J.: IP micro-mobility protocols. SIGMOBILE Mob. Comput. Commun. Rev., pp. 45-53, (2000).
 39. Cambell, A. T.; Gomez, J.; Kim, S.; Chieh-Yih, W.; Turanyi, Z. R.; Valko, A. G.: Comparison of IP micromobility protocols. IEEE Wireless Communications, pp. 72-82, February (2002).
 40. Eddy, W. M.: At what layer does mobility belong? Communications Magazine, IEEE, pp. 155-159, October (2004).
 41. Snoeren, A. C.; Balakrishnan, H.: An end-to-end approach to host mobility. MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking, pp. 155-166, New York, USA (2000).
 42. Snoeren, A. C.; Balakrishnan, H.; Kaashoek, M. F.: Reconsidering Internet mobility. Proceedings of the Eighth Workshop on Hot Topics in Operating Systems, pp. 41-46, 20-22 May (2001).
 43. Matsumoto, A.; Kozuka, M; Fujikawa, K.; Okabe, Y.: TCP Multi-Home Options – Draft. IETF Internet-Draft draft-arifumi-tcp-mh-00.txt(work in progress), October (2003).
 44. Fu, S.; Atiquzzaman, M.: SIGMA – A Transport Layer Handover Protocol for Mobile

- Terrestrial and Space Networks. In: Ascenso, J.; Vasiu, L.; Belo, C.; Saramago, M.: e-Business and Telecommunication Networks, Springer, pp. 41-52,(2006).
45. Fu, S.; Atiquzzaman, M.: Handover latency comparison of SIGMA, FMIPv6, HMIPv6, FHMIPv6. IEEE Global Telecommunications Conference 2005, pp. 3809-3813, December (2005).
 46. Fu, S.; Atiquzzaman, M.; Lee, Y.: Architecture and performance of SIGMA – a seamless mobility architecture for data networks. IEEE International Conference on Communications, pp. 3249-3253, 16-20 May (2005).
 47. Fu, S.; Atiquzzaman, M.; Ma, L.; Lee, Y. J.: Signaling cost and performance of SIGMA – A seamless handover scheme for data networks. Wireless Communications and Mobile Computing, vol. 5, no. 7, pp. 825-845, (2005).
 48. Stewart, R.: Stream Control Transmission Protocol. RFC 4960 (proposed standard), September (2007).
 49. Stewart, R.; Xie, Q.; Tuexen, M.; Maruyama, S.; Kozuka, M.: Stream Control Transmission Protocol (SCTP) Dynamic Address Reconfiguration. RFC 5061 (proposed Standard), September 2007.
 50. Koh, S.; Chang, M.; Lee, M.: mSCTP for soft handover in transport layer. IEEE Communications Letters, pp. 189-191, March (2004).
 51. Koh, S.; Jung, H.; Min, J.: Transport layer internet mobility based on mSCTP. The 6th International Conference on Advanced Communication Technology, pp. 329-333, (2004).
 52. Aydin, I.; Seok, W.; Shen, C.: Cellular SCTP: a transport-layer approach to Internet mobility. Proceedings of The 12th International Conference on Computer Communications and Networks, pp. 285-290, 20-22 October (2003).
 53. Aydin, I.; Shen, C.: Evaluating cellular SCTP over one-hop wireless networks. IEEE 62nd Vehicular Technology Conference 2005, VTC-2005-Fall, pp. 826-830, 25-28 September (2005).
 54. Fitzpatrick, J.; Murphy, S.; Atiquzzaman, M.; Murphy, J.: Evaluation of VoIP in a Mobile Environment using an end-to-end Handoff Mechanism. Mobile and Wireless Communications Summit, 2007, 16th IST (2007).
 55. Moskowitz, R.; Nikander, P.; Jokela, P.; Henderson, T.: Host Identity Protocol. IETF RFC 5201 (2008).
 56. Nikander, P.; Henderson, T.; Vogt, C.; Arkko, J.: End-Host Mobility and Multihoming

with the Host Identity Protocol. IETF RFC 5206 (2008).

57. Laganier, J.; Eggert, L.: Host Identity Protocol (HIP) Rendezvous Extension. IETF RFC 5204 (2008).
58. Zekri, M.; Jouaber, B.; Zeghlache, D.: A review on mobility management and vertical handover solutions over heterogeneous wireless networks. *Computer Communications*, vol. 35, no. 17, pp. 2055-2068 (2012).
59. Dressler, E.; Akan, O. B.: A survey on bio-inspired networking. *Computer Networks*, vol. 54, no. 6, pp. 881-900 (2010).
60. Wang, J. B.; Chen, M.; Wan, X.; Wei, C.: Ant-colony-optimization-based scheduling algorithm for uplink CDMA nonreal-time data. *IEEE Transactions on Vehicular Technology*, vol. 58, no. 1, pp. 231-241 (2009).
61. Meisel, M.; Pappas, V.; Zhang, L.: A taxonomy of biologically inspired research in computer networking. *Computer Networks*, vol. 54, no. 6, pp. 901-916 (2010).
62. Zheng, C.; Sicker, D.: A survey on biologically inspired algorithms for computer networking. *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1160-1191 (2013).
63. Kitano, H.: Biological robustness. *Nature Reviews Genetics*, vol. 5, no. 11, pp. 826-837 (2004).
64. Kitano, H.: Towards a theory of biological robustness. *Molecular Systems Biology*, vol. 3, no. 1, article 137, (2007).
65. Dorigo, M.; Birattari, M.; Stutzle, T.: Ant Colony Optimization. *Computational Intelligence Magazine IEEE*, vol. 1, no. 4, pp. 28-39, November (2006).
66. Deneubourg, J. -L.; Aron, S.; Goss, S.; Pasteels, J. M.: The self-organizing exploratory pattern of the Argentine ant. *Journal of Insect Behaviour*, vol. 3, no. 3, pp. 159-168, March (1990).
67. Goss, S.; Aron, S.; Deneubourg, J. L.; Pasteels, J. M.: Self-organized shortcuts in the Argentine ant. *Naturwissenschaften*, vol. 76, no. 12, pp. 579-581, December 1989.
68. Dorigo, M.; Maniezzo, V.; Colnari, A.: Positive feedback as a strategy. Dipartimento di Elettronica, Politecnico di Milano, Italy, Tech. Rep. 91-016 (1991).
69. Dorigo, M.: Optimization, learning and natural algorithms (in Italian). Ph.D. dissertation, Dipartimento di Elettronica, Politecnico di Milano, Italy, (1992).

70. Dorigo, M.; Maniezzo, V.; Colorni, A.: Ant System – Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man and Cybernatics – Part B*, vol. 26, no. 1, pp. 29-41 (1996).
71. Stutzle, T.; Hoos, H. H.: MAX-MIN Ant System. *Future Generation Computer Systems*, vol. 16, no. 8, pp. 889-914, 2000.
72. Dorigo, M.; Gambardella, L. M.: Ant colonies for the traveling salesman problem. *BioSystems*, vol. 43, no. 2, pp. 73-81 (1997).
73. Dorigo, M.; Gambardella, L. M.: Ant Colony System – A cooperative learning approach to the traveling salesman problem. *IEEE Transactions on Evolutionary Computation*, vol. 1, pp. 53-66 (1997).
74. Gambardella, L. M.; Dorigo, M.: Solving symmetric and asymmetric TSPs by ant colonies. *Proc. 1996 IEEE International Conference on Evolutionary Computation (ICEC'96)*, pp. 622-627 (1996).
75. Di Caro, G.; Dorigo, M.: AntNet – distributed stigmergetic control for communications networks. *Journal of Artificial Intelligence Research*, vol. 9, no. 1, August (1998).
76. Bokor, L.; Novaczki, S.; Imre, S.: A Complete HIP based Framework for Secure Micromobility. *5th International Conference on Advances in Mobile Computing and Multimedia*, pp 111-122, 3-5 December (2007).
77. Arraez, L.; Chaouchi, H.; Gurkas Aydin, Z.: Performance Evaluation and Experiments for Host Identity Protocol. *International Journal of Computer Science Issues*, vol. 8, no. 2, March (2011).
78. Zeadally, S.; Siddiqui, F.: An Empirical Analysis of Handoff Performance for SIP, Mobile IP, and SCTP Protocols. *Wireless Personal Communications*, vol. 43, no. 2, pp. 589-603, October (2007).
79. Mugga, C.; Sun, D.: Performance comparison of multihoming and mobility protocols in IPv6 heterogeneous network environment. *Master Thesis, School of Computing, Blekinge Institute of Technology, SE-371 79 Karlskrona, Sweden*, October (2013).
80. Grilo, A.; Estrela, P.; Nunes, M.: Terminal independent mobility for IP (TIMIP). *IEEE Communications Magazine*, vol. 39, no. 12, pp. 34-41, December (2001).
81. Nyquist, H.: Certain topics in telegraph transmission theory. *Proceedings of the IEEE*, vol. 90, no. 2, pp. 280-305, February (2002).
82. ITU-T Recommendation G.711: Pulse Code Modulation (PCM) of Voice Frequencies,

(1972).

83. ITU-T Recommendation G.720: Coding of speech at 8kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP), (1996).
84. Srinivinsan, K.; Gersho, A.: Voice activity detection for cellular networks. IEEE Workshop on Speech Coding for Telecommunications, pp. 85-86, (1993).
85. Baset, S. A.; Schulzrine, H.: An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol, December (2004).
87. Psytechnics: Estimating E-Model Id within a VoIP Network. Technical Report, Psytechnics (2002).
88. Schulzrinne, H.; Rosenberg, J.: The Session Initiation Protocol – Internet-centric signaling. Communications Magazine, IEEE, vol. 38, no. 10, pp. 134-141, Oct (2000).
89. ITU-T Recommendation G.113: Appendix I – Provisional planning values for the equipment impairment factor I_e . October (2001).
90. ITU-T Recommendation G.114: One-way transmission time. May (2003).
91. Xie, J.; Akyildiz, I.: A novel distributed dynamic location management scheme for minimizing signaling costs in Mobile IP. IEEE Transactions on Mobile Computing, vol. 1, no. 3, pp. 162-175, July (2002).
92. Xie, J.; Akyildiz, I.: A novel distributed dynamic location management scheme for minimizing signaling costs in Mobile IP. IEEE Transactions on Mobile Computing, vol.1, no. 3, pp. 163-175, July (2002).
93. Chung, Y. W.; Sung, D. K.; Aghvami, A.: Steady state analysis of P-MIP mobility management. IEEE Communications Letters, vol.7 no.6, pp. 278-280, June (2003).
94. Bettstetter, C.; Hartenstein, H.; Perez-Costa, X.: Stochastic properties of the random waypoint mobility model – Epoch length, direction distribution, and cell change rate. ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems, pp. 7-14, September (2002).
95. Crovella, ME.; Bestavros, A.: Self-similarity in world wide web traffic: evidence and possible causes. IEEE/ACM Transactions on Networking, pp. 835-846, (1997).
96. Shulzrinne, H.; Rosenberg, J.: The Session Initiation Protocol – Internet-centric signaling. IEEE Communications Magazine, IEEE, vol. 38, no. 10, pp. 134-141, October (2000).

97. Gu, J.; Bae, S. J.; Chung, Y. M. Cheon, K-. Y.; Park, A-. S.: Mobility-based Handover Decision Mechanism to Relieve Ping-Pong Effect in Cellular Networks. *IEEE 16th Asia-Pacific Conference on Communication (APCC)*, (2010).
98. Engelbrecht, A. P.: Artificial Neural Networks, in Computational Intelligence. John Wiley and Sons, pp. 6-7,(2007).
99. A. Kumar, Y. Liu, J. Sengupta, Divya, "Evolution of Mobile Wireless Communication Networks – 1G to 4G," *International Journal of Electronics & Communication Technology*, vol. 1, no. 1, December 2010.
100. S. Dekleva, J. P. Shim, U. Varshney, G. Knoerzer, "Evolution and Emerging Issues in Mobile Wireless Networks," *Communications of the ACM*, vol. 50, no. 6, June 2007.
101. I. Al-Surmi, M. Othman, B. M. Ali, "Review on mobility management for future-IP-based next generation wireless networks," *The 12th International Conference on Advanced Communication Technology (ICACT)*, vol.2, pp.989-994, 7-10 Feb. 2010.
102. I. F. Akyildiz, Jiang Xie, S. Mohanty, "A survey of mobility management in next-generation all-IP-based wireless systems," *Wireless Communications, IEEE*, vol.11, no.4, pp.16-28, Aug. 2004.
103. R. Hussain, S. A. Malik, S. A. Khan and S. ya, "Design, implementation and experimental evaluation of an end-to-end vertical handover scheme on NCTUns simulator," *Simulation Modeling Practice and Theory* , vol. 26, no. 0, pp. 151-167, 2012.
104. S. Abrar, R. Hussain, A. R. Riaz, S. A. Malik, S. A. Khan, G. Shafiq, S. Ahmed, "A new method for handover triggering condition estimation," *IEICE Electronics Express*, vol. 9, no. 5, pp. 378-384, 10 March 2012.
105. S. -J. Yoo, D. Cypher, N. Golmie, "Timely effective handover mechanism in heterogeneous wireless networks," *Wireless Personal Communications*, vol. 52, no. 3, pp. 449-475, 2010.
106. W. -I. Kim, B. -J. Lee, J. -S. Song, Y. -S. Shin, Y. -J. Kim, "Ping-Pong avoidance algorithm for vertical handover in wireless overlay networks" In 66th IEEE vehicular technology conference, fall VTC-2007, pp. 1509–1512, October 2007.
107. T. Kato, R. Takechi, H. Ono, "A study on mobile IPv6 based mobility management architecture," *FUJITSU Science Technology*, vol. 37, pp. 65–71, 2001.
108. Kato, T., Takechi, R., & Ono, H. : Survey and classification of transport layer mobility management schemes. In *IEEE 16th international symposium on personal, indoor and mobile radio communications*, 2005. PIMRC 2005 (Vol. 4, pp. 2109–2115), September

(2005).

109. A. Hasswa, N. Nasser, H. Hassanein, "A seamless context-aware architecture for fourth generation wireless networks," *Wireless Personal Communications*, vol. 43, no. 3, pp. 1035–1049, 2007.
110. X. Yan, Y. A. Sekercioglu, and N. Mani, "A method for minimizing unnecessary handovers in heterogeneous wireless networks," in *Proceedings of the 2008 International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM'08)*, pp. 1-5, USA, June 2008.
111. S. Mohanty, "A new architecture for 3G and WLAN integration and inter-system handover management," *Wireless Networks*, vol. 12, no. 6, pp. 733–745, 2006.
112. Z. -G. Che, T. -A. Chiang, Z. -H. Che, "Feed-forward Neural Networks Training – A Comparison Between Genetic Algorithm and Back-propagation Learning Algorithm," *International Journal of Innovative Computing, Information and Control*, vol. 7, no. 10, October 2011.
113. H. Riaz, A. M. Shahzad, A. Shafayat, A. R. Raja, A. Hassan, A. K. Shahid, "Vertical Handover Necessity Estimation Based on a New Dwell Time Prediction Model for Minimizing Unnecessary Handovers to a WLAN Cell," *Springer journal of Wireless Personal Communications*, pp. 1 - 14 Oct. 2012.
114. D. E. Rumelhart, B. Widrow, M. A. Lehr, "The basic ideas in neural networks," *Communication of the ACM*, vol. 37, no. 3, pp. 87-92, March 1994.
115. T. H. Martin, B. D. Howard, H. B. Mark, "Neural Network Design," PWS Publishing Company, Boston, 1996.
- 116 Jacobson, V.: Congestion avoidance and control. *ACM SIGCOMM Computer Communication Review - Special twenty-fifth anniversary issue, Highlights from 25 years of the Computer Communication Review*, vol. 25, no. 1, pp. 157-187, 1995.
- 117 Perera, C.; Liu, C.H.; Jayawardena, S.; Min Chen: A Survey on Internet of Things From Industrial Market Perspective. *Access, IEEE*, vol. 2, pp. 1660-1679, 2014.