Adaptive Vertical Handoff for Integrated UMTS and WLAN Networks

by

Liping Bai

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2007

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Liping Bai

Abstract

Next-generation wireless networks have been envisioned to be an integration of heterogeneous wireless access networks such as UMTS (Universal Mobile Telecommunication Networks) and the IEEE 802.11 based WLAN (Wireless Local Area Networks). It is an important and challenging issue to support seamless vertical handoff management in such an integrated architecture that provides the mobile users uninterrupted service continuity anywhere, any time. In such a networking environment, the signaling delay of the vertical handoff is not fixed due to the traffic load in the backbone Internet, wireless channel quality and the distance between a mobile node and its home network. However, the currently handoff solutions implicitly considers the signaling delay as a constant value. In this thesis, we study a typical link layer assisted handoff, identifying its deficiency due to the considerably large handoff delay. We propose an adaptive vertical handoff management scheme for integrated UMTS and WLAN networks. The proposed scheme incorporates the idea of pre-handoff with adaptive handoff threshold. We estimate the handoff signaling delay in advance, therefore, providing the delay information required for making an adaptive handoff decision. Instead of setting a fixed threshold, an adaptive handoff threshold value is determined for every single MN based on the estimated handoff signaling delay. The RSS and the RSS's rate of change are used to determine the estimated handoff time instant. Extensive simulation has been conducted to verify the performance of the proposed handoff scheme.

Acknowledgements

The credit for making this thesis possible to accomplish does not belong only to one single person. Instead it is the sincerely appreciated support of many different persons who have helped during my research work that deserves a hats-off from me.

First of all, I would like to express my gratitude and appreciation to my supervisor, Prof. Xuemin (Sherman) Shen, for guidance, inspiring discussions, proofreading, and for giving me the opportunity to do this work.

I would also like to thank Prof. Liang-liang Xie and Prof. Sagar Naik for reviewing this thesis. The department staff, Wendy Boles and Karen Schooley are acknowledged for their administrative support.

Thanks to my friends in BBCR lab for their friendship, help and discussions, especially Stanley Liu, Fen Hou, Chenxi Zhang, and Ying Wang. Special thanks to Jim Guo for his help during my thesis writing.

My deepest gratitude, love and affection belong to my parents, my brother and my relatives for supporting but not understanding my work.

Dedication

to my parents and my brother
for their endless love and support

Contents

List of Figures

Chapter 1

Introduction

Internet has made a significant change in both our daily work and life. In the early days, people use Internet for services such as Web browsing, email and file transfer, which are relatively delay- and bandwidth-tolerant. Nowadays, however, more advanced multimedia applications like IP telephony, video conferencing and online gaming are being widely employed by more and more people [1]. At the same time, more and more people demand these services to be available anywhere, any time, and through any access technologies.

There have been various exiting and emerging wireless access technologies, e.g., IEEE 802.11-based wireless local area networks (WLANs), IEEE 802.16 wireless metropolitan area networks (WMANs), General Packet Radio Service (GPRS), and Universal Mobile Telecommunications System (UMTS). These technologies vary widely in terms of bandwidths, media access technologies, security mechanisms, etc., but they are complementary to each other in nature. Due to the trend toward packet-switched technologies and the increasing use of the Internet , it has been envisioned that these different existing and emerging wireless access networks will be connected together through an IP core network to build up the next-generation all IP-based wireless networks [2]. It incorporates all the best features of the individual network into a single integrated system, thus providing ubiquitous "always best connection" (ABC) [3] to the mobile users. Figure 1.1 shows a typical architecture for next-generation all IP-based wireless networks.

Mobile nodes in such an integrated network environment have to be designed to support multiple or multi-mode wireless interfaces. With these multiple interfaces, the mobile nodes can keep the ongoing traffic session remain uninterrupted while roaming between various wireless access networks. Based on their service needs, the best available networks will be selected for the mobile users when switching between

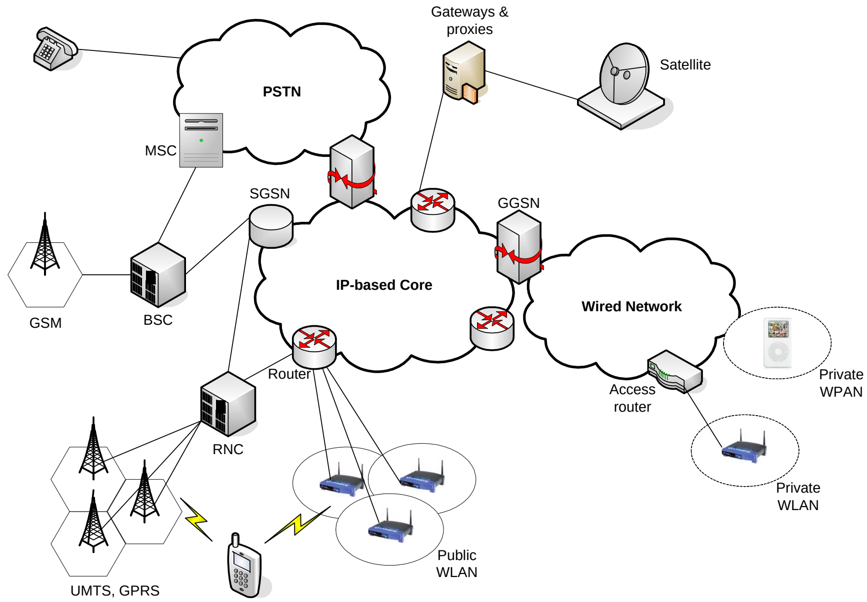

Figure 1.1: An architecture of next-generation all IP-based wireless networks.

the different networks. Apparently, seamless mobility management support plays an important role in providing continuous wireless services to mobile users in next-generation all IP-based wireless networks [4].

There are two components included in mobility management: *location management* and *handoff management* [5]. Location management is used by the network to monitor and get updated information on the location of mobile users while moved into foreign wireless networks. Whereas, handoff management aims at providing continuous connection to mobile users when moving between different wireless networks or base stations (BSs). Before further discussing handoff categorization in detail, we first take a look at the following scenarios where a handoff will happen:

- when a mobile user will soon enter into another network which covers its current serving wireless network
- when a mobile user desires to be switched to the underlying or overlaid network to have some specific servies provided
- when the system decides to have the mobile user handed off to another network in order to balance the overall traffic load among different wireless networks

In general, handoff can be categorized into horizontal handoff and vertical handoff. Horizontal handoff or intra-system handoff is defined as a handoff that occurs between the two BSs of a same network system, for example, handoff between the two cells of a UMTS system, GSM system, etc. Whereas vertical handoff or inter-system handoff is defined as handoff that occurs between two BSs belonging to two different systems, respectively. For example, handoff from a WLAN BS to UMTS BS. Therefore, vertical handoff is concerned with two cells of wireless access systems, which differ in several aspects such as bandwidth, data rate, frequency of operation, coverage, etc.

There have been much research work conducted focused on location management techniques in the literature. However, handoff management for next-generation all IP-based wireless networks is still a crucial and challenging research issue to be solved nowadays [6]. In this thesis, we focus on handoff management.

This chapter is organized as follows. First, we discuss the objective of the thesis by posing some research questions. And then we discuss the scope and methodology applied to this thesis. Finally, the outline of the thesis is presented.

1.1 Main Research Challenges

The next-generation all IP-based wireless networks aim at providing ubiquitous services to mobile users roaming between various wireless networks. Intelligent handoff management techniques are required to support the seamless roaming of the mobile users. However, it has been a challenging research issue to design such an intelligent technique due to the following requirements needed for handoff:

- Extreme low handoff latency is desired to make as less disruption to the ongoing user traffic as possible.
- Near-zero packet loss is expected during handoff so as to provide the desired quality of services.
- Handoff failure probability should be limited to a desired value, e.g., near-zero.

The existing handoff management solutions try to support mobility from different layers of the TCP/IP protocol stack reference model. These handoff management solutions can be classified into different categories according to what TCP/IP protocol layers they were proposed for, such as Mobile IP (MIP) [7] in the network layer, TCP-Migrate [8] in the transport layer, and Session Initiation Protocol (SIP)

[9] in the application layer. Mobile IP and SIP are the two handoff management protocol already standardized by the Internet Engineering Task Force (IETF).

Mobile IP was designed as a protocol for providing mobility support in IP layer. But it is required to modify the existing networking infrastructure as well as the TCP/IP protocol stacks running on Mobile Nodes (MNs). Mobile IP-based handoff solutions have significant handoff latencies [6]. The latency is comprised of handoff movement detection delay and Mobile IP registration delay. The similar problems exist for for TCP-Migrate. In order to implement TCP-Migrate in practice, all the hosts have to modify the TCP protocol running on them. SIP is mainly employed as a signaling protocol for many wireless networks. SIP-based handoff does not require any modifications to the TCP/IP protocol stack. However, the signaling messages of SIP is sent by transport layer protocols such as TCP and UDP. Therefore, the performance of SIP-based handoff solutions are subject to the performance of the transport layer protocols over wireless channels. In addition, SIP introduces additional delay because it involves the processing of application-layer packets.

Research work from link layer point of view was also proposed to support handoff management in literatures. The handoff latency and handoff failure probability can be considerably reduced by using link layer information such as received signal strength (RSS) and the speed of mobile node (MN) [6]. However, the handoff latency in these solution has been assumed to be a constant value. Unfortunately, this assumption does not hold in reality due to the fact that factors such as the traffic load in the backbone network, wireless link quality, and distance of the MN from its home network all contribute to the handoff delay [10]. As a result, these link layer assisted protocols will not be able to achieve satisfactory handoff performance.

1.2 Scope and Methodology

In this thesis, we study the vertical handoff management for integrated UMTS and WLAN networks. First, we discuss the research work proposed to support handoff management from the link-layer point of view. Different from the network-layer handoff solutions, the underlying wireless access technologies take an important role in the link-layer handoff solutions. Fundamentally, information such as RSS or MN's speed are utilized in these link-layer solutions to diminish the delay incurred in handoff movement detection. Such link-layer information helps estimating an upcoming handoff event. Therefore, the handoff can be conducted successfully before the MN moves out of the coverage area of the current serving BS.

Furthermore, we study the existing link-layer assisted handoff solutions and point out their existing deficiency. In those link-layer-assisted handoff solutions, the handoff latency is assumed to be a constant value. And typically, a fixed value of threshold is defined in advance. The handoff is initiated when the RSS of the current serving BS drops below the threshold value. This assumption does not hold in reality. Factors such as the traffic load in the backbone network, wireless link quality , and distance of the MN from its home network all contribute to the handoff delay at the moment when the handoff is initiated [10]. As a result, if the handoff signaling delay is changing from time to time, the performance of these link-layer assisted protocols will degrade significantly. Besides, the MN's speed, which has a significant effect on the performance of the handoff, was not taken into account when evaluating the performance of the link-layer assisted handoff solutions.

Based on the analysis in Section 3.1, we believe that the handoff threshold value need to be adjusted according to the MN's instantaneous moving speed and the dynamically changing handoff signaling delay. A self-adaptive threshold has to be adopted to deal with the problems. Taking that into account, we propose an adaptive vertical handoff management scheme for integrated UMTS and WLAN networks.

Our proposed scheme incorporates the idea of pre-handoff with an adaptive handoff threshold. We estimate the handoff signaling delay in advance, therefore, providing the delay information required for making an adaptive handoff decision. Instead of setting a fixed threshold, an adaptive handoff threshold value has to be determined for every single MN based on the estimated handoff signaling delay. The MN also needs to measure the RSS from the current serving BS periodically. Using the RSS and the RSS's rate of change at a specific time instant, we can estimate the remaining time duration of MN before tearing down its connection with the serving BS due to insufficient signal strength. Based on the information, an appropriate time instant to initiate a pre-handoff process for every single MN will be determined accordingly during the course of its movement.

1.3 Thesis Outline

The remainder of the the thesis is organized as follows:

In Chapter 2, the UMTS and WLAN technologies are briefly reviewed, followed by the handoff management integration architectures. Handoff techniques proposed from different TCP/IP protocol layers are also discussed.

In Chapter 3, a typical link-layer assisted handoff scheme is studied, followed by identifying its deficiency. After that, the proposed adaptive vertical handoff scheme is presented.

In Chapter 4, the simulation is conducted to evaluate the performance of the proposed adaptive vertical handoff scheme. Simulation results demonstrate that our proposed scheme significantly enhances the performance of handoff management.

In Chapter 5, the thesis is summarized, followed by the contribution of this thesis. Several future research directions are suggested for further investigation.

Chapter 2

Background

UMTS and WLAN are two different wireless access technologies. Although they differ in terms of data rate, coverage area and media access technologies, they are complementary to each other. Integration of the UMTS and WLAN networks can provide better services to mobile users. In this chapter, a review on UMTS and WLAN access technologies is conducted. After that, two integration architectures: loose coupling architecture and tight coupling architecture are discussed, which can be used for integration of UMTS and WLAN networks. The handoff techniques proposed from different TCP/IP protocol layers are also presented.

2.1 UMTS and WLAN Networks

In this section, an overview on UMTS and WLAN technologies is presented.

UMTS

Universal Mobile Telecommunication System (UMTS) is a 3G standard with an coverage area of several kilometers and high mobility support. Its low data rate ranges from 64 Kbps to 2 Mbps, which is relative low compared to WLAN. But UMTS can provide much better services and higher data rates than 2G or 2.5G cellular networks because it adopts wide band code-division multiplexing access (WCDMA) technology. NTT DoCoMo launched the world's first WCDMA network in 2001 in Japan, making the UMTS a reality. As described in 3GPP's R6 specifications [11], UMTS infrastructure consists of the core network (CN) and the access network (AN). The UMTS CN is composed of three parts: packet-switched domain (PS-domain), circuit-switched domain (CS-domain) and IP multimedia subsystem (IMS). In PS-domain, the traffic is transmitted in IP-based data packets. In the

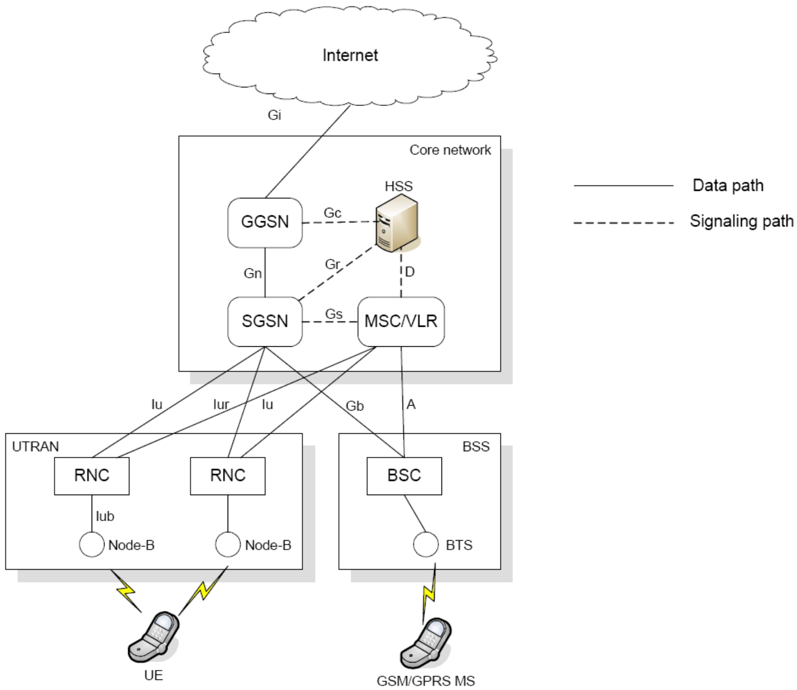

Figure 2.1: UMTS architecture.

CS-domain, the traffic is mainly traditional real-time voice calls. IMS is designed for IP-based multimedia services operated over the PS-domain.

In the CN, the PS-domain deals with the the routing of data packets between the UMTS network and the external network is performed at the serving GPRS support node (SGSN) via the gateway GPRS support node (GGSN). And the CS-domain handles the voive traffic at the mobile switching center (MSC) and the gateway mobile switching centre (GMSC). The home subscriber server (HSS), as as a repository of the users' profiles, which is shared by both domains. Different types of ANs can be concurrently connected to the CN. AN can be a base station system (BSS) or a radio network system (RNS). A BSS consists of base transceiver stations (BTSs), which is responsible for the actual radio communications, and a base station controller (BSC) that are responsible for radio resource control. RNS entities, Node-Bs and the radio network controller (RNC) provide similar functionalities. The UMTS terrestrial radio access network (UTRAN) consists of a number of RNCs and Node-Bs. Figure 2.1 depicts the UMTS architecture with its interfaces to the respective network components.

WLAN

WLAN offers limited coverage usually raniging from a few tens of meters to some hundred meters. IEEE 802.11 based wireless LAN is a local area network. It is not dedicated to offer a large coverage network like GSM or UMTS. Its coverage area can be customized according to the users need. The original IEEE 802.11 standard specified in 1997 is designed for the 2.4 GHz unlicensed band providing data rates up to 2 Mbps [12]. The IEEE 802.11b and 802.11a standards specified in 1999 can provide data rate up to 11 Mbps and 54 Mbps using the 2.4 GHz and 5GHz bands, respectively [13] and [14]. The IEEE 802.11g specified in 2003 can provide data rate up to 54Mbps in the 2.4 GHz band. The emerging IEEE 802.11n standard will provide much higher data rates in the future. Limited by some engineering constraints in the underlying radio access technology, an IEEE 802.11 Basic Service Set (BSS) covers only a few thousand square meters with no specific mobility support.

Two operation modes are defined in the IEEE 802.11 standards: infrastructure mode and ad hoc mode [12]. In an infrastructure mode network, terminals can communicate only with an AP at a given time slot. Whereas in an ad hoc mode network, mobile terminals can communicate with each other as long as there is a radio link available for them to have successful communication. In the ad hoc mode, the network is independent of any other type of networks such as GSM or infrastructure WLAN. The whole network is isolated from other networks due to the fact that no interface to the wired network exists. Typically, infrastructure mode networks are used for integration with UMTS. Authentication and association procedure have to be conducted before a mobile station (MS) can actually have an access to a 802.11 WLAN. A Dynamic Host Configuration Protocol (DHCP) server is also required to configure MS's IP stack. Typically, a MS in WLAN could be a laptop computer or a PDA with a built-in WLAN module or a PCMCIA card. The 802.11 standard only defines the MAC layer and physical layer, and hence the authentication procedures, QoS, and mobility management mechanisms are not always the same for different service providers.

2.2 Integration Architectures

In this section, different coupling architectures for integrating UMTS and WLAN networks are discussed in details.

UMTS provides wide area coverage with high mobility and relatively low data rates. However, many data intensive applications require even higher data rates to run on the MN. WLAN supports higher data rates with a limited coverage areas and limited mobility. A multi-mode UMTS/WLAN terminal can take advantages

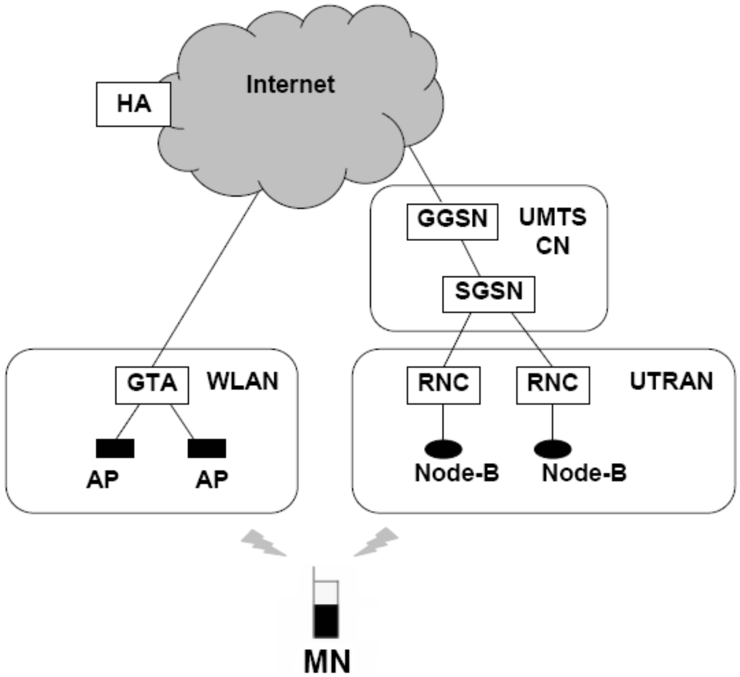

Figure 2.2: Loose coupling architecture.

of the two networks. It can access high bandwidth data services when it is within the WLAN coverage area, otherwise, it can access UMTS while moving out of WLAN coverage area. However, an integrated solution is needed to provide seamless mobility support for the MN so that the ongoing traffic session can be kept uninterrupted while moving between theses wireless networks.

The integration architectures between UMTS and 802.11-based WLANs networks can be broadly divided into two categories: loose- and tight-coupling solutions. We describe the architecture in detail in the following two sub-sections. The difference between these solutions lies in the level of integration between the two networks.

Loose Coupling Architecture

Figure 2.2 shows a typical loose coupling architecture where WLAN and UMTS are not connected directly, instead they are connected via the Internet indirectly. Therefore, the data packets will be routed through the Internet, bypassing the UMTS core network which is the case in tight coupling architecture. Loose coupling architecture also belongs to master/slave architectures. Here, the UMTS network is the master and the 802.11 WLAN is the slave network.

In loose coupling architectures, UMTS and WLAN networks are set up and configured into different network domains with different IP addresses. Therefore, the

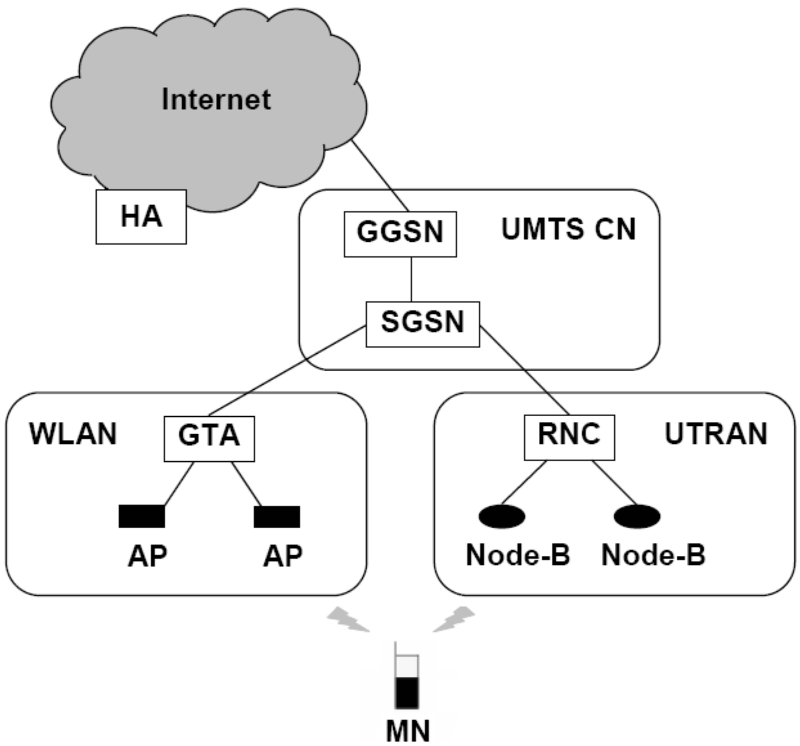

Figure 2.3: Tight coupling architecture.

MN has to request for a new IP address whenever it moves into WLAN from UMTS or vice versa. Since WLAN is connected to UMTS indirectly as a complementary network, the GGSN in UMTS network does not need to set up a tunnel and forward to the data packets to GSN, thus reducing the traffic overhead. UMTS operator can set up its own WLAN or utilized the one built by an independent operator. In addition, the WLAN data traffic never goes through the UMTS core network because the data paths in WLAN and UMTS networks are completely separated. Taking that into account, the WLAN can be set up and configured independent of UMTS networks, saving a lot of efforts to modify the UMTS network architecture as in the case of tight coupling.

Tight Coupling Architecture

Figure 2.3 depicts a tight coupling architecture. In tight coupling architecture, typically, WLAN can be connected to the UMTS network as a Radio Network Controller (RNC) or an SGSN. Similar to the loose coupling architecture, UMTS operator can set up its own WLAN or utilized the one built by an independent operator. The SGSN can support a number of WLAN connections to it. The IP address of WLAN users and UMTS RAN users will be assigned to the same network domains if they are connected to the same Gateway GPRS Support Node

(GGSN). Therefore, when there is a handoff between WLAN and UMTS, the IP address of the MN does not need to be changed. But the UMTS CN will suffer from additional traffic overhead because the signaling traffic and data traffic of WLAN have to go through the UMTS CN. During the handoff from WLAN to UMTS, those undelivered packets to the MN will be forwarded to the new SGSN. THe home subscriber server (HSS) in the UMTS CN works as a repository of the location of the MNs. If the WLAN is deployed by another independent operator instead of the UMTS operator itself, the Gp interface will be used to connect 802.11 WLAN to the UMTS UMTS CN.

There are some drawbacks with tight coupling. Directly connecting the WLAN to the UMTS core network will result in additional system cost and design complexity because the load of UMTS core networks are designed to support cellular users only. Therefore, loose coupling architecture has been widely employed to integrate UMTS and WLAN networks.

2.3 Handoff Techniques

In this section, the reported handoff management schemes operating from the different layers of the TCP/IP protocol stack are presented.

Mobile IP (MIP) [15] is a IP layer mobility management protocol because it was proposed to support mobility management from IP layer point of view. With mobile IP, the MN can remain connected to the Internet while moving to other foreign networks. The mobile IP architecture is shown in Figure 2.4.

There are three main mechanisms introduced in mobile IP: *agent discovery, registration, and tunneling.* When a MN moves from one network to another, mobility agents such as the home agent (HA) and foreign agent (FA) are involved. The MN learns about the reachable networks by listening to the agent advertisement messages sent periodically by the mobility agents. An MN may also send an agent solicitation message and receive agent advertisement message from from any locally attached MAs. The MN will be able to recognize what network domains it has moved into. If the MN found itself in a a foreign network, it will obtains a *care-of address* (CoA) from the foreign network. The CoA information is actually included in the agent advertisement message of a FA. The MN can acquire its CoA from the agent advertisement message. Alternatively, the MN can obtain its CoA by Dynamic Host Configuration Protocol (DHCP). After obtaining a CoA, the MN registers its CoA with its HA in order to obtain service such as receiving packets

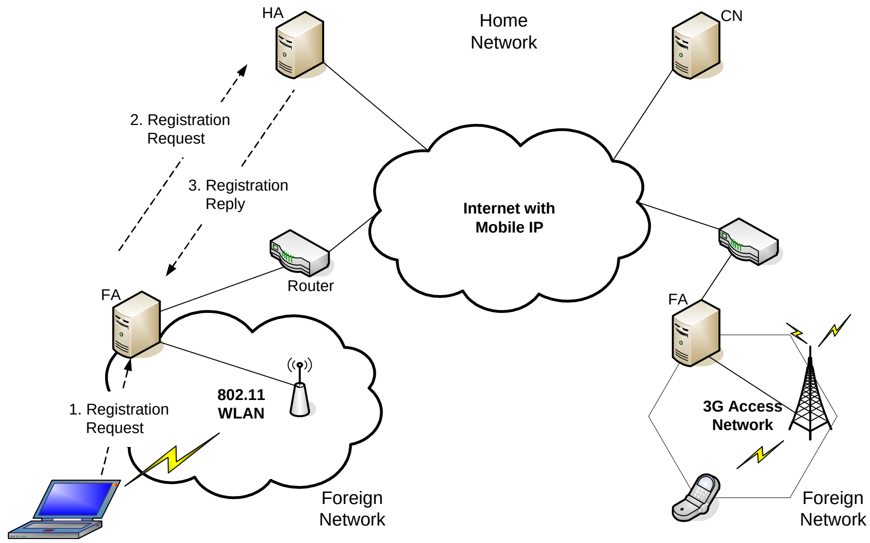

Figure 2.4: Mobile IP network architecture.

for the ongoing session with other correspondence nodes. The MN can register its CoA directly to its HA or ask the FA to forward its request to the HA.

The MN needs a CoA to receive the packets forwarded by its HA when it is currently associated with other FAs. Mobile IP uses a mechanism called tunneling to have the packets forwarded from the MN's HA to its FA and finally to the MN. Packets forwarded from the HA have to be encapsulated and the MN's CoA is added to the encapsulated data as a header. The HA intercepts the packets sent to the MNs home address and tunnels them to the MNs CoA. After receiving the packets, the FA forward them the MN at its current location. The whole procedure is called *triangular routing*. Once the MN receives the first packets at its new location, it will sent the subsequent packets directly to the correspondence nodes, thus the traffic will not traverse the HA any more. Apparently, there is a considerably large delay incurred in triangular routing. This problem can be solved by *route optimization*. The basic idea is that the CN will get informed by the HA about the MNs current CoA. The traffic between MN and the CN will be sent directly to MN, bypassing the HA. In this way, the delay incurred in triangular routing will be diminished. It helps reducing the delay, especially if the MN and CN belong to the same network.

It is common for the MNs to switch between different sub-networks of one domain. Many micro-mobility management solutions have been proposed to reduce the signaling load and delay to the home network when moving between different sub-networks of one domain. These mobility management schemes can be broadly classified into two groups: tunnel-based and routing-based schemes.

Tunnel-based schemes try to reduce the global signaling load and handoff latency by adopting local or hierarchical registration and encapsulation concepts to limit the scope of the signaling messages. Hierarchical Mobile IP (HMIP) [16] and Mobile IP regional registration (MIP-RR) [17] are two examples of tunnel-based micro-mobility protocols. Whereas, routing-based schemes uses the routers to keeps the host-specific routes to forward packets. The host-specific routes are updated when the MN changes its location. Cellular IP (CIP) [18] and handoff-aware wireless access Internet infrastructure (HAWAII) [19] belong to routing-based micro-mobility protocols.

Hierarchical Mobile IP [16] is proposed to extend the Mobile IP to reduce the global signaling delay. In Hierarchical Mobile IP, the MN only perform a single home registration with its HA after obtaining a CoA from the FA. After that, the MN will only perform Regional Registrations with a gateway foreign agent(GFA). These registrations are sent by the mobile to the GFA whenever it changes a new FA (i.e. of WIPPOA). The MN sends a regional registration whcih contains the information about the new local CoA which is used by GFA to reach the MN. The routing in the Hierarchical Mobile IP is operated in a very straight forward way. The HA first intercepts the packets destined to the MN and tunnels them to the GFA. The packets received by the GFA will be de-capsulated and tunneled to the current local CoA of the MN. A multi-level hierarchy of FAs can also be set up in Hierarchical Mobile IP. In this case, every FA in the hierarchy has to keep a binding list for every visiting MN. These binding are updated by the regular registration requests from the MNs. The regional registrations from a MN are only forwarded to the first FA with a binding for the MN already. The details of the MN's movements are out of the interests to the upper levels of the hierarchy since their binding do not need to be updated. During this process, only a few nodes are involved, thus reducing the global registration and signaling delay.

Cellular IP [18] is proposed to support local mobility but efficiently interworks with Mobile IP to provide wide area mobility support. In Cellular IP, each network domain contains multiple MAs. One of the MAs acts as a gateway towards the Internet and a FA for macro-mobility handoff. A routing cache is maintained by every MA for every MN. It includes the information such as the next hop to reach a MN or a gateway. The MN forward packets from the gateway to the MN or vice versa based on the routhing cache. The routing cache is updated when the MA receives the hop-by-hop transmission of two special control packets. Cellular IP adopts hard handoff mechanism for mobility management. Once the radio handoff to set up a new link is finished, the MN will send a route update message to the gateway. A so-called semi-soft handoff was defined to improve the drawbacks caused by hard handoff. The packet loss can be reduced if the MN sends a special packet to the old FA informing its attachment to the new FA. A delay device has to be implemented to reduce a potential synchronization problem which might occure between the two FA. In this way, the packets sent by the new FA during a semi-soft handoff can be postponed.

HAWAII (handoff-aware wireless access Internet infrastructure) [19], is a domain-based mobility management protocol. Similar to Cellular IP, host-based forwarding entries are kept in routing caches of specific routers to support intra-domain micro-mobility. The routing cache get updated by the packets sent by hop-by-hop in the network. The network addresses of the MNs are not changed while they move within the same network domain. Same as Cellular IP, the network topology of HAWAII is organized in a hierarchical architecture with a root gateway seated on top. Paging is employed in HAWAII to support passive connectivity. An IP multi-cast group consists of a number of MNs which belong to the same paging area. The multi-cast group receive the paging requests sent to the corresponding paging area.

All above mentioned micro-mobility management solutions operate from the IP layer and try to improve the Mobile IP signaling delay and registration delay, but with different design methodologies employed. The common practice of these solutions are to localize Mobile IP signaling traffic into one domain to reduce global signaling [20].

In [21], two fast handoff protocols based on bicasting and non-bicasting supporting real-time traffic, such as VoIP, are proposed to reduce the packet loss rate during handoff, but without increasing bandwidth. Typically, fast-handoff protocol try to reduce the packet loss by forwarding the packets to the MN's new FA when the IP Layer handoff is anticipated. However, packet loss will occur if the handoff is performed at a right time. The reported solution bicast packets destined to the MN to one or more potential future FA before the MN actually moves there. The packets will be sent to the MN's current FA and the MN's new FA for a short time. Therefore, there is no need to decide precisely when to start sending data packets to the MN's new FA. It also resolves the problem of of service disruption brought by the ping-pong effect. However, in order to perform the bicast, two wireless links and IP addresses have to be assigned each MN, which result in sending the duplicate packets and consuming excessive wireless bandwidth

In [22], a TCP-based handoff scheme is proposed to improve the handoff performance for hybrid wireless networks which integrate different types of wireless mobile networks. These wireless mobile networks a have a lot of difference in terms of capacity data rate, capacity, coverage area, etc. For example, WLAN and UMTS networks have different bandwidths. When a MN moves into a low data rate wireless network, say, from WLAN to UMTS, its data rate will drop down immediately. It will result in a longer handoff latency due to the authentication process involved. Taking that into account, packet losses over the wireless channel during the handoff is not disclosed. By doing so, these error information will not be taken as an indication of network congestion status by the sender, therefore, no further actions will be taken accordingly to decrease its data rate. In the TCP-based approach, the data transmission and timeout timer will be stopped temporarily in order to to avoid a timeout and packet losses during handoff procedure. Once the handoff is completed, the data transmission will be restored from the TCP slow start state. Apparently, the goal of this approach is to promptly stop the TCP communication session during handoff procedure and restore it from the TCP slow start state once the handoff is completed.

The TCP receiver sends an ACK to the sender as soon as handoff is completed. The ACk message is used to notify the sender to restore the data transmission before a timeout. Based on the received ACK message, the TCP sender will adjust its congestion window size in accordance with the HO option field. If it indicates a horizontal handoff, the timeout timer in the TCP sender will be stopped immediately and the data transmission will be suspended promptly as well. After handoff procedure is completed, the data transmission will be restored with the TCP sender at the CA state with the same congestion window size as before. If the HO option field indicates a vertical handoff, the timeout timer will be stopped immediately with the TCP sender and the data transmission will be put on hold. Once the handoff is completed, the data transmission will be restored at the SS state with the TCP sender.

In [23], a TCP-based handoff approach is proposed to improve the handoff performance. This handoff approach can be applied to the cases where there is a overlapped coverage area from two or even more base stations. And the base stations to which the MN can be handed off are already determined in advance. In another case where there is no overlapped coverage area, the handoff procedure is similar the previous solution. The data transmission during handoff has to be stopped until the new point of attachment is identified so as to achieve satisfied TCP performance. The common practice is to set the TCP receiving window size to zero which can stop the TCP data transmission. Once a new point of attachment is identified, the TCP receiving window size will be set to a non-zero value and advertised which will result in restoring of the TCP data transmission. After the handoff, the TCP fast retransmission can be triggered by sending the ACKs to restore any active TCP data transmission.

In [24] and [25], downward and upward vertical handoffs in integrated wireless LAN and cellular networks are considered to address the problem of Wireless Profiled TCP premature timeouts and false fast retransmission. Wireless Profiled TCP premature timeouts is caused by a dramatic increase of the round-trip time and the false fast retransmission is resulted from packet reordering. To be more specific, a mobile receiver centered loose-coupling cross-layer design is developed. The advantages of this scheme are easy to implement and backward compatible with the Wireless Application Protocol v2 (WAP 2.0), and robust in the situation without the cross-layer information. Two proactive schemes called RTT Inflation and RTT Equalization are proposed. In the RTT Inflation scheme, the retransmission timeout time is carefully adjusted to decrease the influences by the premature timeouts. Whereas, in the RTT Equalization scheme, the false fast retransmission is avoided by equalizing the round-trip delay experienced by all the packets.

SIP-based handoff [26] solution is a typical application layer handoff management protocol, which aims at improving handoff performance from application-layer point of view. SIP is a text-based protocol offering extensibility and the session control provisioning. There are three main components in SIP: user agents, proxy servers and redirect servers. An email address is assigned to each user as its unique username. For example, bob@uwaterloo.ca, where bob is the username and uwaterloo.ca is the domain name. There are two types of terminal mobilities in SIP, pre-call mobility or mid-call mobility. During pre-call mobility, SIP sets up a connection for a new session when the MN moves into a different location. Whereas, mid-call mobility requires SIP to establish a new connection during the middle of a session. Mid-call mobility is similar to a vertical handoff conducted during an ongoing session. After moving into a new network, the MN sends a SIP INVITE message to ask the correspondent nodes (CNs) to join the call, whish uses the same call identifiers as in the original call setup. The new IP address will be included in the SIP INVITE message to inform the new location of the MN. An acknowledge message will be sent to the MN after the CN receive all the updated address information about the MN. And then, the data transmission process will be started. The limitations of SIP-based handoff solution is that it involves the processing of SIP messages in the intermediate and destination servers, which takes a a great deal of time. Therefore, it causes additional handoff delays which is undesired. Besides, in order to implement the SIP-based handoff solution, applications running on all the hosts need to be modified.

The handoff latency for Mobile IP handoff latency is composed of latencies for handoff movement detection and Mobile IP registration. The hierarchical Mobile IP and other micro-mobility solutions particularly achieve reduction in registration signaling delay, but fail to address the problem of handoff movement detection delay [27].

Nowadays, there have been some research work proposed to improve the handoff performance by reducing the movement detection delay. These handoff management solutions operate from the link-layer point of view. These link-layer handoff management solutions rely heavily on the underlying radio access systems. The basic idea is that these solutions use the link-layer information such as RSS or MN's speed to reduce the handoff movement detection delay. With these link-layer information, they can anticipate the possibility of a upcoming handoff in advance so that the handoff procedures can be carried out successfully before the MN moves out of the coverage area of the serving base station (BS).

In [28], a link-layer assisted handoff solution called a two-step vertical handoff scheme is proposed to improve the handoff signaling by pre-handoff mechanisms. The scheme has two steps: pre-handoff and handoff, and intends to manipulate the two physical quantities derived from sensed signal strength: estimated handoff time and location of the MN. Pre-handoff is initiated when MN's estimated time to handoff goes below fixed threshold value. The drawback of this scheme is that it is difficult to implement in reality since it assumes the pre-defined threshold value as constant.

Similar approach using received signal strength (RSS) information to reduce handoff latency and handoff failure probability has been proposed in [29] and [27]. In these link-layer assisted handoff solutions, the handoff delay are implicitly considered as a constant value. Based on this assumption, these solutions initiate a handoff when the RSS of the serving BS drops below a predefined fixed threshold value. However, in a real scenario, signaling delay of the handoff process depends on the traffic load in the backbone network, wireless link quality, and distance between a user and its home network at the handoff instance.

2.4 Summary

UMTS and WLAN are considered as two different wireless access networks in terms of the data rate, coverage area and media access technologies. It is believed that integration of the two networks can provide better services to mobile users. Depending the deployment requirement, the UMTS and WLAN can be integrated using either a loose coupling architecture or a tight coupling architecture. The difference between the two architectures lies in the level of integration between the two networks. In loose coupling architecture, WLAN and UMTS are connected via the Internet. The WLAN data traffic never go through the UMTS core network. However, in the tight coupling architecture, WLAN is connected directly to the UMTS core network, which results in additional system cost and design complexity. Integration of UMTS and WLAN calls for intelligent handoff management techniques to provide seamlessly roaming services for mobile users. Mobility management protocols operating from different layers of the TCP/IP protocol stack, e.g., link layer, network layer, transport layer, and application layer are developed to support mobility in next generation wireless systems. In next the chapter, an adaptive vertical handoff management is proposed to solve the problem in this situation.

Chapter 3

Adaptive Vertical Handoff Scheme

In this chapter, an existing link layer assisted handoff scheme is studied and its handoff performance is analyzed with respect to users' speed and handoff signaling delay, followed by identifying its deficiency in the situations where the handoff signaling delay varies. Based on the insight from this analysis, an adaptive vertical handoff management scheme is proposed. The operation and the implementation of the proposed scheme are presented in details.

3.1 Analysis of Link-layer Assisted Handoff

In this section, a typical link-layer assisted handoff is studied, followed by identifying its deficiency. Based on that, the necessity of applying an adaptive handoff threshold is pointed out to improve vertical handoff performance.

However, the handoff delay in the link-layer assisted handoff management solutions is implicitly considered as a constant value. With this assumption, these protocols initiate a handoff when the RSS of the serving BS drops below a predefined fixed threshold value. However, in a real scenario, signaling delay of the handoff process depends on the traffic load in the backbone network, wireless link quality, and distance between a user and its home network at the handoff instance. Therefore, the protocols that are designed assuming a fixed signaling delay for handoffs have poor performance when the handoff signaling delay varies. Moreover, the existing link-layer-assisted handoff protocols do not consider the influence of user's speed on the performance of the handoff protocols. User's speed has a significant effect on the performance of the handoff protocols. The performance analysis of a typical link-layer assisted handoff management is conducted as follows.

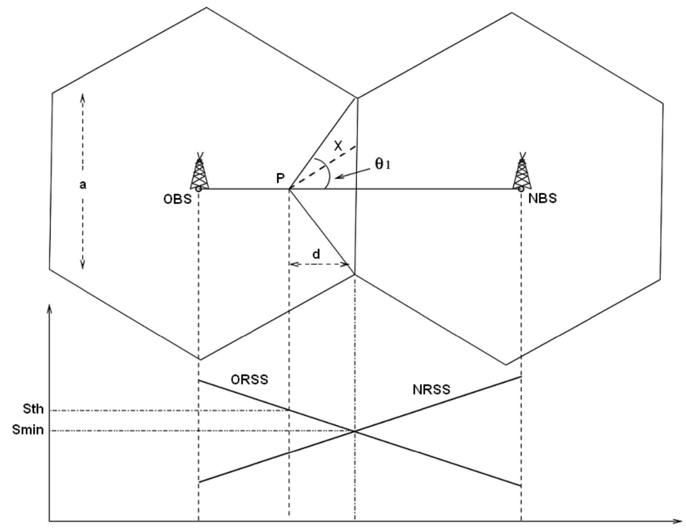

Figure 3.1: A typical handoff process

Figure 3.1 [10] depicts a scenario of a typical handoff process where a MN is moving from its current serving BS (old BS), to the future serving BS (new BS). The followings are the definition of the notations in the figure:

- S_{th}: The RSS threshold value to initiate the handoff.
- S_{min}: The MN's minimum RSS value to communicate successfully with BS.
- OBS and NBS: The old BS and the new one one respectively.
- ORSS and NRSS: The RSS of old BS and the new one respectively.
- a: The cell size served by a BS.
- P: The point when the MN's RSS from the OBS doprs below S_{th}.
- d: The distance from the cell boundary to the point P.
- θ: The motion direction of MN from point P to handed off to NBS.

Since the probability of the MN moving in any direction at point P is the same, the MN will be handed over to NBS if its motion direction from point P is $\theta \in (-\theta_1, \theta_1)$, where $\theta_1 = arctan(\frac{a}{2d})$, otherwise, it runs into a false handoff meaning that the MN will out of the coverage area of the NBS after moving out of the one of OBS.

The handoff failure probability of the MN is given by [10]:

$$p_f = \begin{cases} 1 & \tau > \frac{\sqrt{\frac{a^2}{4}+d^2}}{v}, \\ \frac{1}{\theta_1} arccos(\frac{d}{vt}) & \frac{d}{v} < \tau < \frac{\sqrt{\frac{a^2}{4}+d^2}}{v}, \\ 0 & \tau \leq \frac{d}{v}. \end{cases} \tag{3.1}$$

where τ is the handoff signaling delay. In the remaining part of this section, we present the relevant discussion on handoff performance using the above formulations.

Handoff Failure Probability and User's Speed

Based on equation (3.1), the relationship between the handoff failure probability p_f and the MN's speed v can be illustrated by Figure 3.2 [10] when $\frac{d}{v} < \tau < \frac{\sqrt{\frac{a^2}{4}+d^2}}{v}$. The curves in the figure corresponds to the cases when different threshold values s_{th}, hence different corresponding values of d, are applied. As we can see, for a fixed value of d, the handoff failure probability p_f increases as the MN's speed increases. This is easy to understand because it takes less time for MN to move out of the coverage area of OBS, hence, higher failure probability to be handed over to NBS. In order to obtain a desired handoff failure probability, say 0.1, the value of d, hence the threshold S_{th}, has to be modified accordingly for each and every MN moving in different speeds.

Handoff Failure Probability and Signaling Delay

As we stated before, since the network traffic condition changes dynamically from time to time, core network traffic, wireless channel quality and the MN's distance from HA, are all factors leading to the handoff latency. Based on equation (3.1), the relationship between the handoff failure probability p_f and the handoff signaling delay can be showed in Figure 3.3 [10]. From this figure, we can understand that for a particular threshold value S_{th}, hence a corresponding value of d, the handoff failire probability p_f increases along with the increase of the signaling delay τ. In consequence, with an effor to achieve a desired handoff failure probability, say 0.1, in the value of d, hence the threshold S_{th}, should be adjusted in accordance with the dynamically changing value of handoff signaling delay τ.

As a result, we could draw a conclusion that in order to maintain a particular value of handoff failure probability the handoff threshold value need to be adjusted in accordance with the changing value of the MN's moving speed as well as the dynamically changing handoff signaling delay. Therefore, the application of a self-adaptive threshold is necessary to deal with the problem in this situation. Taking

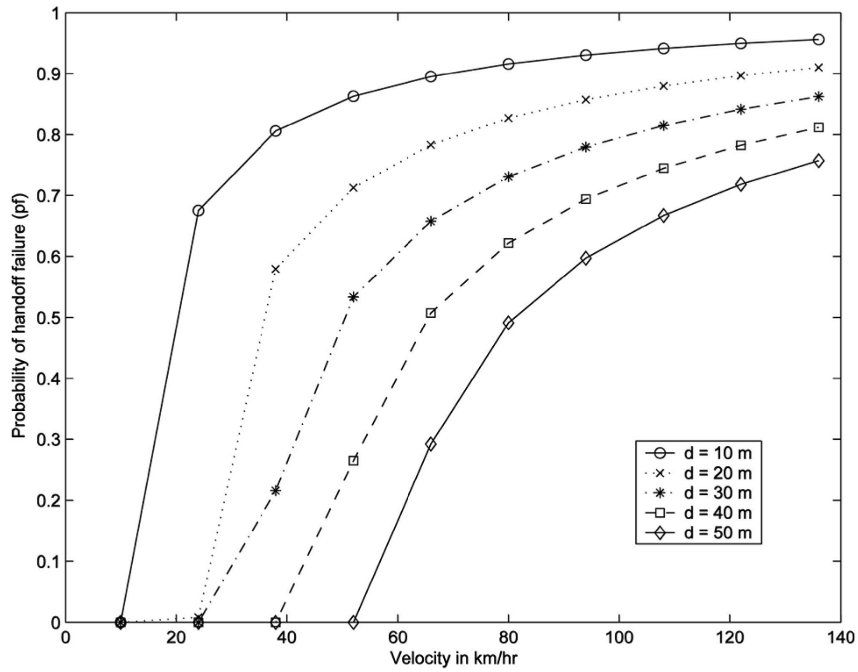

Figure 3.2: Relationship between handoff failure probability and user's speed

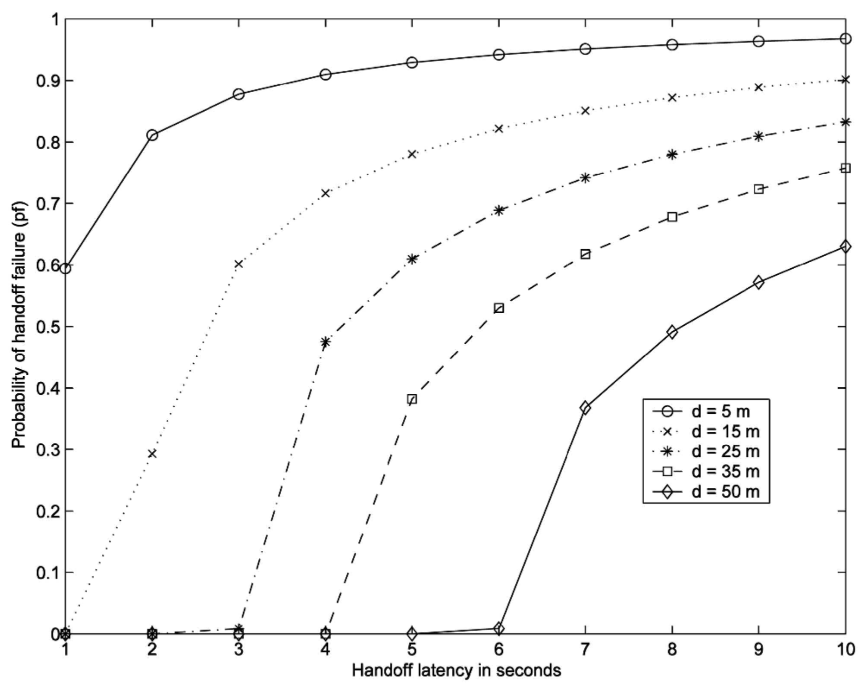

Figure 3.3: Relationship between handoff failure probability and handoff signaling delay

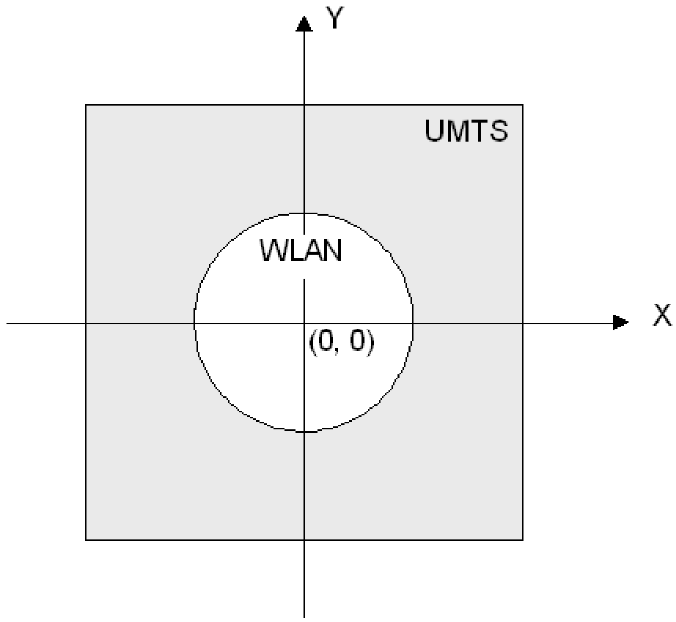

Figure 3.4: The networking environment adopted in the proposed scheme

that into account, in the next section, we propose an adaptive vertical handoff management scheme for integrated UMTS and WLAN networks.

3.2 Adaptive Vertical Handoff Scheme

In this section, the module architecture of the proposed scheme is described, followed by the elaboration of the functionalities of each modules. Our proposed adaptive vertical handoff management scheme is based on the figure as depicted in Figure 3.4. We assume that the networking environment is overlapped and the coverage area of WLAN is surrounded by circle centered around point (0, 0) and the one of of UMTS network by a square geographically covering WLAN. We also assume that the MN is a wireless multimode device capable of supporting multiple wireless access networks by incorporating several network interface cards and the appropriate software for switching between these network interfaces.

The architecture of the handoff management is typically composed of several functional modules. Depending on the handoff management chosen, either it is a mobile-assisted network-controlled handoff (MAHO) or network-assisted mobile-controlled handoff (NAHO). Accordingly, the functional modules can be imple-

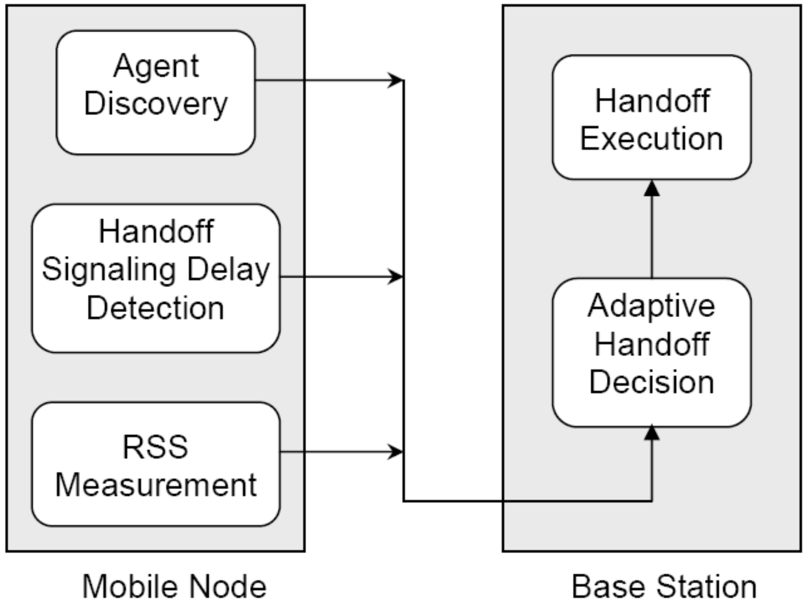

Figure 3.5: Architecture of the proposed handoff scheme

mented on the MN or at the network side. Figure 3.5 shows the diagram of the proposed handoff management module architecture. It consists of multiple modules, each with different functionalities. Some of the modules are responsible for gathering information from link layer and network layer such as RSS and signaling delay respectively. Whereas, other modules use the information achieved to make an adaptive handoff decision to initiate handoff process at the right time and execute the handoff process accordingly. The modules that collect information from link layer and network layer are agent discovery unit, handoff signaling delay detection unit and RRS measurement unit. The modules that use the collected information from link layer and network layer to make the handoff decision and carry out the handoff procedures are adaptive handoff decision unit and handoff execution unit. The following are the brief description of the modules:

- Agent Discovery unit, as the name indicated, learns about the reachable networks for the MN to carry out a possible handoff. The MN learns about the presence of a reachable wireless network by listening to its *ServiceAdvertisement* messages broadcast periodically or actively sending an *AgentSolicitation* message to request an advertisement message.
- Handoff Signaling Delay Detection unit esimates the signaling delay for a possible handoff procedure in advance. This signaling delay information will

be used to set an adaptive handoff threshold for the MNs. The MN can estimate handoff signaling delay by obtaining the round-trip time (RTT) of an invalid handoff registration message from its HA.

- RSS Measurement unit obtains the RSS from the serving BS by a periodical detection when a potential handoff is foreseen. RSS and its rage of change will be used to estimate the MN's moving speed, and more importantly, the remaining time to keep connected with the current serving BS before tearing down its connection due to insufficient signal strength.

- Adaptive Handoff Decision unit makes a decision to initiate the handoff process at the right time based on the information provided by the previous units. An adaptive handoff threshold will be determined for every single MN based on the estimated handoff signaling delay. The handoff process will be initiated once the estimated time duration to handoff goes below the adaptive handoff threshold.

- Handoff Execution unit will start the standard Mobile IP registration process as soon as the pre-handoff initiation message is passed to Handoff Execution unit from Adaptive Handoff Decision unit. MN obtains a new CoA and register with its HA.

3.2.1 Operation of the Proposed Scheme

In this subsection, the detailed description of the proposed adaptive vertical handoff scheme is conducted by presenting the functionalities of each module.

Agent Discovery

The MN uses its agent discovery unit to search for reachable wireless neighbor networks for a possible handoff. These networks should be served by a BS whose geographical coverage area is either overlapped or immediately bordering with the one of the current serving BS. Our system model given earlier falls into the first case. The MN will be aware of the presence of a reachable wireless network by either listening to its *ServiceAdvertisement* messages broadcast periodically or actively send an *AgentSolicitation* message to request an advertisement. The simplest way to discover reachable wireless networks is to keep all network interfaces on. However, when it is attached to the WLAN, the MN with multiple interfaces does not need to activate all the interfaces to receive service advertisements. This is because, in most cases, the MNs prefer to stay in WLAN until they have to move out of its coverage area for whatever reasons. The MN will achieve the details such as IP address of the reachable network, which is UMTS in our study, after the agent discovery process.

Handoff Signaling Delay Detection

The handoff signaling delay is what it takes for the MN to register with its home network while it tries to attache to a new network. It could be significantly large in the case that the MN is far away from its home network. The approach to estimate the handoff signaling delay proposed in [10] is adopted. The basic idea is to use the handoff signaling delay detection unit to estimate the handoff signaling delay in advance, therefore, providing the delay information to the adaptive handoff decision unit to make a decision on handoff initiation at a right time.

Once the agent discovery step is completed, the MN is aware of the future serving BS and its corresponding FA for a possible handoff. And this handoff signaling delay has to be estimated in advance. In order to do that, the MN send out a handoff registration message to its HA with an invalid *Mobile-Home Authentication Extension*. A valid handoff registration message sent by MN is a request for its attachment to a new FA and the corresponding update with the HA. After receiving the invalid handoff registration message with an invalid *Mobile-Home Authentication Extension*, the HA will not proceed with the mobility binding update for the MN. Instead, the HA simply ignores the message and replies to the MN by sending a *Registration Reply* message with an error code indicating an unsuccessful authentication. The MN can estimate the handoff signaling delay, referred to as $D_{signaling}$ by calculating the time difference from sending the invalid handoff registration message to getting the registration reply with an error code from the HA. All these have to be done by the handoff signaling delay detection unit. The handoff signaling delay $D_{signaling}$ can be acquired by the MN in the course of its movement to the BS. It directly reflects the wireless channel quality, backbone Internet congestion status and the distance of the MN from its home network. With these information obtained in advance, the MN can make an adaptive handoff decision to compensate the delay happening on the upcoming handoff process, therefore, improving the system performance. In the next subsection, we present how this information is used to facilitate the proposed adaptive vertical handoff scheme.

RSS Measurement

The functionalities of the RSS measurement unit is to obtain the received signal strength from the serving BS by a periodical detection when a potential handoff is foreseen. In [30] and [31], based on RSS and its rage of change, the moving

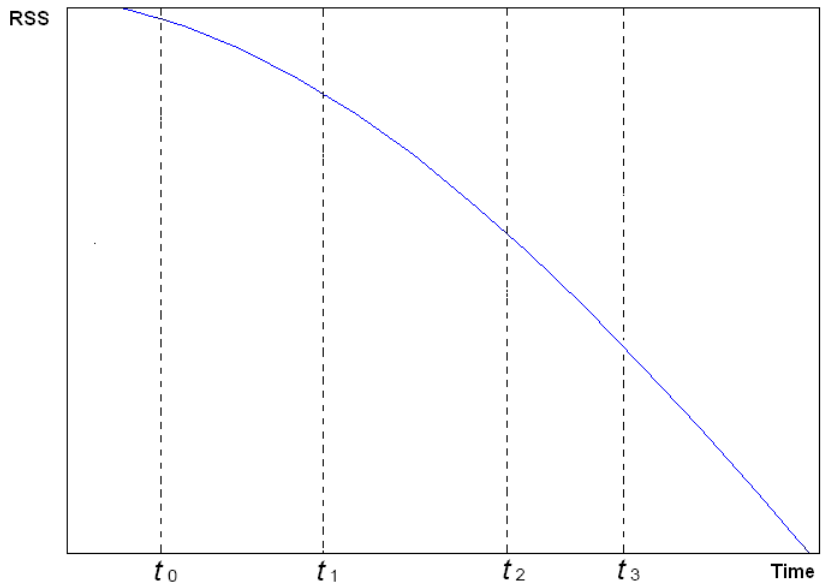

Figure 3.6: Measurement of the RSS and its rate of change

speed of MNs' is estimated, which defines the queuing priority of handover calls. Similarly, in our work, we utilized RSS and its rage of change to estimate the MN's moving speed and, more importantly, the remaining time to keep connected with the current serving BS before tearing down its connection due to insufficient signal strength. Here, we describe the basic idea as follows.

Let time be slotted into intervals, and let RSS be the measurement made during time interval $\Delta t = t_2 - t_1$ as depicted in Figure 3.6. The RSSs rate of change at time t_2 is calculated as the change of RSS over the time interval it occurred. Therefore, based on the RSS and the RSSs rate of change values at time t_2, one can estimate the remaining time duration of MN before tearing down its connection with the serving BS as follows:

$$T_{Est.Handoff}|_{t_2} = \frac{RSS}{\Delta RSS}|_{t_2} \qquad (3.2)$$

$T_{Est.Handoff}$ is the estimated time to handoff, which is the remaining time duration to keep connected with the serving BS with a sufficient signal strength. RSS_{t_2} and ΔRSS_{t_2} are the RSS and its rate of change measured at time t_2, respectively. Intuitively, equation (3.2) uses the RSS and its rage of change, which is the MN's moving speed to predict the time instant when the RSS from the current serving BS goes to zero. The estimated time to handoff along with the estimated handoff signaling delay will be used by the adaptive handoff decision unit to make an decision on handoff initiation at an appropriate time instant.

Adaptive Handoff Decision

When the RSS is not strong enough to guarantee the ongoing communication, the MN has to tear down its connection with the current serving BS. Therefore, the MN has to set up its connection with a future BS for its ongoing session. *Vertical handoff delay* is defined as the time interval from tearing down the connection with the current serving BS to setting up a new connection with a future BS. The whole procedure for setting up a connection with a new BS involves a number of lengthy steps such as agent discovery, registration with HA, authentication and authorization, etc. The performance of handoff can be considerably improved if some of the handoff steps can be conducted pro-actively, i.e., while the MN still senses a strong WLAN signal strength, which is also called pre-handoff [28]. With pre-handoff, the MN can reduce or eliminate the vertical handoff delay by initiating a new connection with the new BS in advance, hence, trying finish the handoff process before its connection with current BS get torn down.

Our proposed scheme incorporates the idea of pre-handoff with adaptive handoff threshold. From the above two modules, we can acquire the estimated handoff signaling delay $D_{signaling}$ and RSS along with its rate of change, hence the MS's estimated time to handoff $T_{est.handoff}$. With these information, the adaptive handoff decision unit makes a decision on pre-handoff initiation at an appropriate time for every single MN in the course of its movement. Instead of setting a fixed threshold, an adaptive handoff threshold value has to be determined for every single MN based on the estimated handoff signaling delay. To be more specific, the adaptive handoff threshold value, referred to as $T_{adpt.threshold}$ is set as the estimated handoff signaling delay $D_{signaling}$. And the pre-handoff process is initiated once $T_{est.handoff}$, the estimated time duration to to handoff, goes below $T_{adpt.threshold}$, the adaptive pre-handoff threshold. It then triggers the pre-handoff execution process.

Handoff Execution

When the pre-handoff initiation messages from the adaptive handoff decision unit is passed to the handoff execution unit, the standard Mobile IP process will be triggered from this point on. It obtains a new care-of address (CoA). This CoA can be obtained by soliciting or listening for FA advertisements (an FA CoA), or contacting Dynamic Host Configuration Protocol (DHCP). The MN registers the new CoA with its HA. The HA sets up a new tunnel up to the end point of the new CoA and removes the tunnel to the old CoA. Then the HA updates the mobility binding by associating the CoA of the MN with its permanent IP address. Once the new tunnel is set up, the HA tunnels packets destined to the MN using the MN's new CoA.

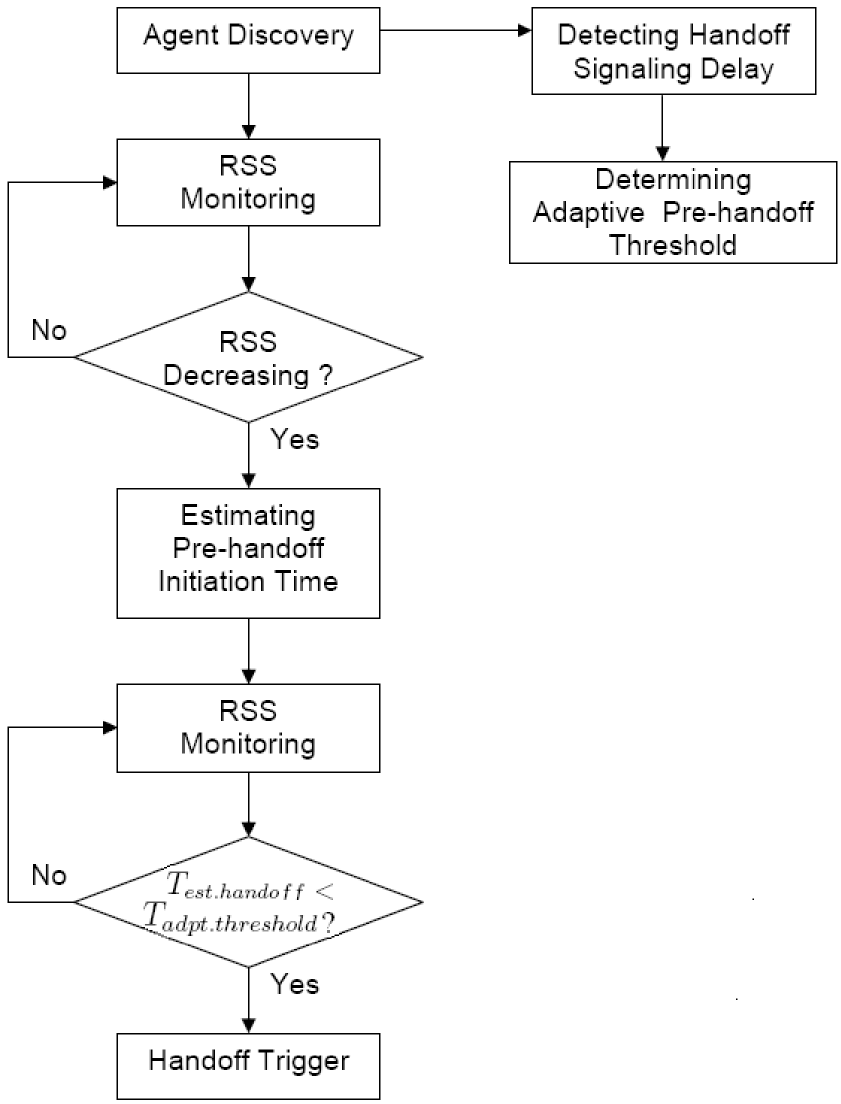

Figure 3.7: Flow chart diagram of the the proposed handoff scheme

We summarize the operation of the proposed adaptive vertical handoff scheme with reference to the flow chart diagram as shown in Figure 3.7.

1. The *Agent Discovery* unit start searching its neighbor networks using a agent discovery protocol.

2. When the neighboring BSs are learned, the associated handoff signaling delay will be estimated by the *Handoff Signaling Delay Detection* unit.

3. An adaptive pre-handoff threshold will be determined by the estimated handoff signaling delay.

4. The RSS of the current serving BS will be monitored by the *RSS Measurement* unit and a pre-handoff will be anticipated if the RSS drops continuously.

5. Based on the previous two steps, an appropriate time to initiate the pre-handoff will be estimated by *Adaptive Handoff Decision* unit.

6. The RSS will be kept on monitored by the *RSS Measurement* unit.

7. The pre-handoff process will be triggered once the estimated time to handoff drops below the adaptive pre-handoff threshold.

8. Pre-handoff process will be carried out by the *Handoff Execution* unit.

3.2.2 Implementation of the Proposed Scheme

In this subsection, the implementation of the proposed adaptive vertical handoff scheme is discussed.

Generally speaking, handoff management can be categorized into network-assisted mobile-controlled handoff (NAHO) or mobile-assisted network-controlled handoff (MAHO) depending on how the functional modules are implemented on the MN and the BS.

If it is a NAHO, the MN take over most of the tasks involved in the handoff. During the agent discovery step, the MN tries to find the reachable networks and selects the future BS with the assistance of the network. The MN estimates the handoff signaling delay by calculating the round-trip time of sending an invalid handoff registration message to the HA. Based on the signaling delay, an adaptive handoff threshold will be set accordingly. RSS Measurement unit obtains the RSS from the serving BS by a periodical detection when a potential handoff is foreseen. RSS and its rage of change can be used to estimate the remaining time to keep connected with the current serving BS. As soon as the estimated time duration to handoff goes below the adaptive handoff threshold, the handoff will be triggered by the MN and initiated by sending an *Proxy Router Solicitation* message to the new FA. However, in this case, since most of the functional modules are implemented on the MN, there will be more processing overhead introduced to the MN, that has limited processing power and speed.

Whereas, in the case of of MAHO, the functional modules like Agent Discovery unit, Handoff Signaling Delay Detection unit and RSS Measurement unit are implemented on the MN as depicted in the Figure 3.5. Accordingly, Adaptive Handoff Decision unit and Handoff Execution unit can be implemented on the BS. Similar to NAHO, the MN learns about the reachable networks by listening to their *Service Advertisement* messages broadcast periodically or sending an *Agent Solicitation* message to request an advertisement. However, different from the NAHO, the MN in the case of MAHO is only responsible collecting relevant information such as handoff signaling delay and RSS with its rate of change. The collected information will be provided to the BS. The BS will set an adaptive threshold according the handoff signaling delay. Also, the BS calculates the remaining time to have the MN kept connected with itself based on the RSS and its rage of change provided by the MN. When the estimated time duration to handoff goes below the adaptive handoff threshold, the network generate the handoff trigger for handoff. Then the network initiates the handoff procedures by sending *Proxy Router Advertisement* message to the MN.

Depending on whether the MN gets an *Agent Advertisement* without or after sending an *Agent Solicitation* message, the handoff can be initiated by either the network or the MN, respectively. Figure 3.8 illustrates the timing diagrams of the proposed handoff scheme where a mobile-assisted network-controlled handoff is considered.

As depicted in the timing diagram, when the pre-handoff requirement is met, the pre-handoff process will be initiated after the current serving BS (old FA) receives the *handoff trigger* message (HO Trigger). It then sends a *Proxy Router Advertisement* (ProxyRtAdv) message to the MN. After receiving the *Proxy Router Advertisement* message, the MN sends a Mobile IP *regional registration request* (Reg. Req.) message to the future FA (new FA). The new FA forwards the message with appropriate extensions to the HA of the MN. After processing the Mobile IP *regional registration request* message, the HA sends a *authentication (auth.) request*

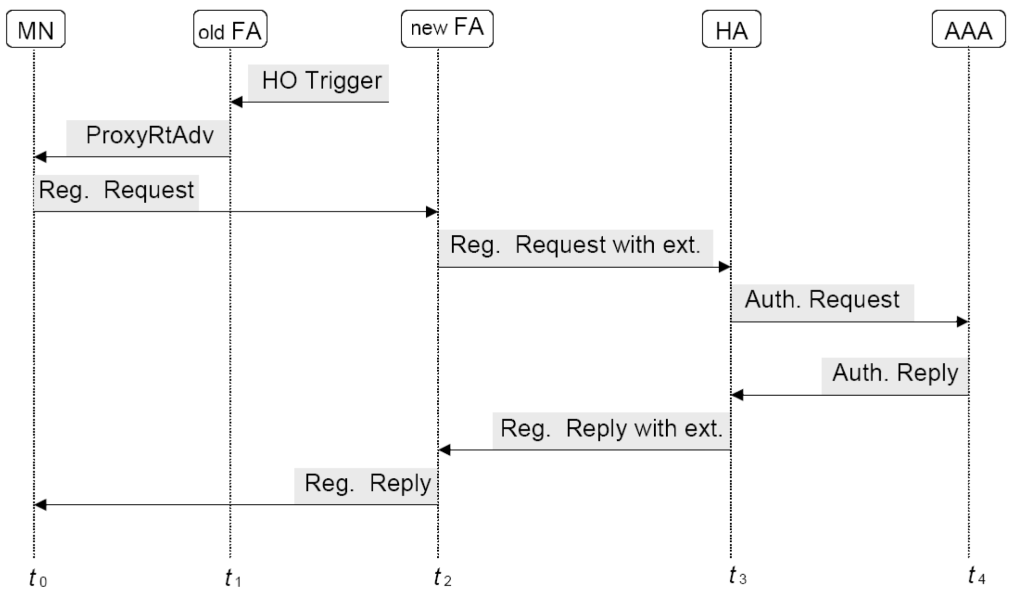

Figure 3.8: Timing diagram for the proposed handoff scheme

message to AAA server. The *authentication reply* will be send to the HA after the MN's information is verified by AAA server. The HA sends a Mobile IP *regional registration reply* message to the new FA, which forwards it to the MN. When the registration is successfully completed, the packets directed to the MN are tunneled from the HA to the new FA to which the MN has attached.

3.3 Summary

In this chapter, the performance of a link layer assisted handoff scheme is studied, followed by pointing out its deficiency in situations with a dynamically changing handoff signaling delay. Furthermore, it is observed that the handoff threshold value need to be adjusted based on the handoff signaling delay so as to meet a particular handoff failure probability. Therefore, an adaptive threshold should be applied to deal with the problem. Based on the insights, an adaptive vertical handoff management scheme is proposed, which incorporates the ideas of pre-handoff and adaptive handoff threshold together. In the proposed schem, the MN estimates the handoff signaling delay of a possible handoff in advance and adjust the handoff threshold accordingly. The RSS and its rate of change are measured to estimate the remaining time to keep connected with the current serving BS. The pre-handoff process will be initiate as soon as the estimated time duration to handoff goes below the adaptive pre-handoff threshold. The implementation of the proposed scheme is also discussed. In the next chapter, the the simulation results will be presented to evaluate the performance of the proposed handoff scheme.

Chapter 4

Simulation Results

In this chapter, two metrics: vertical handoff delay vs. multi-tunnel time are introduced to evaluate the handoff performance of the proposed scheme, followed by presenting the the the simulation model. Extensive simulation are carried out to evaluate the performance of the proposed handoff scheme.

4.1 Vertical Handoff Delay vs Multi-tunnel Time

In this section, two metrics related to handoff performance are discussed.

As we have discussed earlier in Chapter **??**, our proposed scheme incorporates the idea of pre-handoff mechanism, which is a technique used to eliminate or diminish the MN's vertical handoff delay by initiating the handoff process with the new FA in advance while the connection with the old FA still exists.

We first define two metrics used in the simulation: *vertical handoff delay* and *multi-tunnel time*. Figure 4.1 shows the vertical handoff processes in the case with and without pre-handoff mechanism applied. The standard Mobile IP process is shown in Figure 4.1 (a); a pre-handoff process with non-zero handoff delay is shown in Figure 4.1 (b) and a pre-handoff process with non-zero multi-tunnel time is shown in Figure 4.1 (c). We define the following notation with reference to Figure 4.1:

- t_0: The time instant when the signal of the new network is detected by MN.
- t_{disc}: The time instant when MN tears down its connection with old FA.
- t_{setup}: The time instant when MN start setting up its connection with new FA.

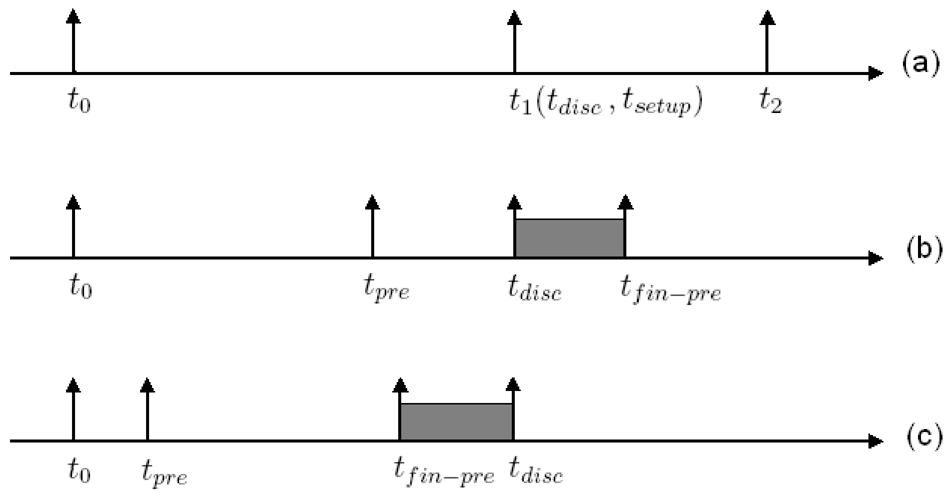

Figure 4.1: Comparison of the vertical handoff delays. (a) The standard MIP Process; (b) a pre-handoff process with non-zero handoff delay; (c) a pre-handoff process with non-zero multi-tunnel time.

- t_1: The time instant represent both t_{disc} and t_{setup}.
- t_2: The time instance when MN start receiving data from via new FA.
- t_{pre}: the time instant when MN initiate a pre-handoff process with new FA.
- $t_{fin-pre}$: The time instant when MN starts receiving data from new FA.

Applying the previous definition of vertical handoff delay to the cases as shown in Figure 4.1, we observe that the vertical handoff delay for the standard MIP is $D_{MIP} = t_2 - t_1$ for the case in Figure 4.1 (a). Whereas the vertical handoff delay and multi-tunnel time are calculated using the following formula, respectively:

$$D_{vertical-handoff} = \begin{cases} t_{fin-pre} - t_{disc} & t_{fin-pre} > t_{disc}, \\ 0 & t_{fin-pre} < t_{disc}. \end{cases} \quad (4.1)$$

$$T_{multi-tunnel} = \begin{cases} 0 & t_{fin-pre} > t_{disc}, \\ t_{disc} - t_{fin-pre} & t_{fin-pre} < t_{disc}. \end{cases} \quad (4.2)$$

If $t_{fin-pre} > t_{disc}$ as shown in Figure 4.1 (b), it means that the MN tears down its old link before start receiving data from the new FA. In this case, the vertical handoff delay is $D_{pre-handoff} = t_{fin-pre} - t_{disc}$, a non-zero variable which is the time duration highlighted by grey band. But we also noticed a zero multi-tunnel time for this case.

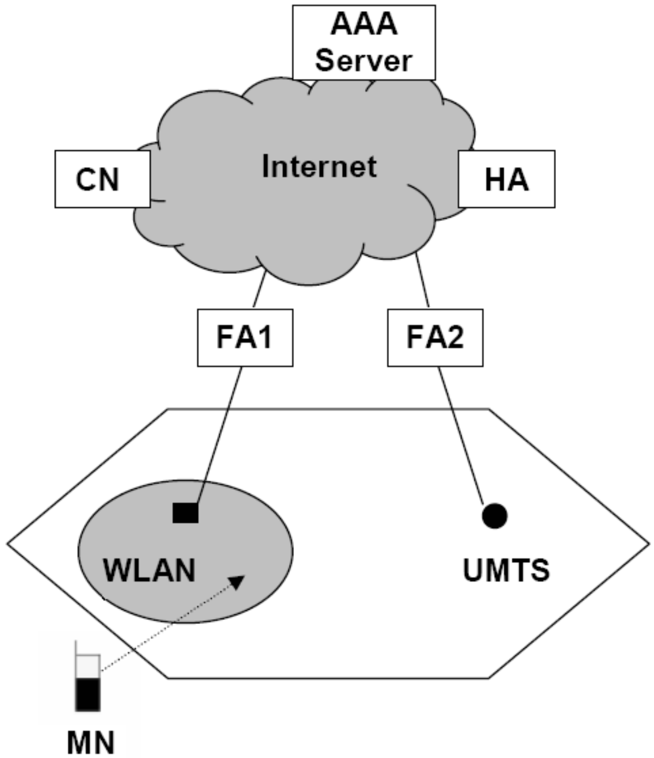

Figure 4.2: A simple simulation model.

However, if $t_{fin-pre} < t_{disc}$ as shown in Figure 4.1 (c), it means that the MN start receiving data from the new FA before its connection with the old FA is torn down. In this case, there is a zero vertical handoff delay and nonzero multi-tunnel time $T_m util - tunnel = t_{disc} - t_{fin-pre}$. The multi-tunnel time duration is highlighted by grey band as well.

The handoff delay $|t_{fin-pre} - t_{disc}|$ and the multi-tunnel time $t_{disc} - t_{fin-pre}$ serve as two key performance metrics in the vertical handoff process with the pre-handoff mechanism. Obviously, the handoff delay affect the service continuity, while the multi-tunnel time represents the redundancy paid for the handoff process. Therefore, a handoff algorithm needs to be designed to make the $|t_{fin-pre} - t_{disc}|$ as small as possible.

4.2 Simulation Model

In this section, the simulation model of the proposed scheme is presented. We evaluate the time duration from when the handoff trigger message is received to when the the registration to the HA successfully completely. After that, the MN can send and receive IP packets through the new FA.

For simplicity, we consider the model illustrated in Figure 4.2. Before building up the handoff delay models for Mobile IPv4 and the proposed handoff schemes, we define the following notations:

- D_{MN-FA}: The one-way delay between the MN and the FA, which is the time to send a message via a wireless link.
- D_{MN-HA}: The one-way delay between the MN and its home network (HA.
- D_{FA-HA}: The one-way delay between the FA and the HA.
- $D_{oldFA-newFA}$: The one-way delay between the old FA and the new FA.
- D_{HA-AAA}: The one-way delay between the old FA and the new FA.

In the standard Mobile IPv4 scheme, the overall handoff delay is composed of the following delays incurred at each step during the MN's registration with its HA. The MN detects a new network by exchanging the agent solicitation and the agent advertisement messages, which totally take $2D_{MN-FA}$. The MN sends an MIP registration request the new FA, which takes D_{MN-FA} totally. The new FA forwards this registration message to the MN's HA taking a time of D_{FA-HA}. When it receives this registration request, the HA update the MN's CoA. The round trip time for a authentication request from HA to AAA server takes $2D_{HA-AAA}$. After that, the HA send a registration reply message to the new FA, again, the process takes a D_{FA-HA}. The new FA sends this registration reply message to the MN taking a time of D_{MN-FA}. At this point of time, the MN finishes the registration with its HA. The overall delay incurred during these process, therefore, is:

$$D_{MIP} = 4T_{MN-FA} + 2T_{FA-HA} + 2T_{HA-AAA} \tag{4.3}$$

Similarly, in the proposed handoff scheme, the overall handoff delay is modeled as follows. Upon reception of the HO-trigger, the old FA sends ProxyRtAdv to the MN, which takes D_{MN-FA}. Then the MN sends a registration request to the new FA via the old FA, which takes $D_{MN-FA}+D_{oldFA-newFA}$. The new FA forwards it to the HA, which takes D_{FA-HA}. The round trip time for a authentication request from HA to AAA server takes $2D_{HA-AAA}$. The registration reply for the HA is routed through the to FA, which takes D_{FA-HA}. The new FA sends this registration reply message to the MN taking a time of D_{MN-FA}. As a result, we can get the overall delay for the proposed adaptive vertical handoff scheme as follows:

$$D_{adaptive} = 3D_{MN-FA} + D_{oldFA-newFA} + 2D_{FA-HA} + 2D_{HA-AAA} \tag{4.4}$$

As we know, D_{MN-FA} is the delay incurred over the wireless link between MN and FA and $D_{oldFA-newFA}$ is the delay generated via the wired link. Generally D_{MN-FA} is much larger than $D_{oldFA-newFA}$. Therefore, comparing equation (4.3) and (4.4), we find that D_{MIP}, the MIP handoff delay, is relatively larger than $D_{Addaptive}$, the handoff delay in the proposed handoff scheme.

4.3 Simulation Results

In this section, the simulation results are presented to evaluate the performance of the proposed handoff scheme.

The objective of our simulation is to make a comparison between the standard Mobile IPv4, and the reported two-step vertical handoff scheme [28] and the proposed adaptive handoff scheme in terms of handoff delay, multi-tunnel time, etc. In our simulation, the following assumption are considered for the delays:

$D_{MN-FA} = 50ms + \tau_1$
$D_{FA-HA} = 50ms + \tau_2$
$D_{HA-AAA} = 150ms + \tau_2$
$D_{oldFA-newFA} = 20ms + \tau_4$

where τ_1, τ_2, τ_3 and τ_4 are exponential r.v. with mean of 20ms, 20ms, 50ms and 10ms, respectively. The delays includes time required for propagation, queuing, processing etc.

The MN's speed and location information are detected every 100 ms. Moreover, we assume that the MN's current location and speed are dependent on the ones in the previous state. The following equations are used to calculate the MN's new direction and speed at the beginning of each time slot:

$$\theta(t + \Delta(t)) = \theta(t) + \Delta(\theta) \tag{4.5}$$

$$v(t + \Delta(t)) = v(t) + \Delta(v) \tag{4.6}$$

where $\Delta(\theta)$ and $\Delta(v)$ are two random variables representing the supplementary angle and speed between the current and the previous state. We assume that $\Delta(v)$ is a discrete random variable with the following possibilities: 0, 0.3 m/s, 0.6 m/s, 0.9 m/s, 1.2 m/s, 1.5 m/s, 1.8 m/s and 2.1 m/s. Moreover we assume that $\Delta(\theta)$ is considered as a discrete random variable with a probabilitt mass function (PMF) as shown in the following table:

$\Delta(\theta)$	0	$\pm\pi/4$	$\pm\pi/2$	$\pm 3\pi/4$	π
$Pr(\Delta(\theta))$	β_1	β_2	β_3	β_4	β_5

apparently, $\beta_1 + 2\beta_2 + 2\beta_3 + 2\beta_4 + \beta_5 = 1$. Here, β_1 is the probability that the MN keeps its moving direction as the previous time slot. Obviously, with a larger β_1, the MN is more likely to move in the same direction as that in the previous moment.

Furthermore, we could obtain the MN's position at every time slot using the following equations:

$$x(t + \Delta(t)) = x(t) + x(t) * \Delta(t) * cos(\theta(t))) \tag{4.7}$$

$$y(t + \Delta(t)) = y(t) + y(t) * \Delta(t) * cos(\theta(t))) \tag{4.8}$$

Therefore, the current distance from the MN to the point (0, 0) can be obtained as follows:

$$d(t) = \sqrt{x(t)^2 + y(t)^2} \tag{4.9}$$

Based on these assumptions, extensive simulations are conducted to demonstrate the performance of the proposed handoff scheme.

We first show the simulation results of the two-step vertical handoff scheme. Figure 4.3 depicts the simulation results of the average vertical handoff delay and multi-tunnel time with respect to the threshold value. On one hand, as we can see, the average vertical handoff delay is significantly reduced as the threshold is increased. On the other hand, the multi-tunnel time is considerably increased as well. Intuitively, both the results are a direct consequence of initiating the pre-handoff earlier. The decreases handoff delay is obtained by investing extra multi-tunnel time, hence more non-used resources occupied by MNs. For the scheme, however, a fixed threshold has to be defined which makes it difficult to implement in reality.

Relationship between Handoff Delay and Handoff Signaling Delay

For the proposed adaptive vertical handoff scheme, we analyze the relationship between the average vertical handoff delay and multi-tunnel time with respect to the handoff signaling delay, thus the adaptive threshold. We first estimate the handoff signaling delay for every single MN. Then, based on that, we set a adaptive

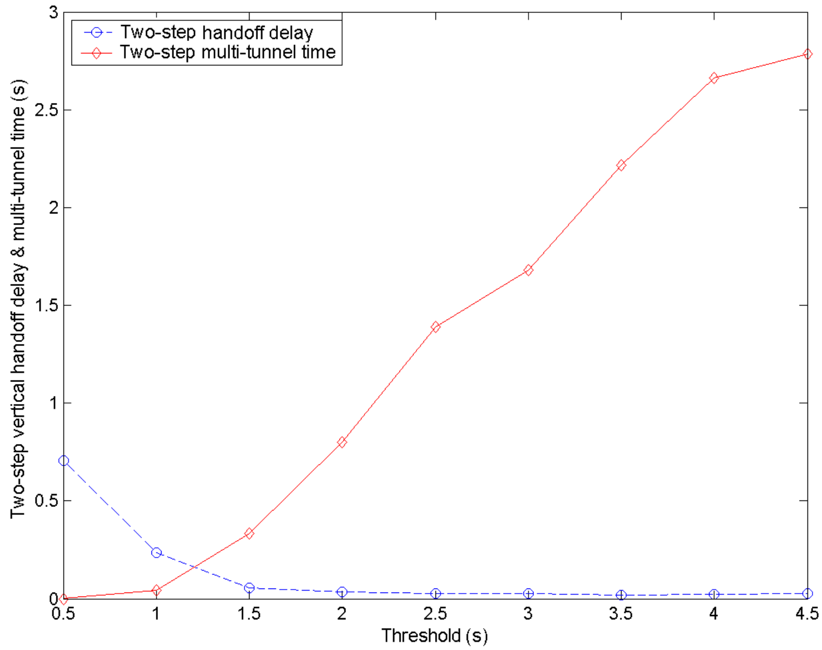

Figure 4.3: Relationship between two-step vertical handoff delay vs threshold.

threshold accordingly. Figure 4.4 illustrates the simulation results of the adaptive vertical handoff scheme. Although the handoff signaling delay is increased due to dynamically changing network conditions, the adaptive vertical handoff delay is staying around a value of 0.16 second. The similar conclusions hold for the the multi-tunnel time against the handoff signaling delay. The reason behind is that the proposed scheme is self-adaptive to the dynamically changing network conditions, therefore being able to adjust its threshold accordingly. A slight variation in the vertical handoff delay is introduced because of the error in speed and handoff signaling delay estimation.

Ratio of Number of Pre-handoff Events and Number of Handoff Events We study the ratio of the pre-handoff events and handoff events. Generally, with a larger threshold, the probability for the MN to enter the pre-handoff stage is getting larger. The MN is more subject to a false pre-handoff with unnecessary multi-tunnel resource consumption. Therefore, it is of interest to observe how often pre-handoff events are initiated and how many of them have led to a real handoff. Fig. 10 shows the ratio of the number of pre-handoff events to the number of handoff events taken in the simulation regarding different values of the signaling delay. However, we can see that the ration is almost a constant number with respect to the increasing signaling delay, thus the increased adaptive handoff threshold. Moreover,

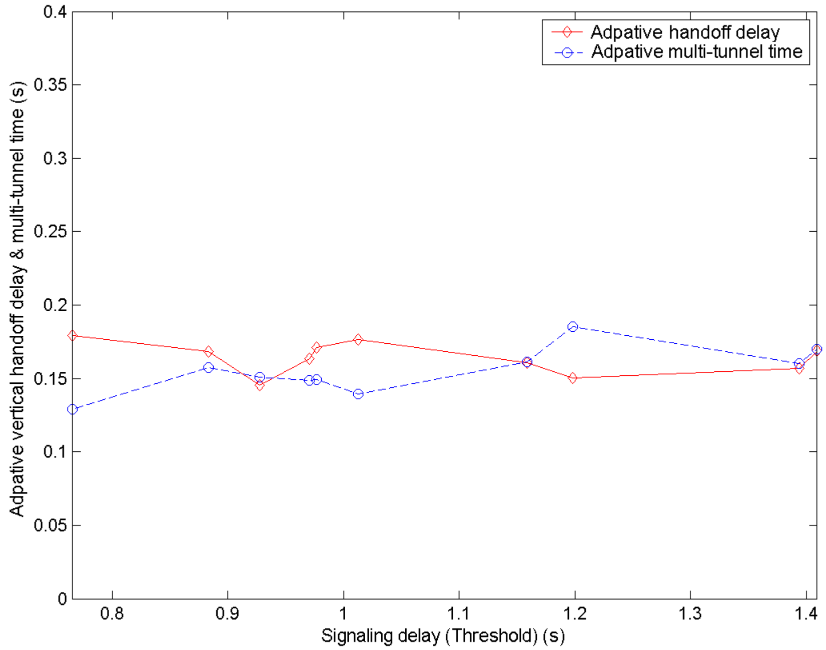

Figure 4.4: Relationship between adaptive vertical handoff delay and vs signaling delay.

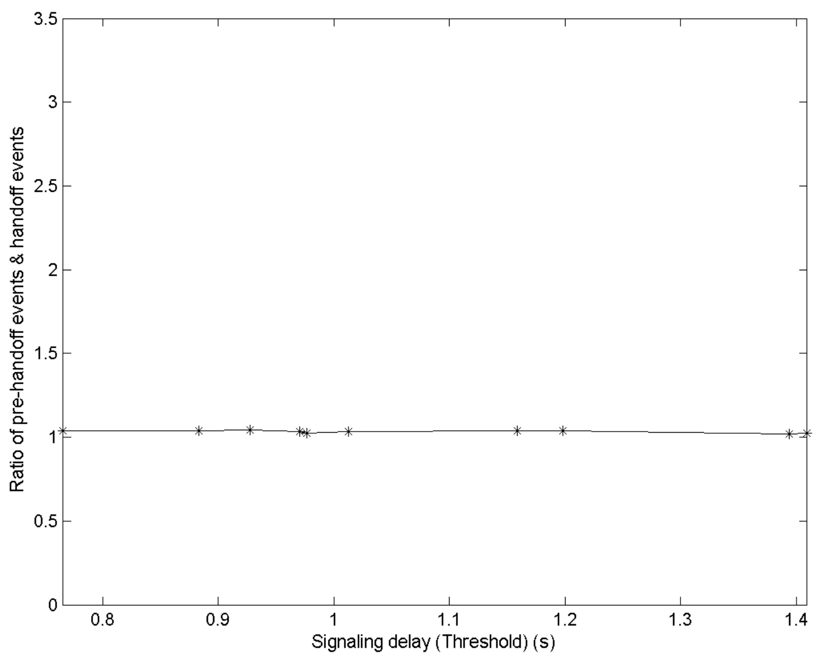

Figure 4.5: Ratio of number of pre-handoff events and number of handoff events.

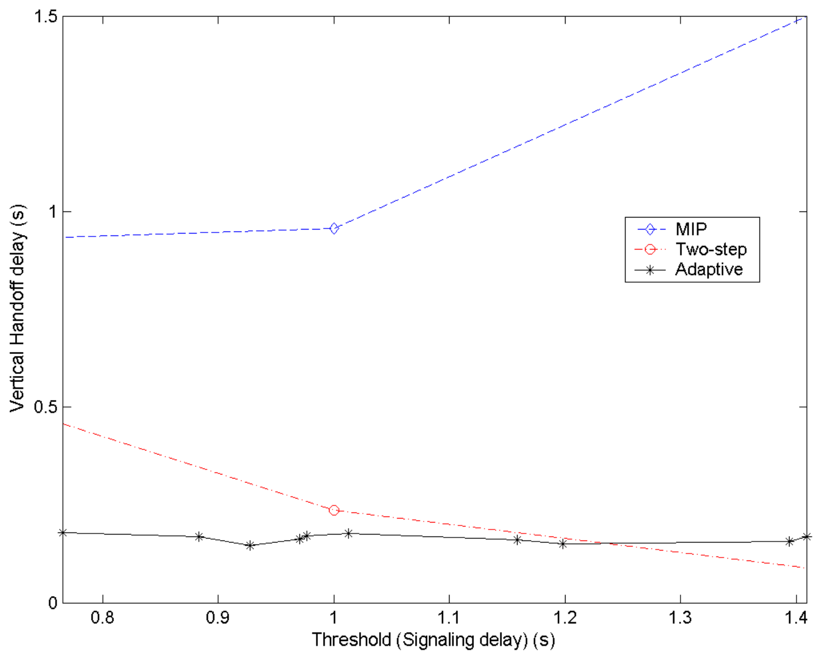

Figure 4.6: Relationship between handoff delay and threshold on previous schemes.

the adaptive handoff scheme has a higher reliability such that the majority of the initiated pre-handoff events successfully lead to handoff events. It shows that the proposed scheme has the self-adaptability to make a relative more precise decision to initiate handoffs even if the signaling delay is dynamically changing.

Relationship between Handoff Delay and Threshold against the Previously Reported Schemes

Figure 4.6 depicts the relationship between the vertical handoff delay and threshold for the proposed scheme against the standard MIPv4 and the two-step vertical handoff scheme. As we can see, the proposed adaptive handoff scheme has a very stable handoff delay with a slight variation when the threshold is increased. It is mainly because the proposed scheme first detect the network signaling delay in advance. Taking that into account, the proposed scheme can make a relatively more precise decision to initiate the handoffs at the right time. The two-step handoff scheme has advantages of fairly low handoff delay when the threshold values get very large. But it fails to demonstrate a better performance for the small values of threshold. The MIP shows the worst performance.

Relationship between Handoff Delay and Threshold with Different Movement Patterns.

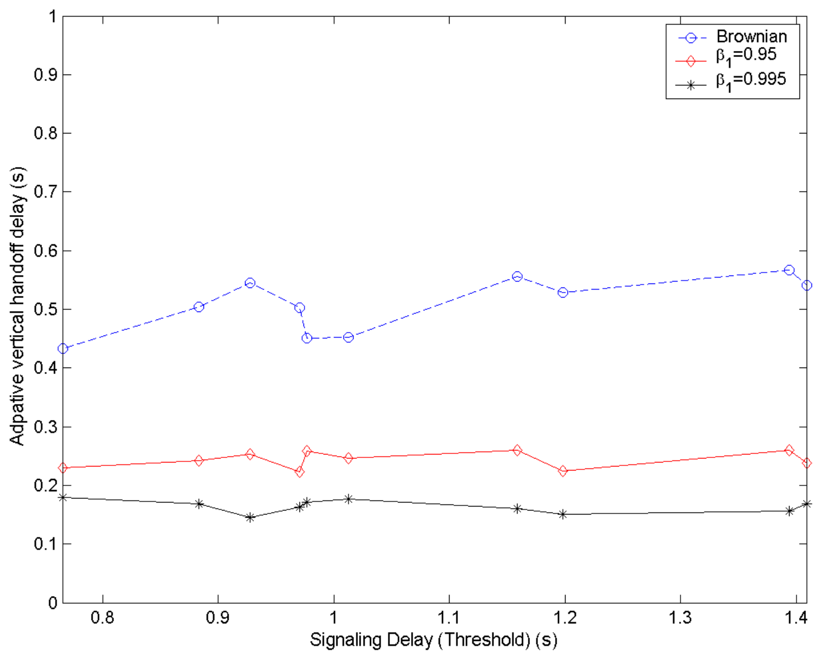

Figure 4.7: Relationship between handoff delay and threshold with different movement patterns.

Figure 4.7 shows the effect of the increased β_1 on the relationship between the average vertical handoff delay and the value of threshold. For comparison, we introduce the Brownian mobility model where the current direction of an MN is independent of that in the previous time slots. As we can see, the moving patterns of the MN have a significant influence of the vertical handoff delay. In the case of Brownian movement pattern, the MN's movement is not randomly changing and independent of the previous one, thus making the adaptive handoff scheme less effective and resulting in larger handoff delay. Whereas, the performance in the cases where $\beta_1 = 0.95$ and $\beta_1 = 0.995$, has gained significant increase partly because the MN's movement pattern in this two cases is much easier to predict and thus the assigned adaptive handoff threshold can lead to better performance. Figure 4.8 shows the impact of the varying moving patterns on the relationship between the average multi-tunnel time and threshold where the multi-tunnel time is getting larger but still stays stable when the user movement is more Brownian. And Figure 4.9 depicts the ratio between the number of pre-handoff events and the number of handoff events versus different values of threshold with regards to different moving patterns.

By observing Figure 4.7, 4.8 and 4.9, we argue that the the proposed adaptive vertical handoff scheme can achieve a relatively very stable handoff delay and multi-

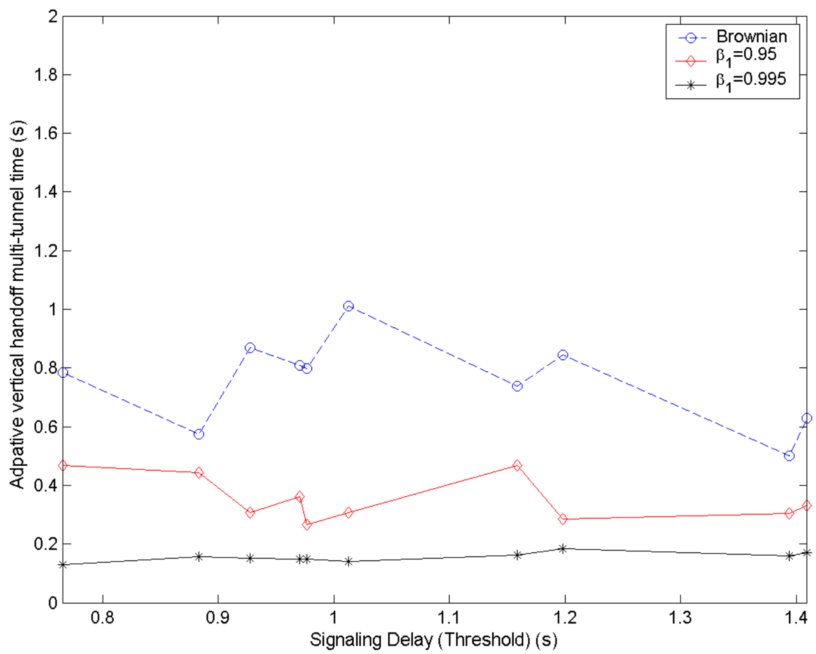

Figure 4.8: Relationship between multi-tunnel time and threshold with moving patterns.

tunnel time for different moving patterns. In an environment where the MN's moving direction and speed in a time slot have a strong correlation with that of the previous time slot, the performance of the proposed scheme could be enhanced further.

4.4 Summary

In this chapter, the vertical handoff delay vs. multi-tunnel time are introduced, followed by the simulation model. Extensive simulation results are presented to evaluate the performance of the proposed scheme. The simulation results show that the proposed scheme outperforms the other two schemes (MIP and two-step handoff scheme) in terms of handoff delay. The relationship between handoff delay and handoff signaling delay, the ratio of pre-handoff events and handoff events are also carried out. The simulation results indicate that the proposed scheme possess the self-adaptability to make a relatively more precise decision to initiate handoffs even if the signaling delay is dynamically changing. A slight variation in the vertical handoff delay and multi-tunnel is introduced due to the errors incurred

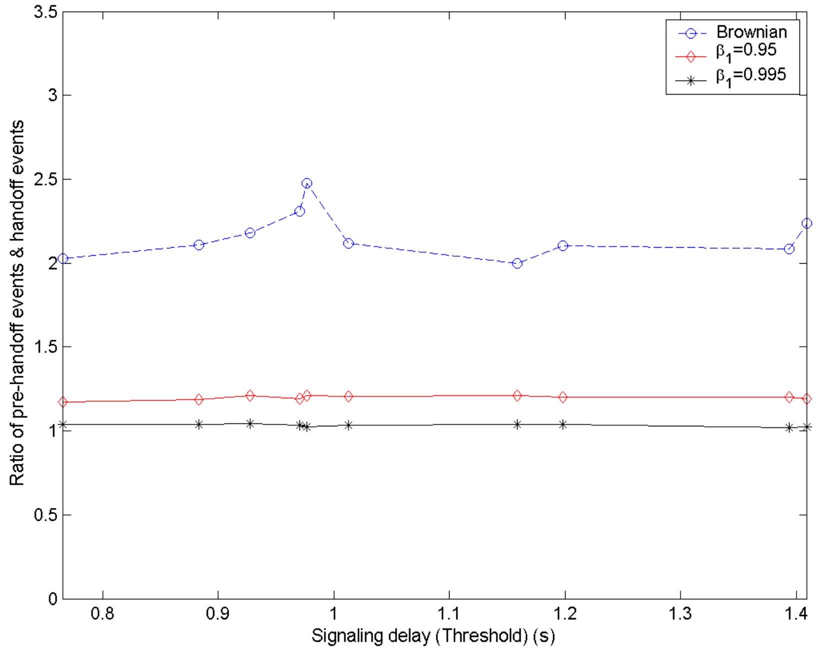

Figure 4.9: Ratio between number of pre-handoff events and number of handoff events vs threshold with different moving patterns.

in estimating handoff signaling delay and the right time to initiate the handoff. The handoff delay as well as multi-tunnel time with respect to different movement patterns are also conducted in the simulation.

Chapter 5

Conclusions

In this thesis, we studied the vertical handoff management for integrated UMTS and WLAN networks. This chapter provides some concluding remarks, highlights the main contributions of this thesis, and presents some possible future research directions.

Highlight of Each Chapter

In Chapter 2, the UMTS and 802.11 based WLAN technologies are reviewed, followed by the discussion on the different coupling architectures for integration UMTS and WLAN networks. The handoff management techniques proposed from different TCP/IP protocol layers are also presented, including IP layer schemes, TCP layer schemes and link-layer schemes. UMTS and WLAN are two different wireless access networks in terms of the data rate, coverage area and media access technologies. But none of them can simultaneously provide the high data rate, low latency, and ubiquitous service for the mobile users. However, since these two wireless networks are complementary to each other, their integration and coordinated operation can provide always best connected high data rate services to mobile users. Depending the deployment requirement, the UMTS and WLAN can be integrated using either a loose coupling architecture or a tight coupling architecture. The difference between these solutions lies in the level of integration between the two networks. In loose coupling architecture, WLAN and UMTS are connected via the Internet. The WLAN data traffic never go through the UMTS core network. And they can be deployed independently saving a lot of efforts to modify UMTS network architecture. However, in the tight coupling architecture, WLAN is connected directly to the UMTS core network, which results in additional system cost and design complexity because the load of UMTS core networks are designed to support cellular users only. Integration of UMTS and WLAN calls for intelligent handoff management techniques to provide seamlessly roaming services for mobile users. Mobility management protocols operating from different layers of the TCP/IP protocol stack, e.g., link layer, network layer, transport layer, and application layer are developed to support mobility in next generation wireless systems. However, some requires to modify the existing TCP/IP protocols stack in order to implement in them in practice. Others set a fixed handoff threshold. And they implicitly assume the signaling delay is a fixed value which is not true in reality due to the dynamically changing network conditions.

In Chapter 3, the performance of a link layer assisted handoff management protocol is discussed, which uses a fixed threshold to initiate the handoff process. The deficiency of using this protocol in situations with a changing handoff signaling delay is pointed out. It is observed that the handoff threshold value need to be adjusted in accordance with the changing value of the MN's moving speed as well as the dynamically changing handoff signaling delay in order to maintain a particular value of handoff failure probability. Therefore, the application of an adaptive threshold is desired to deal with the problem in this situation. Based on the insights, an adaptive vertical handoff management scheme is proposed, which incorporates the ideas of pre-handoff and adaptive handoff threshold. The MN estimates the handoff signaling delay of a possible handoff in advance and adjust the handoff threshold accordingly. The RSS and its rate of change are measured by the MN to estimate the remaining time to keep connected with the current serving BS. The pre-handoff process will be initiate when the estimated time duration to handoff goes below the adaptive pre-handoff threshold. The implementation of the proposed scheme is also discussed.

In Chapter 4, the vertical handoff delay vs. multi-tunnel time incurred in a pre-handoff process are introduced to evaluate the handoff performance of the proposed scheme, followed by the simulation model adopted in our proposed scheme. Extensive simulation results are presented to evaluate the performance of the proposed scheme. In the simulation, the comparison among MIP, a two-step handoff scheme and the proposed handoff scheme are carried out. The simulation results show that our proposed scheme outperform the other two schemes in terms of handoff delay. The relationship between handoff delay and handoff signaling delay, the ratio of pre-handoff events and handoff events are also studied. The simulation results indicate that the proposed scheme possess the self-adaptability to make a relatively more precise decision to initiate handoffs even if the signaling delay is dynamically changing. The reason behind is that the proposed scheme estimates the handoff signaling delay in advance for a possible handoff and adjust the handoff threshold accordingly. A slight variation in the vertical handoff delay and multi-tunnel is introduced due to the errors incurred in estimating handoff signaling delay and the right time to initiate the handoff. The handoff delay as well as multi-tunnel time with respect to different movement patterns are also conducted in our simulation.

Contributions

In our proposed scheme, the idea of pre-handoff and an adaptive handoff threshold have been incorporated to provide mobility management for integrated UMTS and WLAN networks. The handoff signaling is dynamically changing due to the factors such as traffic load in backbone network, wireless channel condition and the distance of MN from its home network. With that in mind, the handoff signaling delay is estimated in advance for a possible handoff process. The adaptive handoff threshold will be adjusted accordingly for every single MN based on the estimated handoff signaling delay. The MN also needs to measure the RSS from the current serving BS periodically. Using the RSS and the RSS's rate of change at a specific time instant, we can estimate the remaining time duration of the MN before tearing down its connection with the current serving BS due to insufficient signal strength. Using these information obtained, an appropriate time instant to initiate a pre-handoff process for every single MN will be determined during the course of its movement accordingly. The pre-handoff process is initiated once the estimated time duration to handoff goes below the adaptive pre-handoff threshold. The handoff performance was evaluated by extensive simulation results.

Future Directions

In this section, we present some future research directions in handoff management.

1. In this thesis, the RSS and the MN's speed are used to make an adaptive handoff decision. However, we need to build a more accurate mobility model to estimate the right time to initiate the handoff procedure so as to improve the handoff performance further.

2. We have estimated the handoff singling delay by adopting a simple method which uses the rough trip time of a packet from the MN to the HA. In the future, we need to find out a better solution without introducing extra signaling overhead to the system by sending those packets.

3. In our simulation, we assume that the MN's speed does not change dramatically during the time duration of the handoff process. In the future, we need to improve the handoff performance of the proposed scheme when the MN's speed is significantly increased.

List of References

[1] S. Ortiz Jr. Internet telephony jumps off the wires. *Computer*, 37(12):16–19, 2004. 1

[2] S.Y. Hui and K.H. Yeung. Challenges in the migration to 4G mobile systems. *Communications Magazine, IEEE*, 41(12):54–59, 2003. 1

[3] E. Gustafsson, A. Jonsson, E. Res, and S. Stockholm. Always best connected. *Wireless Communications, IEEE*, 10(1):49–55, 2003. 1

[4] W. Wu, N. Banerjee, K. Basu, and SK Das. SIP-based vertical handoff between WWANs and WLANs. *Wireless Communications, IEEE*, 12(3):66–72, 2005. 2

[5] R. Chakravorty, P. Vidales, L. Patanapongpibul, K. Subramanian, I. Pratt, and J. Crowcroft. On inter-network handover performance using Mobile IPv6. June 2003. 2

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

[7] C. Perkins et al. IP mobility support for IPv4, 2002. 3

[8] A.C. Snoeren and H. Balakrishnan. An end-to-end approach to host mobility. In *Proceedings of the 6th annual international conference on Mobile computing and networking*. ACM Press New York, NY, USA, 2000. 3

[9] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. SIP: session initiation protocol. 1998. 4

[10] S. Mohanty and I.F. Akyildiz. A cross-layer (layer 2+ 3) handoff management protocol for next generation wireless systems. *IEEE Trans. on Mobile Computing*, 5(10):1347 – 1360, 2006. 4, 5, 21, 22, 27

[11] 3GPP TS 23.002 V6.5.0. *Technical specification 3rd generation partnership Project; technical specification group services and systems aspects; network architecture (Release 6)*. June 2004. 7

[12] IEEE 802.11 WG. *Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specification*. August 1999. 9

[13] IEEE 802.11b WG. *Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specification: high-speed physical layer extension in the 2.4 GHz band*. September 1999. 9

[14] IEEE 802.11b WG. *Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specification: high-speed physical layer in the 5GHz band*. September 1999. 9

[15] C.E. Perkins. Mobile networking in the Internet. *Mobile Networks and Applications*, 3(4):319–334, 1998. 12

[16] H. Soliman, C. Castelluccia, K. El-Malki, and L. Bellier. Hierarchical Mobile IPv6 mobility management (HMIPv6). *draft-ietfmobileip-hmipv6-07.txt, IETF.*, October 2002. 14

[17] E. Gustafesson, A. Jonsson, and C. Perkins. Mobile IPv4 regional registration. 14

[18] A.T. Campbell, J. Gomez, S. Kim, A.G. Valko, C.Y. Wan, and Z. Turanyi. Design, implementation, and evaluation of cellular IP. *IEEE Personal Communications*, 7(4):42–49, 2000. 14

[19] R. Ramjee, K. Varadhan, L. Salgarelli, S.R. Thuel, S.Y. Wang, and T. La Porta. HAWAII: a domain-based approach for supporting mobility in wide-area wireless networks. *IEEE/ACM Transactions on Networking (TON)*, 10(3):396–410, 2002. 14, 15

[20] A.T. Campbell, J. Gomez, S. Kim, C.Y. Wan, Z.R. Turanyi, and A.G. Valko. Comparison of IP micromobility protocols. *IEEE Wireless Communications Magazine*, 9(1):72–82, 2002. 15

[21] M. Shi, L. Xu, X. Shen, J.W. Mark, and A. Saleh. Air interface switching and performance analysis for fast vertical handoff in cellular network and WLAN interworking. *Wiley Journal of Wireless Communication and Mobile Computing (JWCMC)*, to appear. 15

[22] S.E. Kim and JA Copeland. TCP for seamless vertical handoff in hybrid mobile data networks. In *Proceedings of IEEE GLOBECOM'03*, volume 2, pages 661–665, 2003. 15

[23] P. Vidales, L. Patanapongpibul, G. Mapp, and A. Hopper. Experiences with heterogeneous wireless networks, unveiling the challenges. In *Proceedings of the 2nd International Working Conference on Performance Modeling and Evaluation of Heterogeneous Networks(HET-NETs 04)*, 2004. 16

[24] H. Rutagemwa, S. Pack, X. Shen, and J.W. Mark. Robust cross-layer design of wireless profiled TCP mobile receiver for vertical handover. *IEEE Trans. on Vehicular Technology*, to appear. 17

[25] H. Rutagemwa, S. Pack, X. Shen, and JW Mark. Cross-layer design and analysis of wireless profiled TCP for vertical handover. In *Communications, 2007. ICC'07. IEEE International Conference on*, pages 4488–4493, 2007. 17

[26] N. Banerjee, W. Wu, K. Basu, and S.K. Das. Analysis of SIP-based mobility management in 4G wireless networks. *Computer Communications*, 27(8):697–707, 2004. 17

[27] H. Yokota, A. Idoue, T. Hasegawa, and T. Kato. Link layer assisted mobile IP fast handoff method over wireless LAN networks. In *Proceedings of the 8th Annual International Conference on Mobile Computing and Networking*, pages 131–139. ACM Press New York, NY, USA, 2002. 18

[28] Y. Wang. A novel two-step vertical handoff scheme for cellular/wlan interworking. Master's thesis, University of Waterloo, 2006. 18, 29, 39

[29] M.M. Buddhikot, G. Chandranmenon, S. Han, Y.W. Lee, and S. Miller. Design and implementation of a WLAN/CDMA2000 interworking architecture. *IEEE Communications Magazine*, pages 90–100, 2003. 18

[30] HG Ebersman, OK Tonguz, M.D. Syst, and NY Rochester. Handoff ordering using signal prediction priority queuing inpersonal communication systems. *Vehicular Technology, IEEE Transactions on*, 48(1):20–35, 1999. 27

[31] AE Xhafa and OK Tonguz. Dynamic priority queueing of handover calls in wireless networks: an analytical framework. *Selected Areas in Communications, IEEE Journal on*, 22(5):904–916, 2004. 27