NAIST-IS-DD0761034

Doctoral Dissertation

Handover Management for Real-time Application over Multi-homed Broadband Wireless Access Networks

Muhammad Niswar

February 4, 2010

Department of Information Systems
Graduate School of Information Science
Nara Institute of Science and Technology

A Doctoral Dissertation
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of ENGINEERING

Muhammad Niswar

Thesis Committee:

Professor Suguru Yamaguchi	(Supervisor)
Professor Minoru Okada	(Co-Supervisor)
Associate Professor Youki Kadobayashi	(Co-Supervisor)

Handover Management for Real-time Application over Multi-homed Broadband Wireless Access Networks*

Muhammad Niswar

Abstract

Heterogeneous broadband wireless access networks will be deployed in many areas. Recently, IEEE 802.11g has been dominant Wireless Local Area Network (WLAN) standard and widely used to provide high data rates in a limited area such as office, cafe, hotels, school and airport. On the other hand, the emerging mobile WiMAX (IEEE 802.16e) has gained serious attention as a means of providing wireless broadband access to mobile users in a wide area and it provides QoS for various applications. 802.11g and 802.16e will become a key technology as means of an economically viable solution for providing wireless broadband access to mobile user. These two different wireless access technologies will coexist while complementing each other in the near future, hence, a mobile node (MN) with dual interfaces will be likely to execute many handovers between 802.11g networks as well as between 802.11g and 802.16e networks with different IP subnets. Meanwhile, there is a huge demand for Voice over IP (VoIP) service over wireless network. VoIP applications are delay and loss sensitive application. An acceptable VoIP call must have a total end-to-end delay (E2E) not exceeding 150-200 ms. However, wireless access networks exhibit significant variations in packet delay. Furthermore, when moving to a network that offers a lower bit rate, a VoIP application may experience congestion situations with packet loss and service degradation. Therefore, preserving VoIP communication quality over wireless network is a challenging and important issue particularly in mobile wireless environment.

This dissertation presents an end-to-end handover management for VoIP over 802.11g networks as well as intermingled 802.11g and 802.16e networks considering wireless link condition and congestion state of both wireless networks. The handover management exploits Request-To-Send (RTS) retries and Round-Trip-Time (RTT) of 802.11g interface

*Doctoral Dissertation, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0761034, February 4, 2010.

as well as Carrier-to-Interference-plus-Noise-Ratio (CINR) level and MN queue length of 802.16e interface as handover decision metrics. Our handover management implemented on transport layer controls handover according to handover decision metrics obtained through cross layer approach. Moreover, we also employ multi-homed MN that can support single-path and multi-path transmission for achieving seamless handover. Single-path transmission means that MN communicate with CN using single interface and multi-path transmission, on the other hand, means that MN sends duplicated packets to a CN using two interfaces for supporting seamless handover. Our proposed methods aim to preserve VoIP quality during handover between the networks with different IP subnets. We conducted simulation experiments to investigate the effectiveness of our proposed handover management using Qualnet 4.5. Our simulation results show that our proposed handover management can preserve VoIP quality during MN's handover between two 802.11g networks as well as between 802.11g and 802.16e networks.

Keywords:

Handover Management, Handover Decision Criteria, Multi-homed, Cross Layer, VoIP, 802.11g, 802.16e

Contents

1 Introduction **1**
1.1. Research Motivation . . . 1
1.2. Emerging Broadband Wireless Access . . . 2
1.3. Co-existence of 802.11g and 802.16e . . . 4
1.4. Mobility Management . . . 5
1.5. VoIP over Wireless Network . . . 7
1.5.1 VoIP CODEC . . . 8
1.5.2 VoIP Quality Assessment . . . 8
1.6. Research Contribution . . . 9
1.7. Outline of Dissertation . . . 11

2 Related Work and Problem Statement **12**
2.1. Existing Mobility Management . . . 12
2.1.1 Network Layer Solution . . . 12
2.1.2 Transport Layer Solution . . . 14
2.1.3 Application Layer Solution . . . 16
2.1.4 Additional Layer Solution . . . 17
2.2. Problem Statement . . . 17

3 Proposed Multi-homed Handover Management Architecture **18**
3.1. Multihoming . . . 18
3.2. Motivation and Benefits of Multihoming . . . 19
3.3. Handover Management Architecture . . . 20
3.4. Single-Path and Multi-Path Transmission . . . 20

4 Handover Management for VoIP over 802.11g Networks **22**
4.1. Introduction . . . 22
4.2. Related Work . . . 23

4.3. Proposed Handover Decision Criteria . . . 25
4.3.1 RTS Frame Retries . . . 26
4.3.2 AP Queue Length . . . 29
4.3.3 Transmission Rate . . . 30
4.4. Proposed Handover Management . . . 31
4.4.1 Single-Path and Multi-Path Transmission . . . 33
4.4.2 Dealing with Ping-Pong Effect (Extension Method 1) . . . 34
4.4.3 Elimination of Redundant Probe Packets (Extension Method 2) . . 39
4.5. Performance Evaluation . . . 40
4.5.1 Evaluation of Basic Methods . . . 40
4.5.2 Evaluation of Extension Method 1 . . . 44
4.5.3 Evaluation of Extension Method 2 . . . 45
4.5.4 Random Movement Environment . . . 45
4.5.5 Limitation . . . 46
4.6. Summary . . . 46

5 Handover Management for VoIP over Intermingled 802.11g and 802.16e 47
5.1. Introduction . . . 47
5.2. Related Work . . . 47
5.3. Overview of 802.16e . . . 48
5.4. VoIP over Wireless Networks . . . 50
5.4.1 VoIP over 802.11g . . . 51
5.4.2 VoIP over 802.16e . . . 51
5.5. Proposed Handover Decision Criteria . . . 52
5.5.1 Handover Decision Criteria for 802.16e . . . 52
5.6. Evaluation of Handover Decision Criteria for 802.16e . . . 57
5.6.1 Evaluation of CINR . . . 58
5.6.2 Evaluation of MN Queue Length . . . 59
5.7. Proposed Handover Management . . . 64
5.7.1 Single-Path and Multi-Path Transmission . . . 65
5.7.2 Assumption and Considered Environments . . . 67
5.8. Simulation Experiments and Results . . . 67
5.8.1 Performance Evaluation of handover based on Link State Handover Decision Criteria . . . 68
5.8.2 Performance Evaluation of Handover based on Congestion State Handover Decision Criteria . . . 68

5.9. Summary 74

6 Conclusion and Future Direction **78**

6.1. Conclusion 78

6.2. Future Direction 80

Acknowledgments **81**

Bibliography **83**

Publications **88**

List of Figures

1.1 Wireless Technologies . . . 3
1.2 Co-existence of 802.11 and 802.16 . . . 5
1.3 Classification of Mobility Management . . . 6
1.4 Handover Process . . . 7

2.1 Mobile IP: MN moves to foreign network . . . 13
2.2 Communication Path of Mobile IP . . . 13
2.3 Session Initiation Protocol . . . 16

3.1 Proposed Handover Management Architecture . . . 21

4.1 Proposed Handover Architecture . . . 25
4.2 RTS Retry Rate vs. MOS over Distance . . . 26
4.3 Relationship among # of MNs, AP queue length, and MOS . . . 27
4.4 Relationship among AP queue length, W-RTT, and MOS . . . 27
4.5 RTT between AP and MN (W-RTT) . . . 28
4.6 Simulation Model 1 . . . 31
4.7 Switching to Single/Multi-Path Transmission . . . 32
4.8 Switching from Multi-Path Transmission to Single-Path Transmission . . . 32
4.9 Handover based on RTS Retry Ratio . . . 33
4.10 Handover based on Transmission Rate . . . 34
4.11 Calculate W-RTT from existing probe packet . . . 35
4.12 Obtaining a right to send the probe packet . . . 36
4.13 Simulation Models 2 . . . 38
4.14 MN's MOS over distance . . . 40
4.15 Use Rate of AP1 over distance . . . 41
4.16 Variation of MN's MOS . . . 41
4.17 Variation of AP1 queue length . . . 42

4.18 Variation of MN's MOS . . . 42
4.19 Variation of AP1 queue length . . . 43
4.20 Variation of AP1 queue length of our previous method . . . 43
4.21 Variation of AP1 queue length of the extension method 2 . . . 44

5.1 802.16 MAC Frame . . . 50
5.2 Uplink Data Transmission . . . 51
5.3 CINR as a reference for changing burst profile . . . 54
5.4 Simulation Model 3 . . . 55
5.5 Simulation Model 4 . . . 56
5.6 Simulation Model 5 . . . 56
5.7 MOS vs. CINR . . . 58
5.8 Packet Loss vs. CINR . . . 59
5.9 Relationship among Uplink MOS, MN Queue Length . . . 60
5.10 Relationship among Uplink MOS, Delay and ♯ of Stationary MN . . . 61
5.11 Relationship among Uplink MOS, Packet Loss and ♯ of Stationary MN . . . 61
5.12 Relationship among Uplink MOS, MN Queue Length and ♯ of Moving MNs 62
5.13 Relationship among Uplink MOS, Delay and ♯ of Moving MNs . . . 62
5.14 Relationship among Uplink MOS, Packet Loss and ♯ of Moving MNs . . . 63
5.15 Ave. MN-QL, Uplink Delay, MOS of Stationary MNs . . . 63
5.16 Ave. MN-QL, Uplink Delay, MOS of Moving MNs . . . 64
5.17 Algorithm of Switching from Singlepath to Singlepath/Multipath . . . 66
5.18 Algorithm of Switching from Multipath to Singlepath/Multipath . . . 67
5.19 Simulation Model 6 (Move from 802.11g to 802.16e) . . . 69
5.20 Link State Criteria Charachteristics (802.11g to 802.16e) . . . 69
5.21 Simulation Model 7 (Move from 802.16e to 802.11g) . . . 70
5.22 Link State Criteria Charachteristic (802.16e to 802.11g) . . . 70
5.23 MOS of CN (802.11g to 802.16e) . . . 71
5.24 MOS of MN (802.11g to 802.16e) . . . 71
5.25 MOS of CN (802.16e to 802.11g) . . . 72
5.26 MOS of MN (802.16e to 802.11g) . . . 72
5.27 Simulation Model 8 . . . 73
5.28 Congestion State Criteria Charachteristics (802.11g to 802.16e) . . . 74
5.29 MOS of CN (802.11g to 802.16e) . . . 75
5.30 MOS of MN (802.11g to 802.16e) . . . 75
5.31 Congestion State Criteria Charachteristics (802.16e to 802.11g) . . . 76

5.32 MOS of CN (802.16e to 802.11g) 76
5.33 MOS of MN (802.16e to 802.11g) 77

List of Tables

1.1 Voice Quality Classes 10

4.1 Simulation Parameters 1 31
4.2 System Parameters 37

5.1 Simulation Parameters 2 57
5.2 Threshold Value 68

Chapter 1

Introduction

Future wireless network will be consisting of various broadband wireless access technologies such as IEEE 802.15 wireless personal area network (WPAN), IEEE 802.11 wireless local area network (WLANs), IEEE 802.16 wireless metropolitan area networks (WMANs), IEEE 802.20 wireless wide area networks (WWANs), General Packet Radio Service (GPRS), and Universal Mobile Telecommunication System (UMTS). These various broadband wireless access technologies will co-exist in some areas and mobile users will traverse several different broadband wireless access technologies while establishing VoIP communication.

Recently, IEEE 802.11g has been dominant WLAN standard and widely used to provide high data rates in limited area such as office, cafe, hotels, school and airport. On the other hand, the emerging mobile WiMAX (IEEE 802.16e) has gained serious attention as a means of providing wireless broadband access to mobile users in a wide area and it provides QoS for various applications. Therefore, in the near future, since 802.11g and 802.16e networks co-exist while complementing each other, a mobile node (MN) will be likely to execute many handovers between 802.11g and 802.16e networks with different IP subnets.

1.1. Research Motivation

Handover management between different broadband wireless access technologies has been a hot issue and development topic in the past few years. In heterogeneous wireless network, an MN will be equipped with multiple interfaces to access different wireless networks. This MN has a great flexibility for network access and connectivity but introduce the challenging problem of mobility support among different wireless networks. The MN expects to maintain the connectivity without any disruption when they move from one network to another. To achieve seamless handover, we need to develop a handover management for heterogeneous wireless networks.

Prior to developing the handover management, we need to define handover decision criteria for heterogeneous wireless network. Many technologies use receive signal strength (RSS) or Carrier-Interference-plus-Noise Ratio (CINR) as handover decision criteria. However, in heterogeneous network or vertical handover, we cannot compare directly the CINR level of two different wireless technologies since each technology present distinct characteristics in terms of physical and MAC layers. For instance, they might have same CINR level but differ in data rate. Therefore, it is challenge to develop a vertical handover management considering suitable handover decision criteria from two different wireless networks that give an indication of wireless link and congestion state in order to achieve seamless handover and optimal wireless resources utilization in heterogeneous environment.

In this dissertation, we study the handover decision criteria for both 802.11g and 802.16e and propose an effective handover management for Voice over IP (VoIP) over 802.11g networks as well as intermingled 802.11g and 802.16e networks. The proposed handover management aims to maintain adequate VoIP quality during handovers between two different wireless access technologies.

1.2. Emerging Broadband Wireless Access

Recently, several emerging broadband wireless access technologies have been deployed in many areas. These wireless access technologies can be categorized based on their coverage area: WPAN, WLAN, WMAN and WWAN as shown in Fig.1.1

WPAN is a wireless data network used for communication among data devices close to one person. WPAN is designed to provide wireless access in few meters area with the radius of less than 10m. Example of WPAN technologies are Bluetooth (IEEE 802.15) and Zigbee.

WLAN is designed to provide wireless access in local area with radius of 50-100m. The most widely deployed WLAN standard is IEEE 802.11. Another one is the HiperLAN by ETSI. Both technologies are united under the Wireless Fidelity (WiFi) alliance [1]. A typical WLAN consists of an access point (AP) and MN. Usually, the MN communicates with its peer through AP when infrastructure mode is applied. Furthermore, the MN also can communicate directly to another MN when ad-hoc mode is applied.

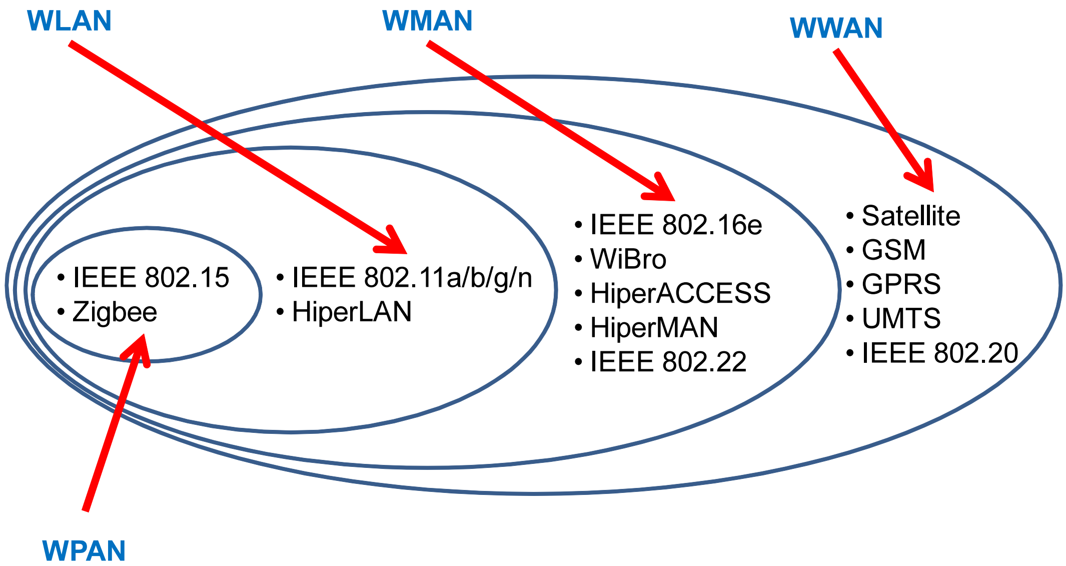

Figure 1.1. Wireless Technologies

Although WLAN is widely deployed, WMAN is expected to be deployed with a large number in the near future. IEEE 802.16 has been developed by IEEE in order to provide broadband wireless access to fixed Line of Sight (LOS) subscribe station (SS) from a Base Station (BS). IEEE 802.16e, the current version, is designed to provide broadband wireless access to mobile SS with non-LOS (NLOS) from BS. This technology expected to provide data rate up to 100 Mbps with coverage area of 3-50Km depending on operated frequency band. Furthermore, Telecommunication Technology Association of South Korea has developed Wireless Broadband (WiBro), which is based on IEEE 802.16. On the other hand, ETSI developed HiperACCESS for LOS and HiperMAN for both LOS and NLOS user support. Another recent WMAN technology is IEEE 802.22 standard for Wireless Regional Area Network (WRAN). This technology uses unoccupied TV channels in the 54-862 MHz range, depending on the region of operating. IEEE 802.22 is expected to support data rates up to 18 Mbps with coverage are of 20-40 km.

Satellite system provides service in widest coverage for WWAN. Most of current satellite system is capable for downlink broadcast communication only. Next Generation Satellite Systems (NGSSs) are expected to have on-board routing capabilities and support upload channel. These NGSSs will play an important role in future broadband system. Furthermore, IEEE 802.20 is another emerging WWAN technology. This technology aims to provide a broadband wireless access to highly mobile vehicles with speeds up to 250 km/h. But expected data rate is not as high as WLAN and WMAN due to the high speed of MN.

1.3. Co-existence of 802.11g and 802.16e

These days, the IEEE 802.11g standard has been the dominant WLAN standard and is widely deployed to provide high data rates in a limited coverage area such as office, cafe and school. Currently, 802.11g is embedded in most laptops and handheld. On the other hand, emerging mobile IEEE 802.16e standard promises to guarantee broadband access to many mobile users within a wide coverage area and it provides QoS for various applications.

IEEE802.11g and IEEE802.16e technologies leverage Orthogonal Frequency Division Multiple Access (OFDMA) technology and use IP-based technologies to provide connection services to the internet. Although both technologies provide broadband wireless connectivity, they have been optimized for different usage in terms of coverage area: 802.11g for high-speed WLAN connectivity and 802.16e for high-speed WMAN connectivity. Many enterprises, governments, and universities have deployed 802.11g in-buildings for their employee and students. 802.16e offers broadband connectivity beyond individual buildings to provide blanket coverage of an entire campus. Dual interface of both 802.11g and 802.16e in a single MN enable users to connect to either in-building 802.11g networks or campus-wide 802.16e networks allowing them to stay connected as they move. Moreover, using this dual-mode MN, network administrators can also reduce the number of 802.11g APs needed to attain full campus coverage, thereby reducing maintenace costs. We believe that deploying 802.11g and 802.16 in same place can offer users a more complete suite of broadband services in more places.

Furthermore, these two technologies will become key technology as means of an economically viable solution to providing wireless broadband access to mobile user. Therefore, these two different wireless access technologies co-exist while complementing each other in the near future, hence, an MN with dual interfaces will be likely to execute many handovers between 802.11g and 802.16e networks with different IP subnets Figure 1.2 shows that an MN traverses 802.11g and 802.16e networks. When the MN handovers to same wireless access technology (from 802.11g to 802.11g), this is called horizontal handover. On the other hand, when the MN handovers to different wireless technology (from 802.11g to 802.16e), this is called vertical handover.

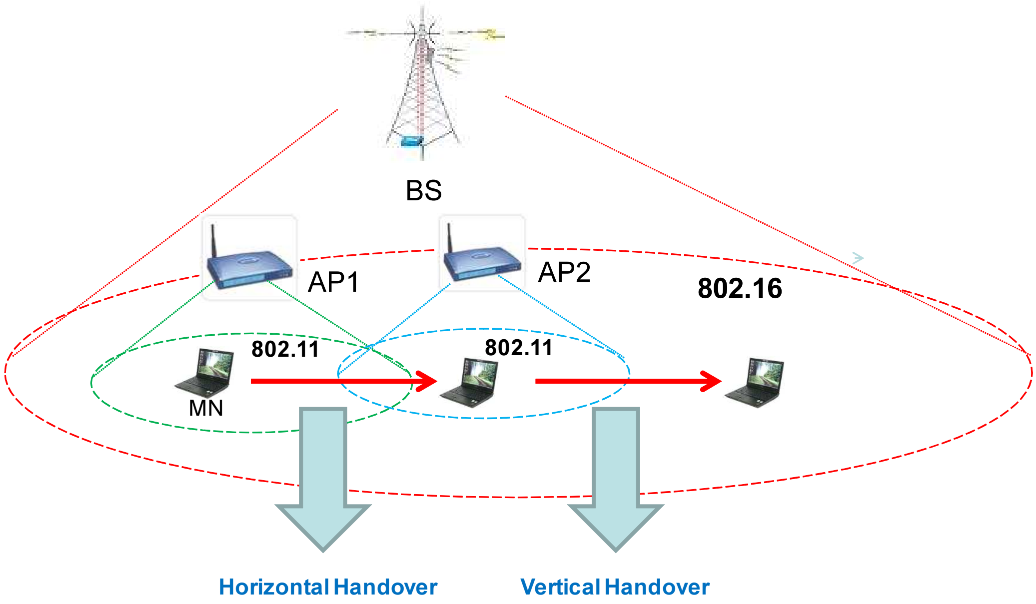

Figure 1.2. Co-existence of 802.11 and 802.16

1.4. Mobility Management

Mobility management plays an important role in current and future wireless mobile network for delivering service to mobile user that move from one to another network. Mobility management consists of two component operations: handover and location management. Handover occurs when the MN changes the point of attachment while it is still communicating with its peer. Location management refers to locate mobile users in order to deliver data packets to them despite the fact that their locations may change from time to time. Figure 1.3 shows the classification of mobility management. In this study, we only consider handover management solution for emerging broadband wireless access technology.

Handover management is a fundamental operation for any mobile network. Although its functionality and implementation differs among the various technologies, some basic characteristics are common. The handover process can be divided into three stages (Fig.1.4): Handover initiation is a stage to triggering handover according to some parameters/specific conditions such as link deterioration due to the movement of MN or network congestion. During handover decision stage, the decision for changing the AP is made based on comparison of several parameters such as received signal strength of both serving

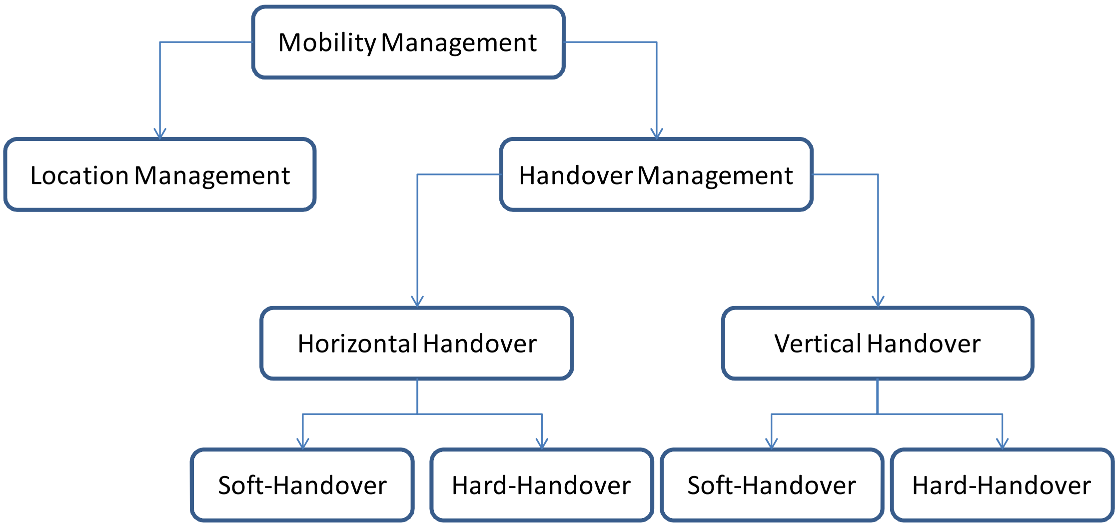

Figure 1.3. Classification of Mobility Management

and neighboring AP. In this stage, path/route selection or signaling exchange is conducted in order to determine the best path to delivery the data. Finally, handover execution is made through data re-routing to new path for communication re-establishment.

The main objective of handover management is to preserve the communication quality during the handover of MN. To achieve this, we must address the issue how the way of the old link is released and the new ones are established. In this respect, there are two types of handover: soft and hard handover. The soft handover refers to handover that the MN still remain communicating with the serving AP and used for a while in parallel with the communication with the target AP. In this case, the connection to the target AP is established before the connection to the serving AP is broken, hence this handover is called make-before-break. On the other hand, a hard handover refers to handover that the MN released the communication with the serving AP and only then the communication with the target AP is engaged. Thus the connection to the serving AP is broken before the connection to the target AP is made, hence this handover is also known as break-before-make. In this study, we consider the soft handover and discuss the mechanism that can support seamless handover for VoIP application.

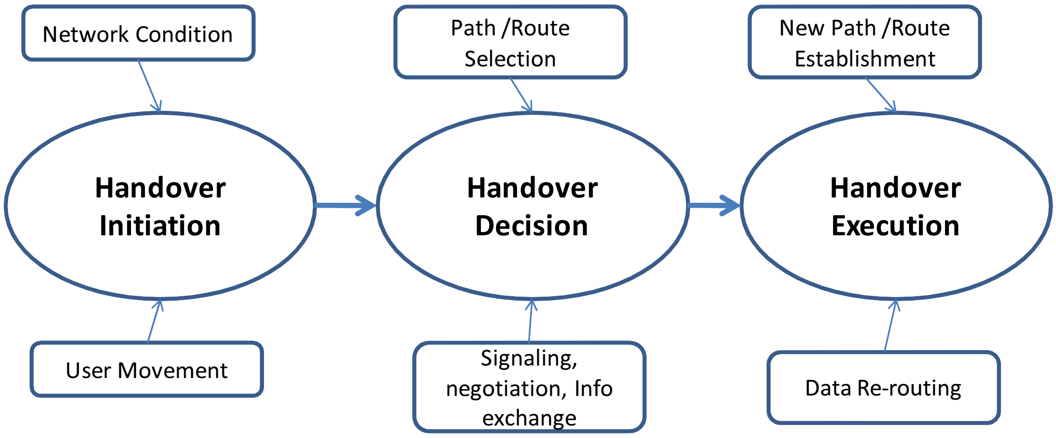

Figure 1.4. Handover Process

1.5. VoIP over Wireless Network

VoIP is probably the most important technology we are facing now because VoIP makes a significant change in the way that people communicate. VoIP can easily and efficiently route the phone calls over the existing data networks and users does not need to maintain separate voice and data networks. VoIP is certainly an all-in-one application that provides various features like conference calling, call forwarding, caller ID features, Interactive Voice Response (IVR), etc with the flexibility of using the phone and the internet at the same time. VoIP allows users in reducing both communication and infrastructural expenses to a great extent, hence, the demand for VoIP service continues to be strong on both the residential and business side.

VoIP uses the Internet Protocol (IP) to transmit voice as packets over an IP network. In the VoIP, the voice signal is digitized, compressed and converted to IP packets and then transmitted over the IP network. Signaling protocols are used to set up and tear down calls, carry information required to locate users and negotiate capabilities. VoIP applications are delay and loss sensitive application. An acceptable VoIP call must have a total end-to-end delay (E2E) not exceeding 150-200 ms [2]. However, wireless access networks exhibit significant variations in packet delay. Enforcing a limit on the E2E delay is not easy when passing from one access network to another. VoIP applications must synchronize the speech playout. With a limited amount of buffering due to E2E delay limitations, a change in the network delay might cause a loss of synchronization, thus deteriorating the quality of a VoIP call also in the absence of packet loss. This is the case when packets arrive too late to be played back, or when they arrive too soon and the application cannot perform packet reordering. When moving to a network that offers a lower bit rate, a VoIP application may experience congestion situations with packet loss and service degradation. An acceptable VoIP call must have a packet loss ratio not exceeding 5 percent. Therefore, preserving VoIP communication quality over wireless network is a challenging and important issue particularly in mobile environment.

1.5.1 VoIP CODEC

VoIP uses Compression/Decompression (CODEC) technology for converting audio signals into a digital bit stream and vice versa. Today, most VoIP uses codec that standardized by ITU for the sake of interoperability across vendors. The most popular CODECs are G.711, G.723 and G.729. G.711 uses Pulse Code Modulation (PCM) of voice frequencies, sampled at the rate of 8000 samples/second with 8 bits is used to represent each sample, resulting in a 64 kbps bit rate. G.723 uses Code Excited Linear Prediction (CELP) technique to produce a 5.3 Kbps voice stream, and G.729 uses Conjugate Structure Algebraic Code-Excited Linear Prediction (CS-ACELP) technique to produce an 8 Kbps voice stream. In this study, we employ VoIP CODEC of G.711 that is most commonly used by many vendors for VoIP application.

1.5.2 VoIP Quality Assessment

In voice communication, particularly in VoIP, E-model [52] provides a powerful method to assess speech transmission quality over IP network. E-model is a computational model, standardized by ITU that uses transmission parameters such as delay and packet loss to predict the subjective quality of VoIP. The output of an E-model calculation is a single rating called R-factor that is derived from delays and equipment impairment factors. R-factor is expressed as follow:

$$R = (R_o - I_s) - I_d - I_e \tag{1.1}$$

where I_e and I_d represent the impairment caused by equipment and mouth-to-ear delay, respectively. Because other impairments include loud connection and quantization impairment I_s, the basic signal-to-noise ratio R_o, and the advantage factor (user willingness to accept some quality degradation in return for ease of access) A are irrelevant for assessing speech-transmission quality, hence, we can reduce the expression:

$$R = 94.2 - I_d - I_e \tag{1.2}$$

$$I_d = 0.024d + 0.11(d - 177.3)H(d - 177.3) \tag{1.3}$$

$$I_e = 30ln(1 + 15e)H(0.04 - e) + 19ln(1 + 70e)H(e - 0.04) \tag{1.4}$$

Here d is the one-way delay (in milliseconds) and H(x) is the Heavyside (or Step) function:

$$H(x) = 0 : x < 0 \tag{1.5}$$

$$H(x) = 0 : x > 0 \tag{1.6}$$

The R-factor ranges from 0 to 100 and value of more than 70 indicates a sufficient quality of VoIP communication. This R-factor can be further translated into mean opinion score (MOS), which is subjective numerical indication of the perceived quality of conversational voice call that ranges from 1 (lowest quality) to 5 (highest quality). Relation between R-factor and MOS can be expressed as follows.

$$R < 0 : MOS = 1.0 \tag{1.7}$$

$$0 < R < 100 : MOS = 1 + 0.035R + R(R - 60)(100 - R)7.10^-6 \tag{1.8}$$

$$R > 100 : MOS = 4.5 \tag{1.9}$$

Table 1.1 shows a details of VoIP quality assessment based on MOS. Normally, $MOS >$ 3.60 provide an adequate VoIP call quality. Since we propose a handover management for VoIP over 802.11g and 802.16e, we employ MOS as a performance metric for measurement of VoIP quality in this study.

1.6. Research Contribution

Providing an effective handover management for mobile user is the subject of this dissertation. Handover initiation and decision mechanism are our main discussion in this dissertation. The primary contributions of this dissertation are as follows:

Table 1.1. Voice Quality Classes

R	User Satisfication	MOS
$90 < R < 100$	Very Satisfied	4.3-4.5
$80 < R < 89$	Satisfied	4.0-4.2
$70 < R < 79$	Some Users Dissatisfied	3.6-3.9
$60 < R < 69$	Many Users Dissatisfied	3.1-3.5
$50 < R < 59$	Nearly All Users Dissatisfied	2.6-3.1
$0 < R < 49$	Not Recommended	1.0-2.9

- We explore several metrics from physical, MAC and upper layer of both 802.11g and 802.16e and select the best metrics that can indicate the wireless link condition and network condition of both 802.11g and 802.16e networks. Selected metrics are used as handover decision criteria for initiating or triggering handover.

- After selecting the handover decision criteria, we propose an effective handover management for real-time application (VoIP) over 802.11g networks as well as intermingled 802.11g and 802.16e networks. Our proposed handover management aims to preserve on-going VoIP session while MN moves from one network to another.

- Utilization of multihoming function can enhance the session preservation and share traffic load among networks by selecting the best available connection or by enabling multiple connections. We utilize the benefit of multihoming function in order to support seamless handover in heterogeneous wireless networks.

- Performance evaluation is done based on simulation experiments in order to validate our proposed handover management. We use MOS as a performance metric to evaluate whether our proposed handover management can preserve on-going VoIP session.

- Although our proposed handover management is applied for G.711 VoIP codec, other VoIP codecs as well as another real-time application such as video streaming can be also applied but it requires parameter tuning for handover decision criteria in order to achieve seamless handover.

1.7. Outline of Dissertation

The remaining part of the dissertation is organized as follows:

Chapter 2 describes existing mobility managements take place in different protocol stack and problems arise in existing mobility managements.

Chapter 3 introduces multi-homed networks and proposes handover management architecture. It describes motivation behind using multi-homed network and handover management architecture for multi-homed MN.

Chapter 4 proposes and evaluates horizontal handover management for VoIP over 802.11g networks. This handover management exploits request-to-send (RTS) retries ratio and estimation of queue length of access point (AP) by measuring round trip time (RTT) between MN and AP (W-RTT).

Chapter 5 proposes and evaluates vertical handover management for VoIP over intermingled 802.11g and 802.16e networks. This handover management exploits RTS retries ratio and W-RTT of 802.11g as well as CINR level and MN queue length of 802.16e.

Chapter 6 concludes this dissertation, followed by the direction of further work.

Chapter 2

Related Work and Problem Statement

2.1. Existing Mobility Management

Mobility management in wireless networks can take place in different layers of the OSI protocol stack reference model. The following describes several mobility management protocols and their solutions have been proposed for IP-based wireless system.

2.1.1 Network Layer Solution

Mobile IP [13] is a dominant network layer protocol to solve the problem of changing of MN attachment point while maintaining a permanent IP address of MN. The Mobile IP requires a home agent (HA) and foreign agent (FA) in the home network where MN's addressed belongs to and in the foreign network where MN moves to, respectively. Since IP address comes with the two versions: IPv4 and IPv6, the IETF working group in Mobile IP also proposed two versions: Mobile IPv4 and Mobile IPv6.

In Mobile IPv4 [13], when MN moves to foreign network, the MN obtain a care-of-address (CoA) in the foreign network. The CoA can be obtained through FA's advertisement or some external mechanism such as a Dynamic Host Configuration Protocol (DHCP). The MN performs a binding update to its HA where the MN registers its new CoA with its HA through exchange of registration request and reply messages (Fig.2.1) The HA intercepts and tunnels all datagrams destined to the MN using MN's CoA. Here, communication to MN leads to triangular routing problem. In reverse direction, datagram sent by MN are normally delivered through standard IP routing mechanism without

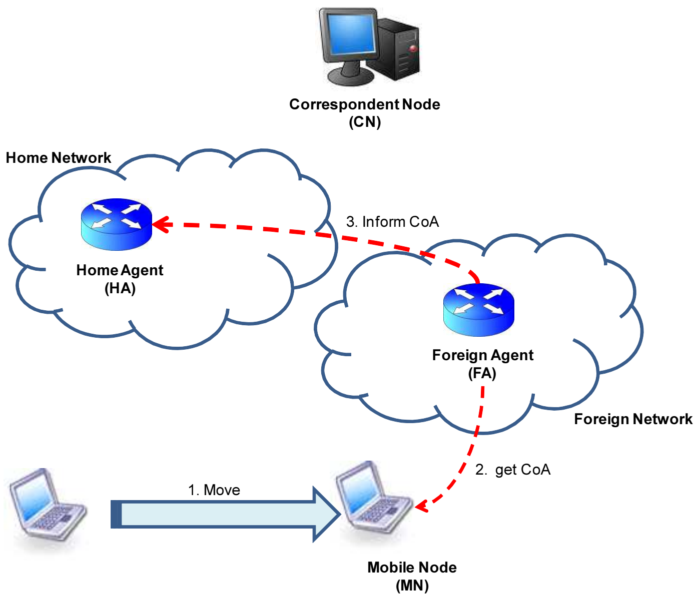

Figure 2.1. Mobile IP: MN moves to foreign network

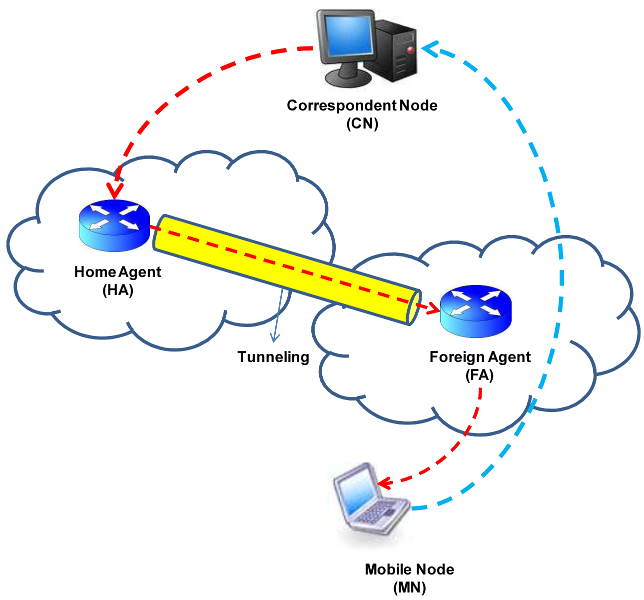

Figure 2.2. Communication Path of Mobile IP

passing through the HA (Fig.2.2).

In Mobile IPv6 [14], a binding update message from the MN is used to inform the HA and CN about the change of point of attachment. Both HA and CN maintain their own biding cache. The Mobile IPv6 provides two alternatives communication between CN and MN that move away from home network: direct communication (route optimization method) and bidirectional tunneling. The route optimization method is a solution to alleviate triangular routing problem in Mobile IPv4. In this method, the datagram can be sent directly from the MN to the CN since the MN registers its current binding at the CN. This method can be applied if CN is Mobile IPv6-capable. On the other hand, if the CN is not Mobile IPv6-capable, bidirectional tunneling is applied. In this method, Data packets from the CN are routed to the HA and then tunneled to the MN. Data packets to the CN are tunneled from the MN to the HA and then routed normally from the home network to the CN.

However, Mobile IP has some drawbacks. The Mobile IP requires a significant signaling cost of the location update to the HA and CN when the number of MNs increases. Furthermore, the handover latency may be large if the foreign and home networks are far away from each other.

2.1.2 Transport Layer Solution

Transport layer mobility protocols are designed to provide an end-to-end approach mobility support that independent of under laying network layer protocols, hence, it keeps the network infrastructure unchanged. Several transport layer mobility protocol have been proposed including MSOCKS, I-TCP, M-TCP, M-UDP, MIGRATE, MMSP , mSCTP.

MSOCKS [3] uses TCP Splice [4] for connection migration and supports multiple IP addresses for multiple interfaces. TCP Splice is a mobility management that split a TCP connection at a proxy by dividing the end-to-end communication into end-proxy and proxy-end communications. When an MN disconnects itself from a subnet during handover, it obtains a new IP address from the new subnet using DHCP, and establishes a new connection with the proxy using its second interface. The communication between proxy and CN, however, remains unchanged. The data flow between MN and CN thus continues, with the CN being unaware of the mobility. Location management is done through the proxy which is always aware of the location of the MN and this limits the mobility within the coverage of the proxy.

Migrate [5] is a transparent mobility scheme which is based on connection migration using Migrate TCP [5], and uses DNS for location management. In Migrate TCP, when an MN initiates a connection with a CN, the end nodes exchange a token to identify the particular connection. A hard handover takes place when the MN reestablishes a previously established connection using the token, followed by migration of the connection.

Indirect TCP (I-TCP) [7] is a mobility scheme that requires a gateway between the communication path of the CN and MN to enable mobility. In this scheme, a TCP connection between CN and gateway and an I-TCP connection between the gateway and MN is established to provide CN to MN communication. In the I-TCP, when the MN moves from one subnet to another one, a new connection between MN and the gateway is established and the old one is replaced by the new one. This scheme requires modification on the transport layer of the MN. I-TCP does not support IP diversity and soft handover as well as location management.

Mobile TCP (M-TCP) [8] is an enhanced version of I-TCP. M-TCP is implemented at MN that works like a link layer one hop protocol that connects to the gateway via wireless network. The gateway maintains a regular TCP connection with the CN and redirects all packets coming from CN to MN. This redirection is unnoticed by both the MN and CN. M-TCP is less complexity in the wireless part of the connection with compare to I-TCP. Similar to I-TCP, M-TCP does not support IP diversity or location management but ensures transparency to applications.

Mobile UDP (M-UDP) [9] is similar to I-TCP and M-TCP but it is an implementation for UDP protocol. Like M-TCP, M-UDP uses a gateway to split the connections between MN and CN to ensure one unbroken gateway to CN connection and continuously changing MN to gateway connection. This scheme also does not support IP diversity or location management.

TCP Redirection (TCP-R) [10] is a connection migration scheme that maintains active TCP connections during handover by updating end-to-end address pairs. Whenever MN gets a new IP address, TCP-R updates the address at CN and the already existent connection continues with the new address. TCP-R does not implement connection timeout to support long disconnection. Transport layer at both the ends needs modification for this support, yet it gives application transparency. Like Migrate, TCP-R proposes to use DNS as location manager. Combined with a handover management scheme, this scheme might be deployed as a complete mobility scheme.

Mobile Stream Control Transmission Protocol (mSCTP) [21] is a mobility extension of SCTP that supports a multi-homing feature and dynamic addition and deletion of addresses to and from an existing association. The multi-homing feature of mSCTP can be used to provide seamless handover for MNs that are moving into different wireless

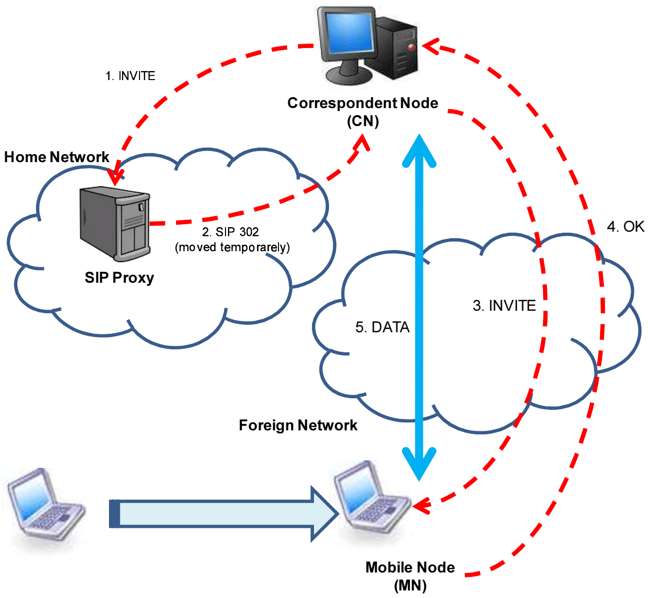

Figure 2.3. Session Initiation Protocol

networks during the active session. However, in this scheme, the issue of handover decision is not discussed in details.

2.1.3 Application Layer Solution

Session Initiation Protocol (SIP) [11] is an application layer mobility that standardized by IETF. SIP is an end-to-end oriented signaling protocol and can be applied for real time applications such as voice and video calls over IP. SIP is a text-based protocol with syntax similar to that of HTTP. SIP defines an extensible set of request methods, including in the base specification INVITE to initiate a session, ACK to confirm a session establishment, BYE to terminate a session, OPTIONS to determine capabilities and CANCEL to terminate a session that has not been established yet. The session itself is typically described using the Session Description Protocol (SDP) that lists media stream addresses, ports and the encodings supported.

SIP can be used for mobility support and SIP-based mobility support is only suitable for UDP-based application. Figure 2.3 shows SIP signaling exchanges. When MN moves to foreign network, MN acquires a new IP address prior to receiving or making a call. MN simply registers with home SIP proxy each time it obtains a new IP address. When CN want to establish communication to MN, CN send INVITE request to home SIP proxy.

Home SIP proxy informs the new address of MN. CN sends another INVITE request to the MN without going through SIP proxy. This INVITE message contains updated session description with a new IP address.

However, SIP-based mobility, in particularly, the handover procedure may introduce latency for the signaling messages procedure and overhead for IP encapsulation.

2.1.4 Additional Layer Solution

Media Independent Handover (MIH) is a standard being developed by IEEE 802.21 to enable handover between heterogeneous access networks. MIH facilitates the network discovery and selection process by exchanging network information that helps mobile devices determine which networks are in their current neighborhoods. IEEE 802.21 introduces additional layer called MIH function located between layer 2 and 3 (Layer 2.5). MIH function provides abstracted services to higher layers by means of a unified interface and it defines three different services: Media Independent Event Service (MIES), Media Independent Command Service (MICS) and Media Independent Information Service (MIIS). MIES provides triggered events corresponding to dynamic changes in the link condition. MICS helps the MIH user to manage and control the link behavior relevant to handovers and mobility. MICS uses the information obtained from the event service as part of the subscription and notification process and acts upon it accordingly. MIIS provides an information model that passes on the information regarding the neighboring networks and their capabilities. The drawback is that since MIH requires modification/additional layer to the existing protocol stack of mobile device and infrastructure networks leading to cost inefficiency.

2.2. Problem Statement

Many existing mobility management have been proposed and deployed in different protocol stacks. So far, they are lack of discussion of handover decision criteria to initiate handover and handover decision making process in details. Hence, we do not know whether existing mobility management can preserve communication quality during handover.

This dissertation address these issues because they are a key point to execute the handover and how we implement them in order to preserve the quality of communication during handover. In this study, we consider transport layer approach since it require no change to network infrastructure. The change/modification is only required in end nodes.

Chapter 3

Proposed Multi-homed Handover Management Architecture

Recent mobile devices often integrate/embed several wireless technologies. These mobile devices can access the Internet ubiquitously since a permanent Internet connectivity is now required by some applications. Unfortunately, there is no network interfaces assuring global scale connectivity. Nodes must thus use various types of network interfaces to obtain wide area network connectivity [15].

Beside ubiquitous access, recent mobile devices also integrate several access technologies in order to increase bandwidth availability or to select the technology the most appropriate according to the type of flow or choices of the user. Basically, each network interface has different cost, performance, bandwidth, access range, and reliability. Users should thus be able to select the most appropriate set of network interface(s) depending on the network environment, particularly in wireless networks which are mutable and less reliable than wired networks. Users should also be able to select the most appropriate interface per communication type or to combine a set of interfaces to get sufficient bandwidth. Therefore, multi-homing function plays an important role to achieve ubiquitous access.

3.1. Multihoming

Multihoming refers to a node that has multiple interfaces or multiple IP addresses. The node can be assigned with more than one IP addresses with same or different domains. There are various types of multihoming including multi-prefixed, multi-interfaced, multi-linked and multi-sited [16]. Multi-prefixed means that multiple prefixes are advertised on the link(s) the node is connected to. Multi-interfaced is MN has multiple interfaces to choose between, on the same link or not. Multi-linked is similar to multi-interfaced but all interfaces are not connected to the same link. Multi-sited means that a node attached to different sites or domain.

3.2. Motivation and Benefits of Multihoming

There are two basic motivations for multihoming [17]. First, multihomimg enhances the session preservation between MN and CN. Since the wireless link is not as stable as a wired link, session preservation is an important issue in the case of mobile networks, particularly, when MN handovers to another network. Secondly, multihoming can share traffic load more efficiently by selecting the best available connection or enabling multiple connections simultaneously.

Furthermore, there are several benefits of multihoming including ubiquitous access, redundancy/fault-recovery, load sharing, load balancing, bi-casting and preference setting [23]. In terms of ubiquitous access, multihoming provides an extended coverage area. Multiple Interfaces bound to distinct technologies can be used to ensure a permanent connectivity is offered. Redundancy/fault-recovery means that multihoming can act upon failure of one point of attachment, i.e., the functions of a network are assumed by secondary system components when the primary component becomes unavailable (e.g. failure). Connectivity is guaranteed as long as at least one connection to the Internet is maintained. Multihoming can support load sharing and balancing in order to spread network traffic load among several networks and balance load between multiple points of attachments. Multihoming can also support Bi-casting (n-casting) that duplicates a particular flow for simultaneous transmission through different routes. It minimizes packet loss typically for real-time communication and burst traffic. It also minimizes delay of packet delivery caused by congestions and achieves more reliable real-time communication than single-casting. For mobile computing, bi-casting is useful not to drop packets when a mobile node changes its interface during communication. Finally, multihoming provide the user or the application or the ISP the ability to choose the preferred transmission technology or access network for matters of cost, efficiency, politics, bandwidth requirement, delay, etc.

When considering these benefits, one has to consider whether these goals can be achieved with transparency or without transparency. Transparency is achieved when switching to a different point of attachment does not cause on-going sessions to break.

Therefore, this dissertation discusses how to preserve on-going session, in particularly, real-time session (VoIP) during handover and proposes some solutions. We believe that multihoming is one of possible solution for achieving seamless handover.

3.3. Handover Management Architecture

We propose an end-to-end multi-homed handover management for 802.11g networks as well as intermingled 802.11g and 802.16e networks. Our proposed Handover Management is implemented on transport layer of MN and obtains effective handover decision criteria from low layer using cross layer approach as shown in Fig. 3.1. The handover management switches between single-path and multi-path transmission modes in response to wireless network condition of both 802.11g and 802.16e. The handover management appropriately selects the path according to handover decision criteria.

In our study, we consider an MN has two interfaces and each interface is assigned single IP address with different IP subnet. As shown in Fig. 3.1, an MN has dual interfaces connected to two access point. The MN establishes VoIP call with the correspondent node (CN). One or both endpoints (CN or MN) can be multi-homed and multi-addressed but we assume that only MN is multi-homed in this study.

3.4. Single-Path and Multi-Path Transmission

Since the MN has dual interface, MN can transmit the packet through either one or both interfaces. When an MN communicates with a CN using only one and both interfaces, we called single-path and multi-path transmission mode, respectively. Although multi-path transmission generate redundant packets by sending duplicated packets to a CN through two interfaces but it is useful for supporting soft-handover. In this dissertation, we use terminology of single-path and multi-path transmission, instead of terminology of single/bi-casting used by [23] as describe in section 3.2.

In this study, the main purpose of employing single-path/multi-path transmission is that MN may use its two interfaces either alternatively or simultaneously in order to achieve seamless handover. MN has a choice to select the best interface in order to maintain on-going session while changing the point of attachment. Path selection is made based on some metrics/criteria that indicate the wireless link condition or network congestion.

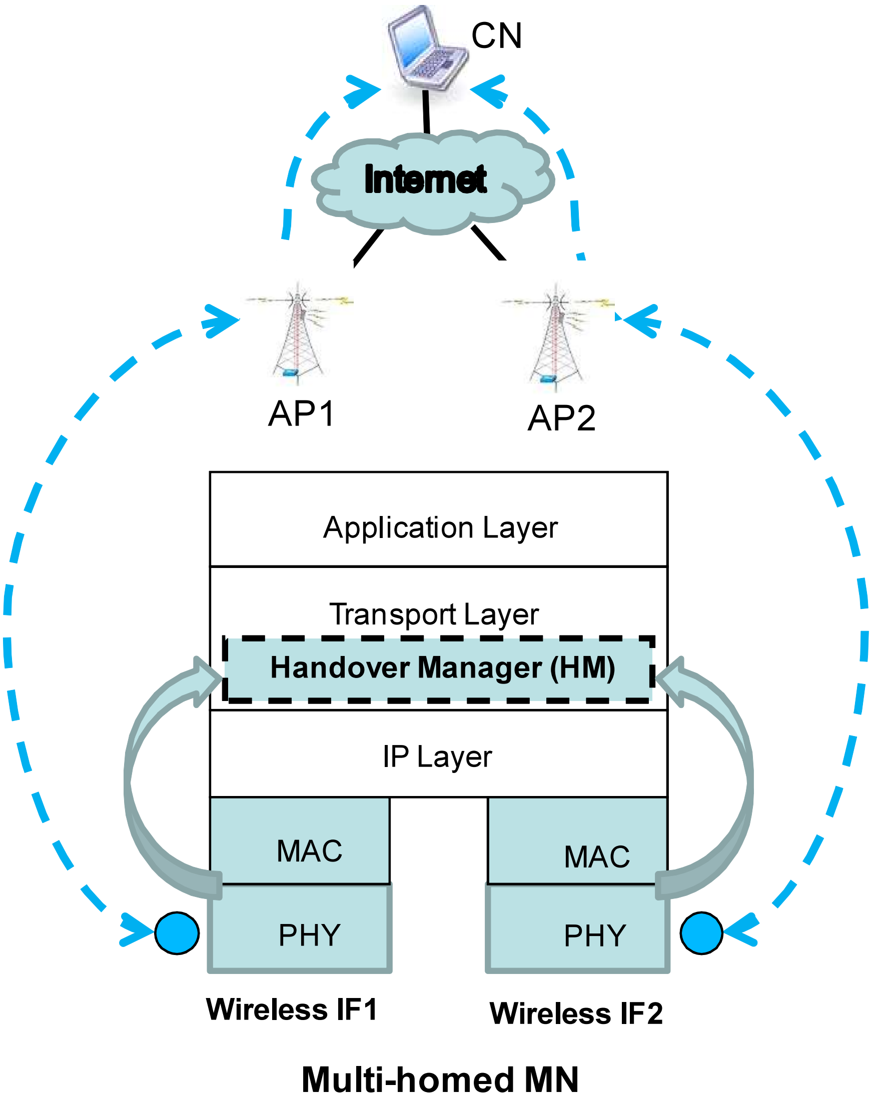

Figure 3.1. Proposed Handover Management Architecture

Chapter 4

Handover Management for VoIP over 802.11g Networks

4.1. Introduction

Wireless LAN (WLAN, IEEE802.11a/b/g/n) has been the dominant wireless technology and is extensively deployed today. Meanwhile, there is a huge demand for Voice over IP (VoIP) service over WLANs. However, delivering VoIP over WLANs (VoWLANs) has many challenges because VoIP is a delay and packet loss sensitive application. In some metropolitan areas, WLANs (WiFi hotspots) have already provided Internet connectivity to mobile nodes (MNs) in many locations. In such an environment, the MNs are likely to traverse several WLANs with different IP subnets during a VoIP call because the coverage of an individual WLAN is relatively small. Consequently, VoWLAN quality could be drastically degraded due to the severe wireless network condition caused by the movement and increase of MNs. Therefore, to preserve VoWLAN quality, MNs need to appropriately and autonomously execute handovers in response to the wireless network condition.

In such a mobile environment, typically, two main factors degrade VoWLAN quality: (1) degradation of wireless link quality and (2) congestion at an AP. First, as an MN freely moves across WLANs, the communication quality degrades due to the fluctuation of wireless link condition. Second, as VoIP is a bi-directional communication, an AP becomes a bottleneck with the increase of VoIP calls. That is, VoIP packets to MNs are liable to experience large queuing delay or packet loss due to increase in queue length or buffer overflow in the AP buffer because each MN and AP has almost the same priority level of frame transmission by following the CSMA/CA scheme. In addition, in multi-rate

WLANs, although a rate adaptation function changes the transmission rate in response to wireless link condition, a low transmission rate occupies a larger amount of wireless resources than that of a high transmission rate. Thus, compared with a high transmission rate, a low transmission rate tends to cause a congestion at an AP. Therefore, to preserve VoWLAN quality, we need to develop a handover strategy considering these two factors in WLANs.

So far, many researchers have studied handover strategies. Although most of them focus on the mechanism for switching wireless networks, they do not sufficiently study a handover strategy considering both wireless network condition and characteristics of an application. In a bi-directional real-time communication such as VoIP, packets routed to MN and queued in the AP buffer experience queuing delay or packet loss, thereby resulting in degradation of VoIP quality for MN. However, common APs, which are already widespread, do not have a mechanism to report the congestion state to MNs. Thus, MNs need to estimate the occurrence of the congestion at the AP for avoiding degradation of VoIP quality.

In this chapter, first, we study a way of estimating AP queue length at an MN side to detect the congestion in a WLAN. Then, we propose a new handover strategy method considering wireless network conditions, i.e., the deterioration of wireless link condition and congestion at the AP. Finally, we show the effectiveness of our proposed method through simulation experiments.

4.2. Related Work

Many handover decision strategies have been studied for various layers of the protocol stack where network and transport layers are most widely studied. Mobile IP [13] is a network layer scheme utilizing and relying on network infrastructures including Router advertisement, Home Agent (HA) and Foreign Agent (FA). However, a handover process in Mobile IP takes a significant time period including the period for acquisition of the IP address in a new WLAN and registration request to an HA and a CN. For example, the layer 2 handover period is 50-400 ms [18], acquisition of IP address from DHCP takes about 300 ms, and registration request is one way delay. Therefore, it is clear that aggregation of interruption period of layer 2 and 3 contributes to deterioration of VoIP communication quality. Moreover, although FMIPv6 [19] and HMIPv6 [20] have been proposed to reduce the handover processing period, they are difficult to deploy in WLANs administrated by different organizations. This is because they require additional network element such as the HA that introduce a burdensome administration and require additional cost. Then, we consider the end-to-end basis approach, which is not require any change of network infrastructure.

On the transport layer, mobile Stream Control Transmission Protocol (mSCTP) [21], which is a mobility extension of SCTP, has been proposed. Although mSCTP supports multi-homing and dynamic address reconfiguration for mobility, the issue of the handover decision is not discussed in detail. The authors in [22] proposed an SCTP based handover scheme for VoIP using a MOS [52] as a handover decision criterion. The handover mechanism employs a probe message called a heartbeat to estimate a Round Trip Time (RTT) and then calculates MOS value based on the RTT. However, since upper layer (above layer 3) information such as packet loss, RTT, and MOS indicate end-to-end communication quality, the information is varied due to both the wireless and wired networks. Therefore, the existing studies could cause unnecessary handovers due to temporal congestions in wired networks.

In a mobile environment, MNs need to promptly and reliably detect wireless link condition. Our practical experiments in [25] proved that the number of frame retries on the MAC layer has the potential to detect the wireless link degradation during movement because packets over WLAN inevitably experience frame retries before being treated as packet loss. Reference [26] proposed a handover mechanism employing the number of frame retries as a handover decision criterion through analytical study. This method, however, only considers the frame retransmission caused by the collision with frames transmitted from other MNs in a non-interference environment. On the other hand, we proposed a handover strategy considering the number of data frame retries on the MAC layer [27] [28] [29]. This strategy employs multihoming enabling to execute multi-path transmission mode for supporting inter-domain soft-handover between two WLANs with different IP subnets. However, although our previous method can detect the degradation of wireless link condition due to both movement of MN and radio interference, it cannot detect congestion at both serving AP and targeted AP. This is because our previous method detects wireless link condition based on only data frame retries without considering congestion at both serving AP and targeted AP. Therefore, in our previous method, an MN could execute a handover to a congested AP as well as leads to imbalanced traffic load among APs, thus, VoIP quality would be degraded.

We need a handover strategy considering congestion of AP and the load balancing among the APs. In [30], authors proposed a decentralized AP selection strategy to achieve a load balancing among the APs by exploiting the packet error rate that can be obtained

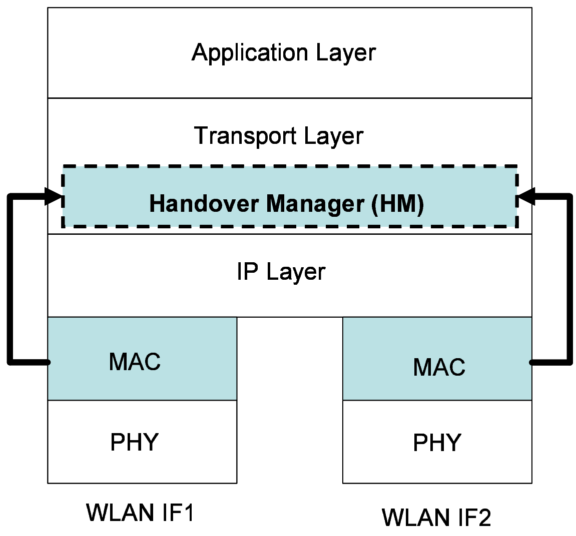

Figure 4.1. Proposed Handover Architecture

from the RSS. However, this strategy only considers TCP traffic and it requires a slight modification in both MN and AP. We consider a handover strategy based on end-to-end basis for real-time application and the handover strategy aim no modification of network infrastructure such as AP.

In this paper, we propose a handover strategy considering congestion of both targeted AP and serving AP in addition to wireless link condition in order to avoid VoIP quality degradation and achieve a load balancing among APs.

4.3. Proposed Handover Decision Criteria

We discuss handover decision criteria that can precisely indicate wireless network condition. In particular, many handover technologies employ the RSS on PHY layer as a handover decision criterion. However, our previous research [25] shows that RSS is very difficult to use to properly detect deterioration in communication quality because it fluctuates abruptly due to distance and interfering objects. It also cannot detect the degradation due to radio interference. Furthermore, in [25], we showed that the information of the MAC layer, i.e., frame retry has the potential to serve as a significant metric. However, it cannot satisfactorily detect the wireless networks condition. In this section, we then describe the following three handover decision criteria employed in our new proposed method.

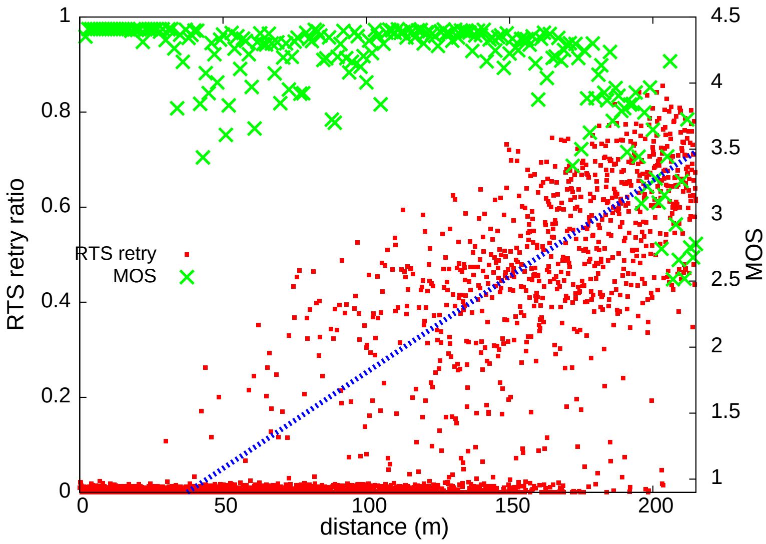

Figure 4.2. RTS Retry Rate vs. MOS over Distance

4.3.1 RTS Frame Retries

In the IEEE802.11 standard, a sender confirms a successful transmission by receiving an ACK frame in response to the transmitted data frame. When a data or ACK frame is lost, the sender periodically retransmits the same data frame until achieving a successful transmission or reaching a predetermined retry limit. The standard supports two retry limits: long-frame and short-frame retry limits. In addition, the standard also includes the Request-to-Send (RTS)/Clear-to-Send (CTS) function to prevent collisions caused by hidden nodes. If RTS/CTS is applied, a long-frame retry limit of four is applied, otherwise, a short-frame retry limit of seven is applied. When frame retries reach the retry limit, the sender treats the data frame as a lost packet. That is, we can detect the occurrence of packet loss in advance by utilizing the frame retries. Moreover, unlike the RSS, frame retries can promptly and reliably detect the wireless link degradation due to not only reduction of signal strength but also radio interference and collisions [25]. Therefore, frame retry allows an MN to detect wireless link condition promptly and reliably.

In [27], we employed data frame retry as a handover decision criterion in WLANs with a fixed transmission rate (11 Mb/s). However, in a real environment, almost all WLANs employ a multi-rate function that can change the transmission rate according to wireless link condition. If the transmission rate is dropped by the multi-rate function, a more robust modulation type is selected and thus data frame retries are further decreased. As a result, an MN cannot properly detect the degradation of wireless link quality only from

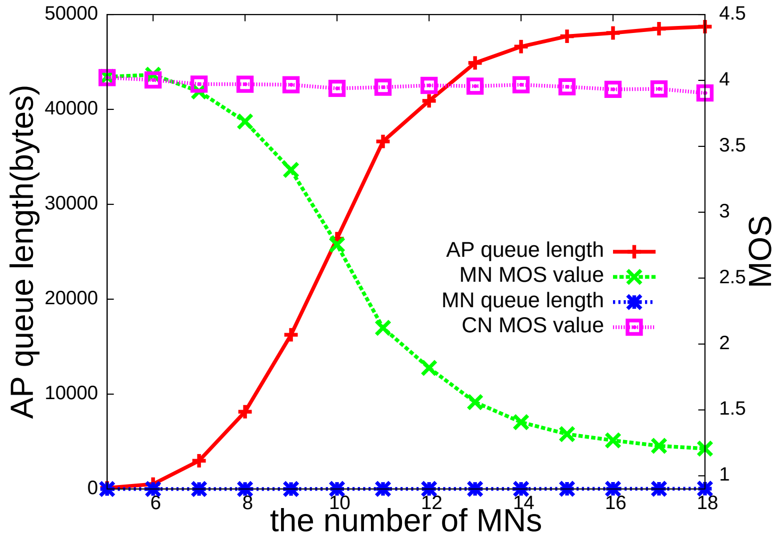

Figure 4.3. Relationship among # of MNs, AP queue length, and MOS

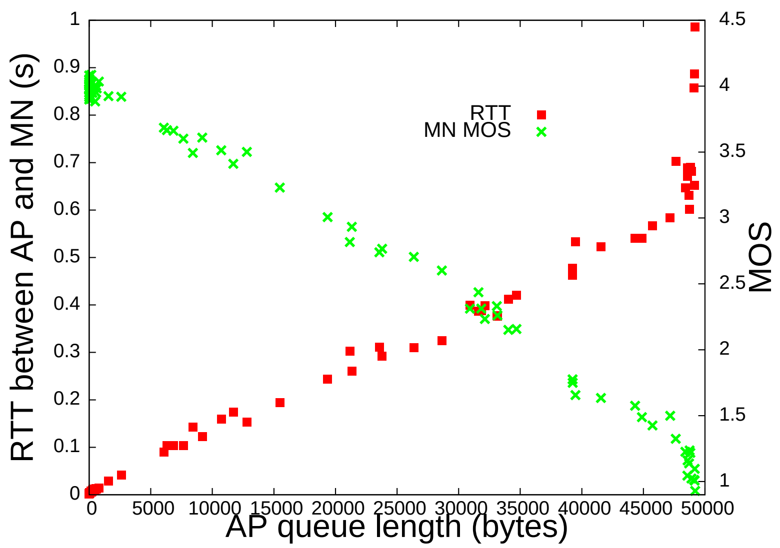

Figure 4.4. Relationship among AP queue length, W-RTT, and MOS

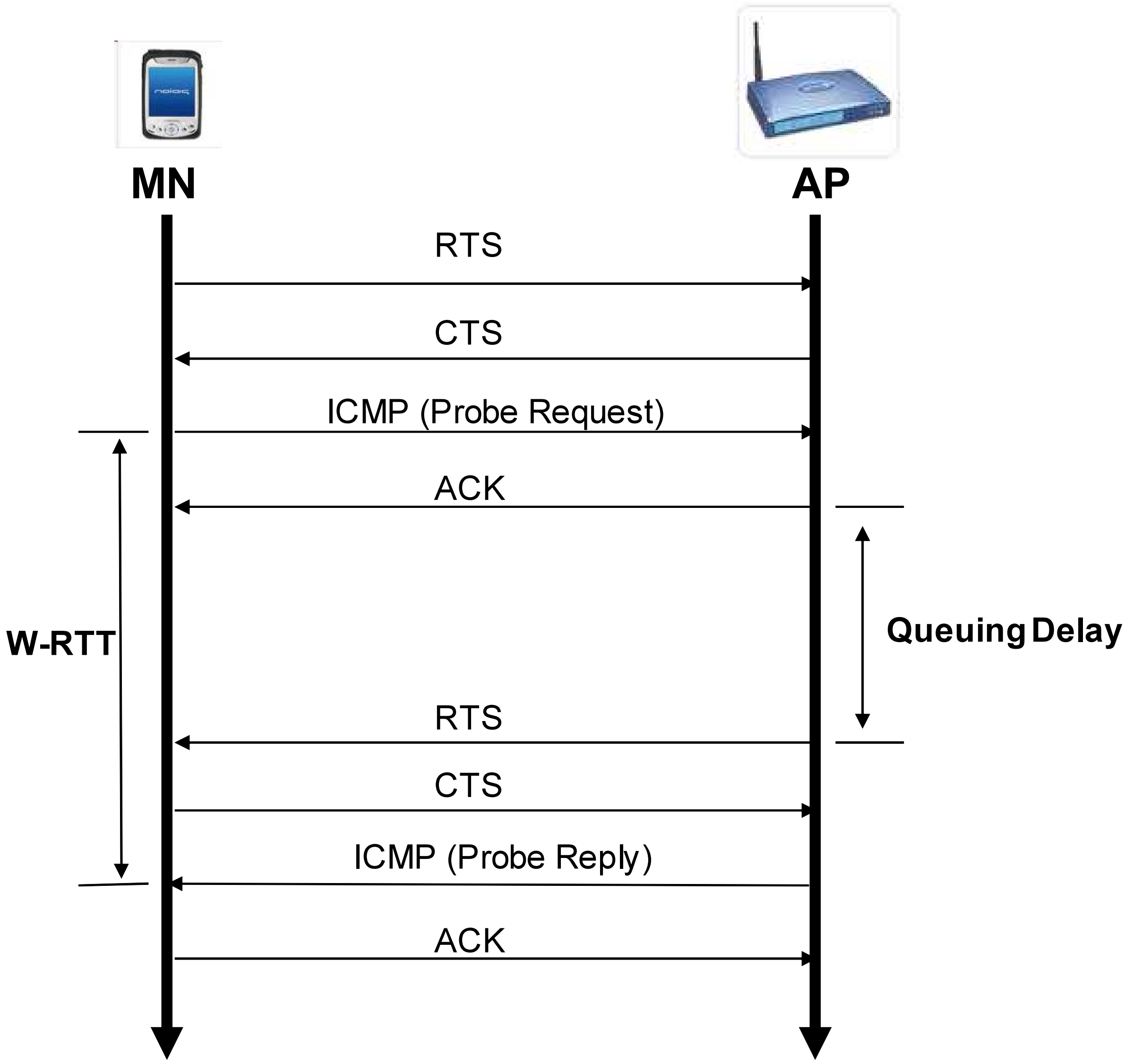

Figure 4.5. RTT between AP and MN (W-RTT)

data frame retries in multi-rate WLANs.

We then consider an RTS frame as an alternative metric of data frame retries. As an RTS frame is always transmitted at the lowest rate (e.g., 6 Mb/s in 802.11a/g and 1 Mb/s in 802.11b), an MN can appropriately detect the change of wireless link quality. Moreover, RTS frame is basically employed to prevent collisions in wireless network due to hidden nodes. However, according to the IEEE802.11 standard, as RTS threshold is 2347 bytes by default, thus, RTS is not sent in case of VoIP packet size (160 bytes). Therefore, in our proposal, all MNs must set RTS threshold to 0 in order to enable the MNs send the RTS frame. To show the effectiveness, we investigate the behavior of the RTS retry ratio when an MN moves away from an AP through simulation experiments using Qualnet 4.0.1 [55]. Figure 4.6(a) and Table 5.1 show a simulation model and parameters, respectively. Note that we employ MOS [52] to assess the VoIP quality. In our study, RTS retry ratio is employed instead of the frequency of RTS retries. The RTS retry ratio is calculated as follows:

$$RTS\ Retry\ Ratio = \frac{Number\ of\ RTS\ Frame\ Retries}{Total\ Transmitted\ Frames} \tag{4.1}$$

The number of RTS frame retries and the total transmitted frames are calculated at every distance point.

Figure 4.2 shows the relationship between the MOS and RTS retry ratio as a function of distance between the AP and the MN. Note that MOS of more than 3.6 indicates an adequate VoIP call quality. From Fig. 4.2, we can see that the MOS is degraded with the increase in the RTS retry ratio when the MN moves away from the AP. Since the RTS retry ratio is varied due to the fluctuation of wireless link quality, we employ a least-squares method to grasp their trend and estimate the best fit of the occurrences of RTS retry ratio over the distance, shown as a straight line. The straight line shows that the RTS retry ratio of 0.6 indicates the starting point of VoIP quality degradation. Therefore, we set the RTS retry ratio of 0.6 as one of the thresholds to execute the handover in this study.

4.3.2 AP Queue Length

With the increase of VoIP calls in a WLAN, the AP queue length increases because all traffic to and from WLAN has to go through the AP. The 802.11 CSMA/CA mechanism provides the AP and MNs with the same number of transmission opprtunities. Suppose we have N_{mn} number of MNs transmitting upstream traffic, the MNs have a $N_{mn}/(N_{mn}+$

AP) share of bandwidth while the AP has only a $1/(N_{mn} + AP)$ share bandwidth for transmitting downstream traffic. Moreover, the AP typically needs to send more packets than MN, hence, each packet routed to MN (downlink transmission) and queued in the AP buffer may experience a large queuing delay or packet loss due to asymmetry of bandwidth share. Consequently, the queuing delay and the packet loss due to buffer overflow severely affect the VoIP quality of MNs. Unfortunately, the IEEE802.11 (a/b/g/n) standard does not provide a mechanism that can inform MNs of the AP queue length. Therefore, to preserve VoIP quality, an MN needs to detect the congestion of the AP by itself.

We then investigate the relationship between the number of MNs (VoIP calls) and AP queue length through simulation experiments. Figure 4.6(b) and Table 5.1 show a simulation model and parameters, respectively. In the simulation scenario, we randomly locate from one to 18 MNs in a WLAN. Each MN communicates with a CN using VoIP. Figure 4.3 shows the relationship among the number of MNs, AP queue length, and MOS. From Fig. 4.3, we can see that although VoIP quality of CNs (CN MOS value) is kept adequate even if VoIP calls increase, the MN MOS value adversely degrades.

From Fig. 4.3, we found the significance of the AP queue length. However, how can MNs detect AP queue length without modifying an AP? Therefore, we propose a method to estimate AP queue length based on RTT between MN and AP. Note that in this paper, the RTT between MN and AP is called Wireless RTT (W-RTT). As illustrated in Fig. 4.5, the MN periodically sends a probe packet (ICMP message) to an AP and then calculates W-RTT. The W-RTT increases in response to the increase of AP queuing delay because a probe response packet experiences queuing delay in the AP buffer. Therefore, the W-RTT can be used to derive information about AP queuing delay.

We then investigate the relationship among MOS, AP queue length and the W-RTT using the simulation model in Fig. 4.6(b). From Fig. 4.4, we can see that the W-RTT increases with the increase of AP queue length and the AP queue length of less than 7,500 bytes satisfies an adequate VoIP call. The graph also shows that the W-RTT should be kept under 200 ms to satisfy adequate VoIP quality. Therefore, in our proposed method, we employ W-RTT to estimate AP queue length and set the W-RTT threshold ($W\text{-}RTT\_thr$) of 200 ms to maintain the adequate VoIP quality.

4.3.3 Transmission Rate

IEEE 802.11 supports a rate adaptation function that can dynamically and automatically change the transmission rate based on wireless link condition. In the case where wireless link quality degrades, as the transmission rate decreases caused by the change of the

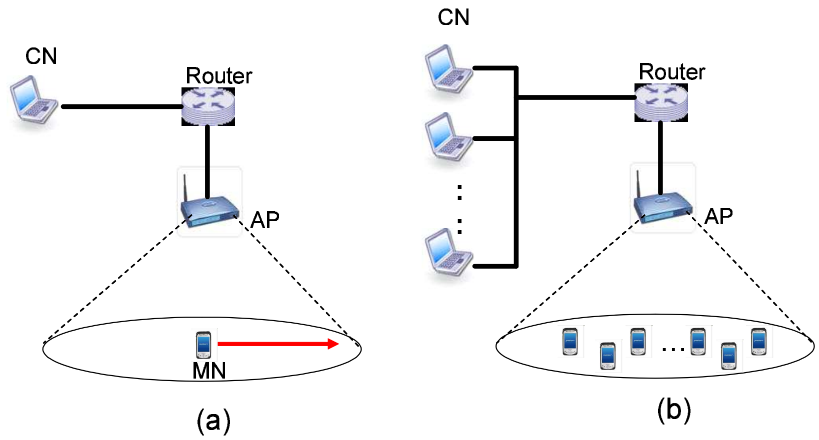

Figure 4.6. Simulation Model 1

Table 4.1. Simulation Parameters 1

VoIP Codec	G.711
WLAN Standard	IEEE802.11g
Supported Data Rate (Mbps)	6 9 12 18 24 36 48 54
Tx Power (dBm)	5.0 5.1 6.7 7.9 8.1 9.3 10.6 10.1
Fading Model	Nakagami Ricean K = 4.84
SIFS	16 μs
Slot Time	9 μs
CWmin, CWmax	15, 1023

modulation type, the wireless resource is more occupied because of the long transmission delay. As a result, the lower transmission rate is likely to cause congestion of an AP. Therefore, to alleviate congestion of an AP, the transmission rate can also be treated as a potential handover decision criterion.

4.4. Proposed Handover Management

In our study, we employ RTS frame retry, estimation of AP queue length (W-RTT), and transmission rate as handover decision criterion. We then propose a handover strategy method based on [27]. In [27], an MN has two WLAN interfaces (IFs), and a handover Manager (HM) implemented on transport layer controls handover based on the handover decision criteria obtained through cross layer approach (see Fig. 4.1).

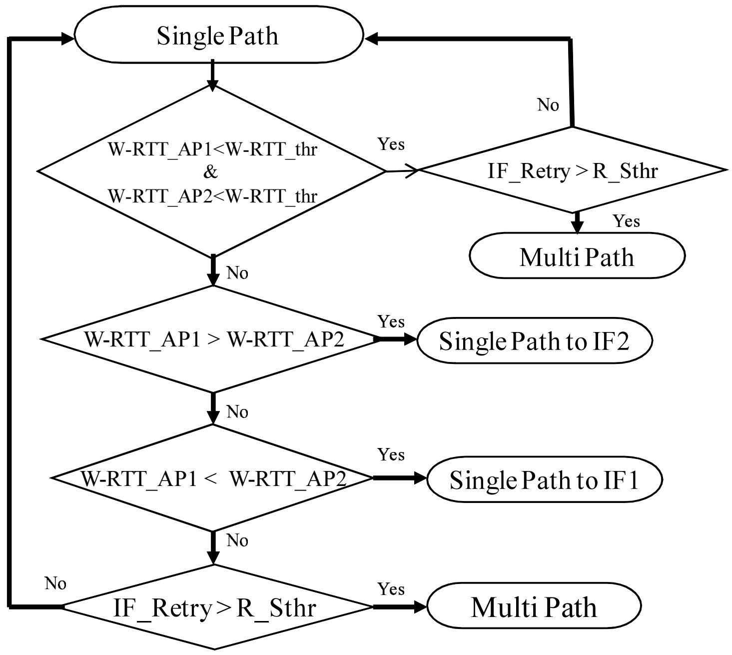

Figure 4.7. Switching to Single/Multi-Path Transmission

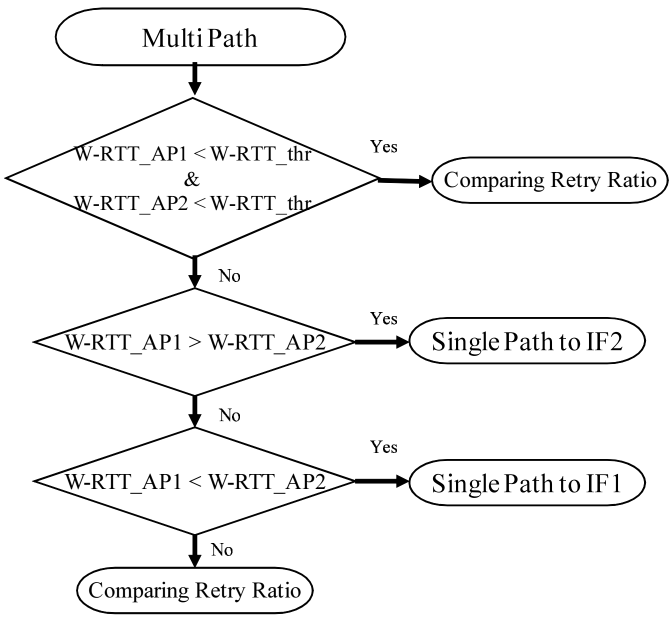

Figure 4.8. Switching from Multi-Path Transmission to Single-Path Transmission

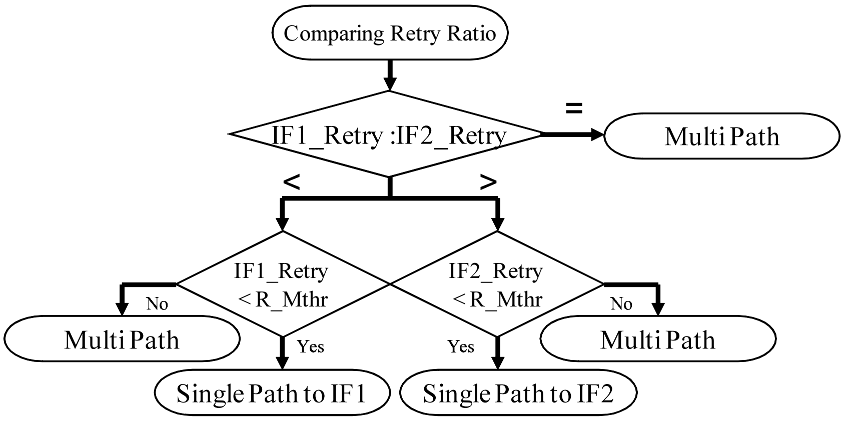

Figure 4.9. Handover based on RTS Retry Ratio

4.4.1 Single-Path and Multi-Path Transmission

Our proposed handover method employs multi-homing similar to [27] in order to support soft-handover. HM properly switches between single-path and multi-path transmission modes in response to wireless network condition. Single-path transmission mode means that an MN communicates with a CN using only one IF. Multi-path transmission, on the other hand, means that an MN sends duplicated packets to a CN through two IFs for supporting soft-handover.

Figure 4.7 shows an algorithm of switching to single/multi-path transmission when an MN moves into an overlap area of two APs (AP1 and AP2). An MN associated with two APs transmits a probe packet to both APs at 500 ms intervals to estimate AP queue length of each AP. If both W-RTT values are below an W-RTT threshold (*W-RTT_thr*), an MN detects that both APs are not congested. Then, the MN investigates the RTS frame retry rate of the current (single) active IF since its movement also affects wireless link condition. If the RTS frame retry ratio reaches a retry ratio threshold of single-path (*R_Sthr*), an HM switches to multi-path mode to investigate the wireless link condition of both IFs as well as supporting soft-handover. On the other hand, if the W-RTT of AP1 reaches *W-RTT_thr*, i.e., detection of congestion at AP1, an MN switches to the AP2 directly without switching to multi-path mode because multi-path mode may cause more serious congestion in WLANs. Finally, if both measured W-RTTs reach *W-RTT_thr*, an MN then investigates the wireless link condition by using the RTS frame retry ratio of the active single IF.

In multi-path transmission, to preserve VoIP quality, an MN sends the same data

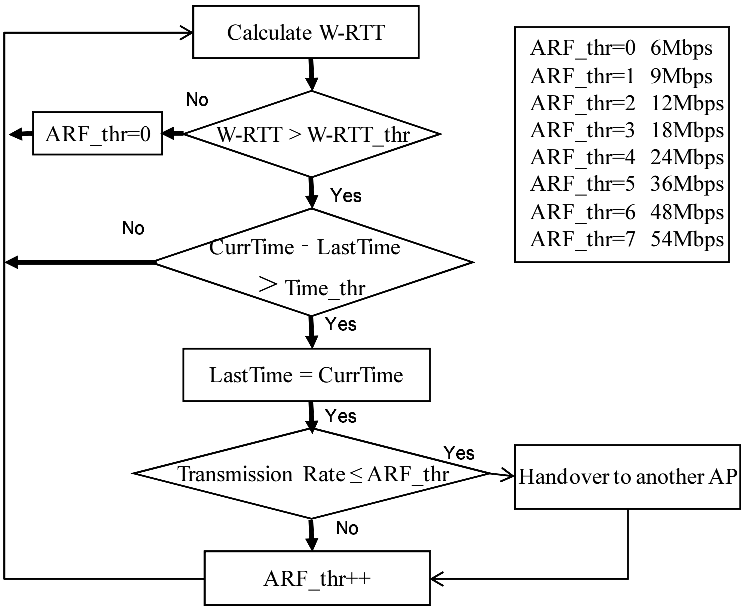

Figure 4.10. Handover based on Transmission Rate

packets to both IFs. Hence, the MN needs to switch back to single-path transmission as soon to prevent redundant network overload. As shown in Fig. 4.8, an algorithm of switching to single-path transmission works as follows. An MN measures W-RTTs of both APs at all times. If either of the W-RTTs is below the $W\text{-}RTT\_thr$, the MN switches to an IF with the smaller W-RTT. If both W-RTTs are simultaneously below the $W\text{-}RTT\_thr$, the MN then compares the RTS frame retry rate of both IFs. Figure 4.9 shows an algorithm to compare RTS frame retry ratios of both IFs. If both frame retry ratios of the IFs are equal, the MN continues multi-path mode. On the other hand, if either of the frame retries is below the ratio threshold of multi-path (R_Mthr), the MN switches to single-path mode through the IF with the smaller frame retry ratio. Note that, this proposed method is further called the basic method.

4.4.2 Dealing with Ping-Pong Effect (Extension Method 1)

If all MNs send probe packets to measure the W-RTT in the basic method, the MNs may unfortunately detect congestion of the serving AP (e.g., AP1) at nearly the same time. Then, all MNs may switch the communication to a neighbor AP (e.g., AP2) and leave AP1. As a result, the neighboring AP2's queue length is drastically increased, and then, all MNs switch back to AP1 again. This phenomenon is typically known as the ping-pong

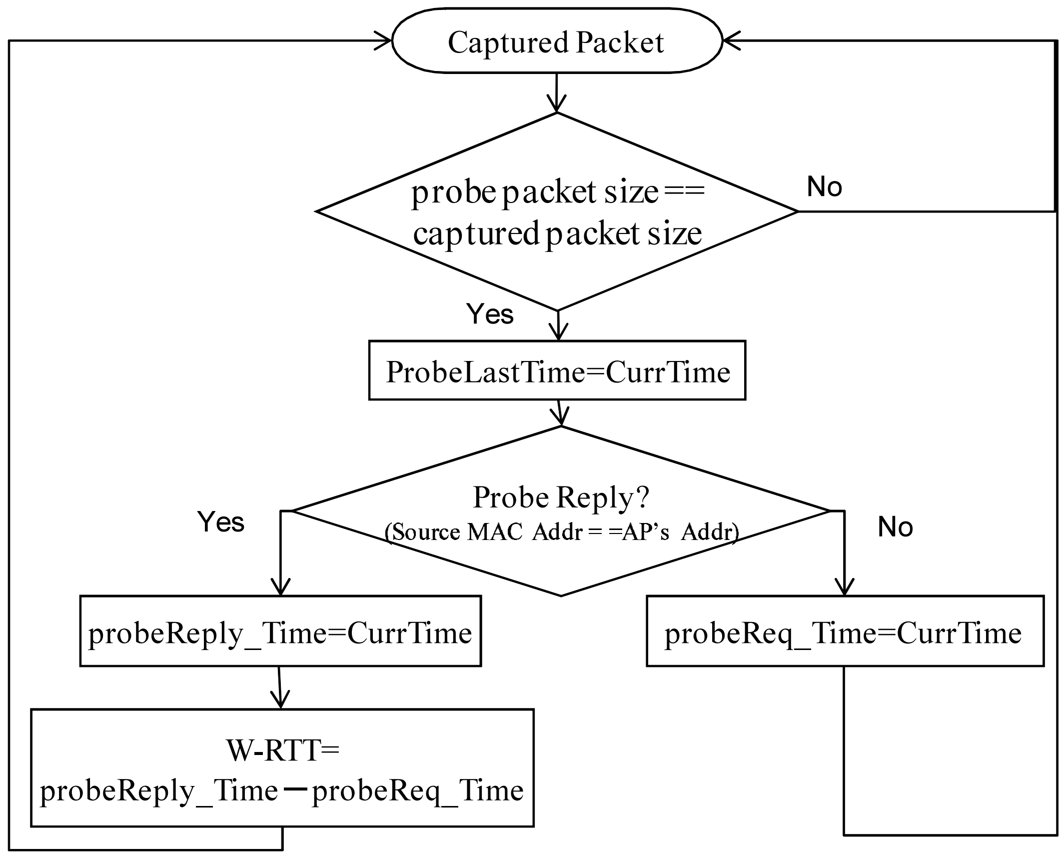

Figure 4.11. Calculate W-RTT from existing probe packet

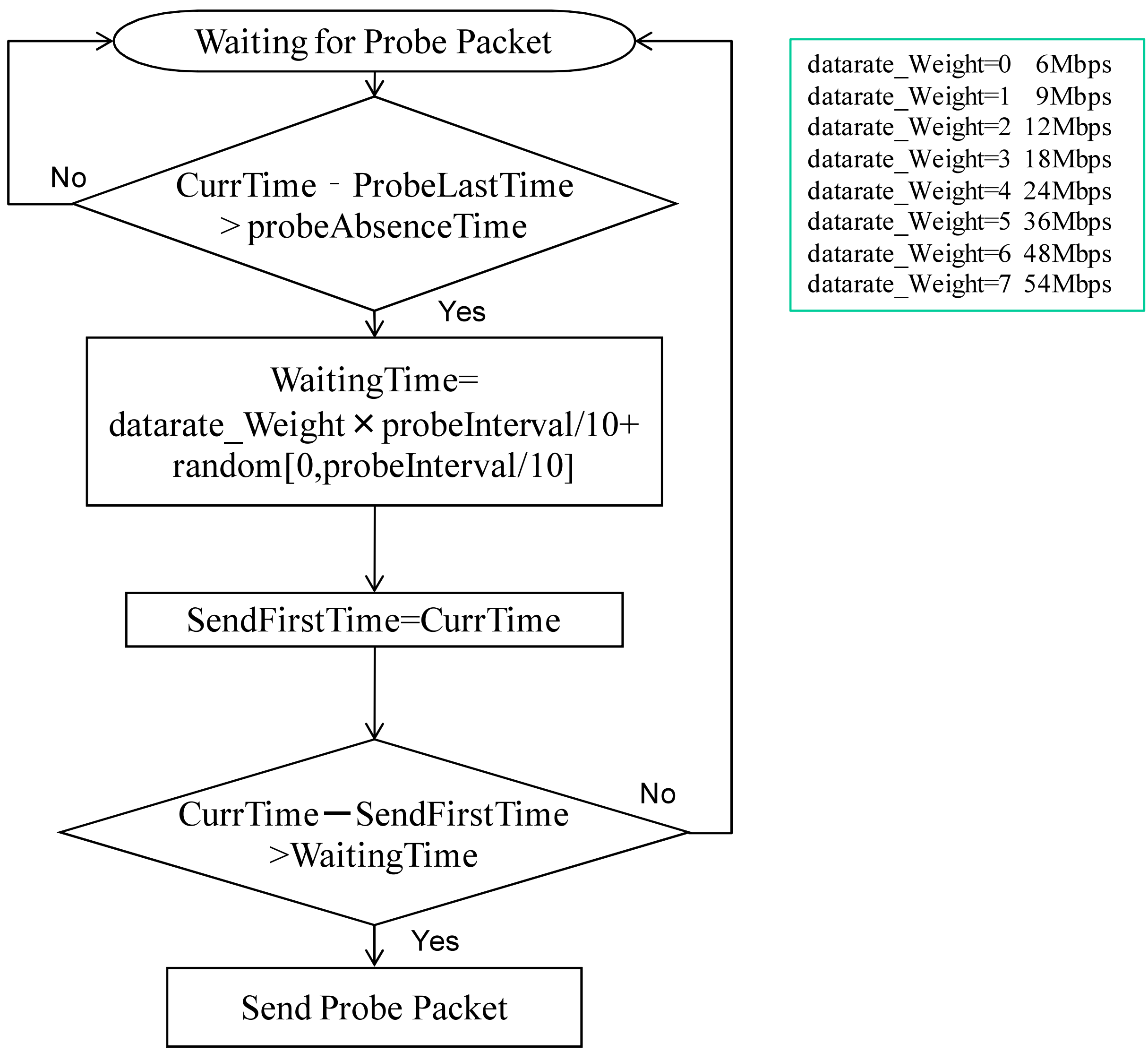

Figure 4.12. Obtaining a right to send the probe packet

Table 4.2. System Parameters

$WRTT\_thr$	200 ms
R_Sthr	0.6
R_Mthr	0.4
$probeAbsenceTime$	1 second
$Time\_thr$	2 second

effect and leads to degradation of all VoIP quality due to fluctuation of both APs queue length.

To avoid the ping-pong effect, we extend the basic method. In extension method 1, to avoid the simultaneous handover among all MNs, they first examine their own current transmission rate before executing handover. Figure 4.10 shows a handover algorithm based on transmission rate. A WLAN provides a multi-rate function that can change the transmission rate dynamically based on wireless link condition. As mentioned earlier, since an MN with lower transmission rate occupies a large amount of wireless resources, the MN is liable to lead to congestion of an AP. Moreover, as MNs with the lowest transmission rate typically are far away from the connected AP, that is, near the edge of its coverage, they must execute handover as soon as possible to preserve their communication quality. Therefore, in extension method 1, MNs with the lowest transmission rate (6 Mb/s) first execute handover. Then, if the AP queue length is still high even after $Time\_thr$, MNs with the next lowest transmission rate (12 Mb/s) start to execute handovers. The extension method 1 actually works as follow. When the serving AP is congested, all MNs detect the congested AP through RTT measurement ($W\text{-}RTT > W\text{-}RTT\_thr$). First, all MNs with transmission rate of 6Mb/s (the lowest rate) execute the handover since ARF_thr is set to 0 by default. Note that ARF_thr of 0 indicates 6 Mb/s. After that, if the AP is still congested in $Time\_thr$, the all remaining MNs increase the ARF_thr by one, i.e., the ARF_thr of one indicates 9 Mb/s. Then, MNs with transmission rate under 9 Mb/s execute handover. This handover process is repeated until congestion of the AP is alleviated. If the congestion is alleviated, ARF_thr of all MNs set back to default value of 0. Thus, an MN does not need to know whether another MN with lower transmission rate has executed the handover or not. Therefore, to execute handover, all MNs only monitor their own transmission rate and compare the rate with the current ARF_thr.

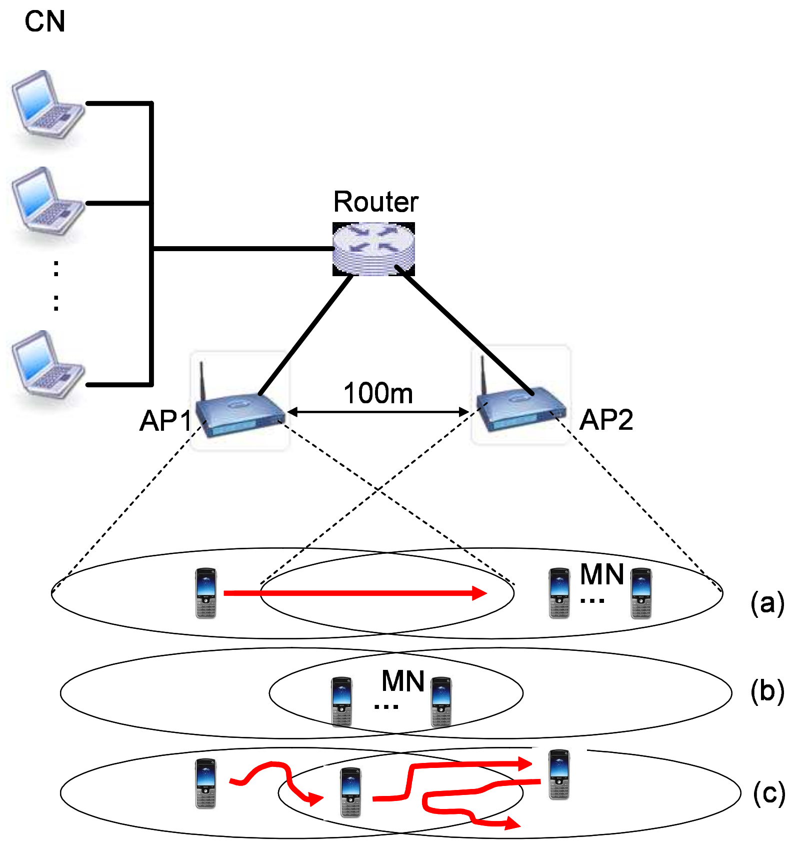

Figure 4.13. Simulation Models 2

4.4.3 Elimination of Redundant Probe Packets (Extension Method 2)

If every MN measures W-RTT by using probe packets according to extension method 1, these probe packets may aggravate congestion in a WLAN. To eliminate the redundant probe packet, we further extend the extension method 1. In extension method 2, only one representative MN sends a probe packet to the AP, and then other MNs measure W-RTT by capturing the probe and probe ACK packets that the representative MN sends and receives, respectively. This method works as follows (see Fig. 4.11).

Each MN first monitors all packets over a wireless link before sending a probe packet by itself. If it receives a probe packet sent by another MN, it cancels the transmission of a probe packet and measures W-RTT by using the probe and probe ACK packets, which another MN exchanges. Each MN can then identify whether a captured packet is a probe packet or not, by observing the ICMP message frame length (64 bytes). Furthermore, an MN can also identify whether a probe packet is for request (ICMP Request) or for reply (ICMP Response) by observing the MAC address of the captured ICMP packet because all MNs connected to an AP can identify the MAC address of the AP. If the *destination MAC address* of the captured packet is that of the AP, each MN can judge the packet as a *probe request* packet. On the other hand, if the *source MAC address* is an AP's, then each MN judges the packet as a *probe reply* packet.

In Fig. 4.11, $probeReq\_Time$ and $probeReply\_Time$ are the receiving time of the probe request transmitted from another MN and that of the probe reply transmitted from the AP, respectively. As every MN can identify whether a captured packet is a probe request or probe reply, it can calculate the W-RTT ($probeReply\_Time - probeReq\_Time$) properly. In this way, this method can eliminate the redundant probe packets because MNs calculate the W-RTT by capturing the probe packets that one representative MN sends.

If the representative MN leaves a WLAN, one of the remaining MNs needs to start periodically sending a probe packet as a representative MN. Here, we describe how an MN obtains the right to send probe packets in Fig. 4.12. First, all MNs always examine the difference between the last receiving time of a probe packet ($ProbeLastTime$) and the current time ($CurrTime$). If the difference is greater than $probeAbsenceTime$, that is, a probe packet can not be captured for a while, first, MNs with the lowest transmission rate in a WLAN try to send a probe packet. This is because a probe packet sent at the lowest transmission rate can be captured by almost all MNs in a WLAN due to its inherently longer transmission range. The timing to send a probe packet among MNs is determined

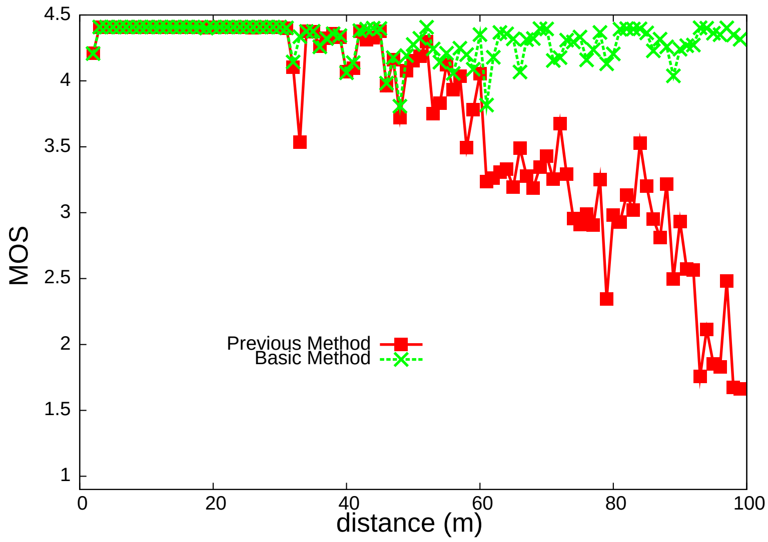

Figure 4.14. MN's MOS over distance

based on $WaitingTime$. Basically, an MN with the smallest $WaitingTime$, will be a representative MN because $WaitingTime$ is calculated based on $datarate\_Weight$, which indicates its weight of transmission rate (see Fig 4.12). Thus, if the $datarate\_Weight$ is lower, then $WaitingTime$ gets small. If several MNs with the same transmission rate are existed, then random value in $WaitingTime$ helps to distinguish who will be the representative MN among them.

4.5. Performance Evaluation

We evaluate our proposed handover schemes through simulation experiments. We implement our proposed methods in Qualnet simulator 4.0.1 [55]. Tables 5.1 and 4.2 show the simulation parameters and system parameters, respectively. In our study, we employ MOS to assess the VoIP quality.

4.5.1 Evaluation of Basic Methods

Figure 4.13 (a) shows a simulation model to evaluate effectiveness of our basic method based on AP queue length and RTS retry ratio. In this simulation, an MN with two WLAN IFs moves from AP1 to AP2 at the speed of 1 m/s. AP2 is assumed to be congested due to existence of fixed 15 MNs establishing VoIP calls. The 15 MNs are randomly located in the AP2 coverage area. That is, each MN may have a different

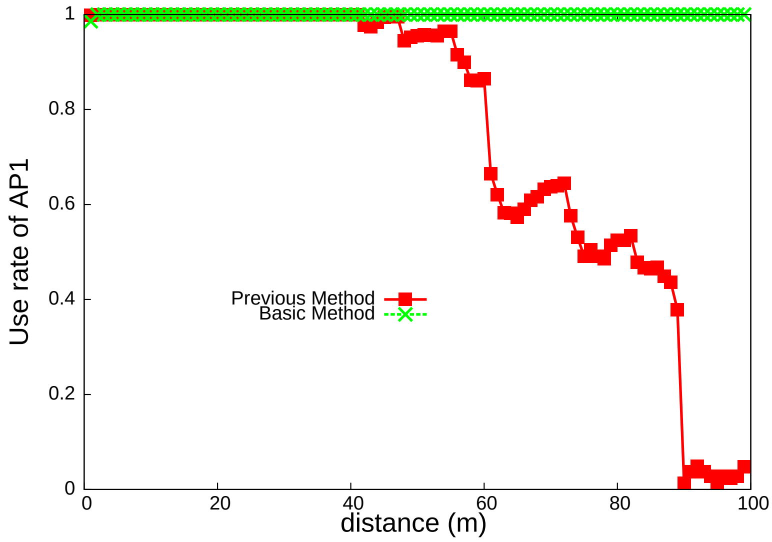

Figure 4.15. Use Rate of AP1 over distance

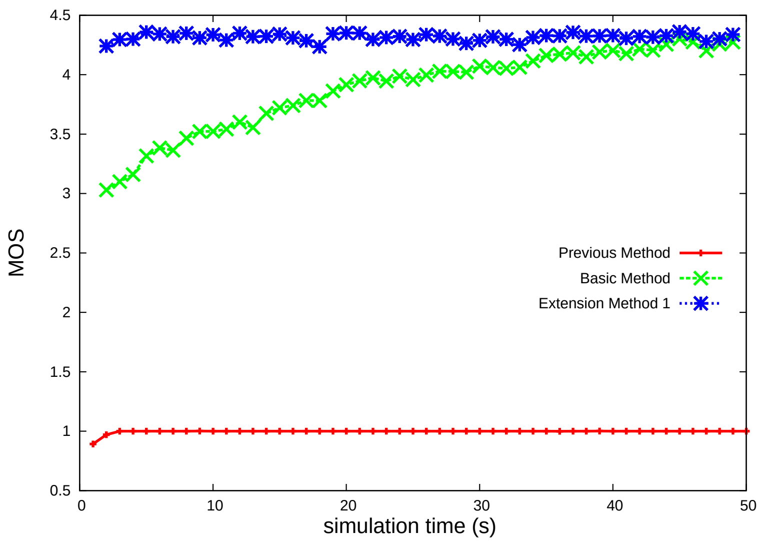

Figure 4.16. Variation of MN's MOS

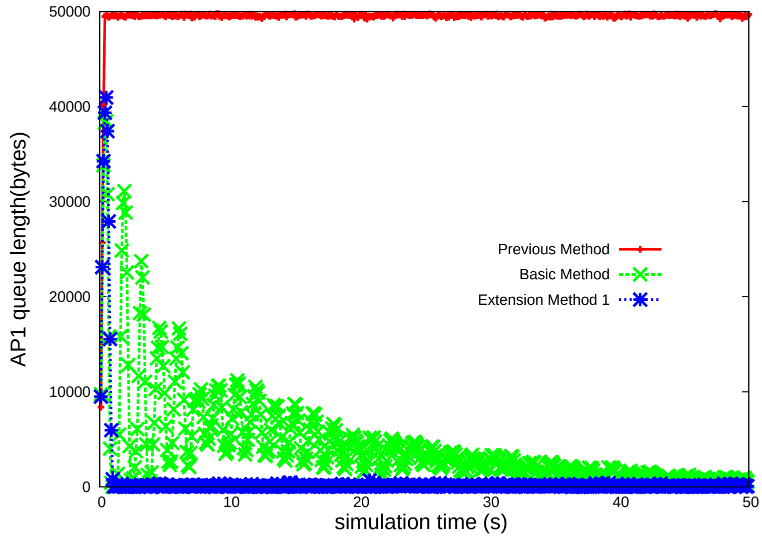

Figure 4.17. Variation of AP1 queue length

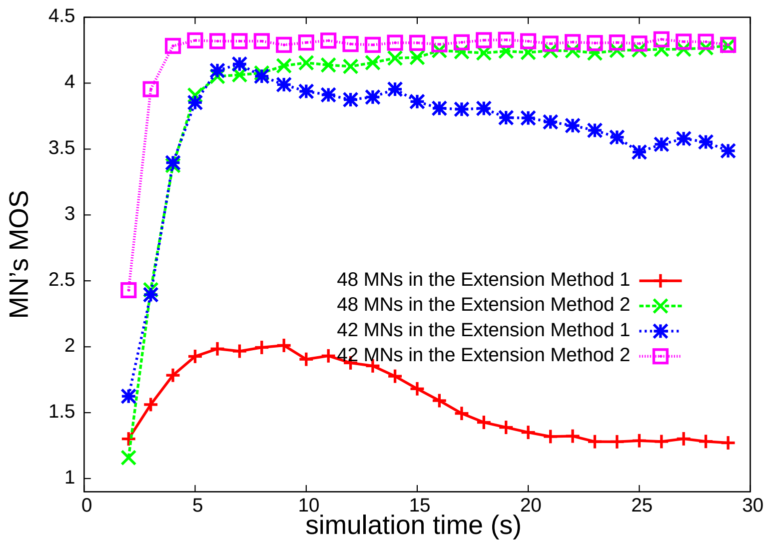

Figure 4.18. Variation of MN's MOS

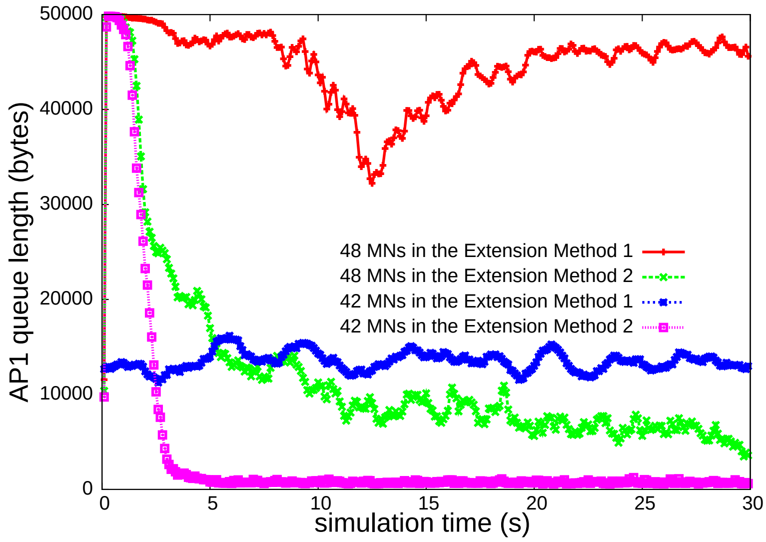

Figure 4.19. Variation of AP1 queue length

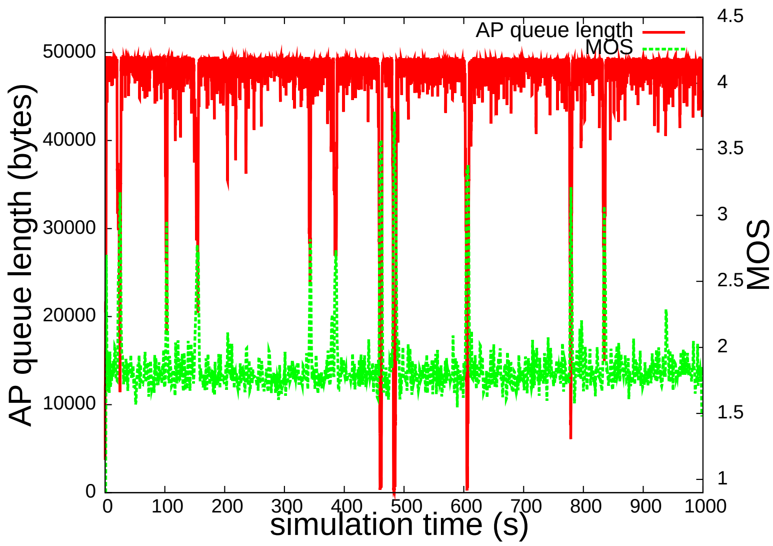

Figure 4.20. Variation of AP1 queue length of our previous method

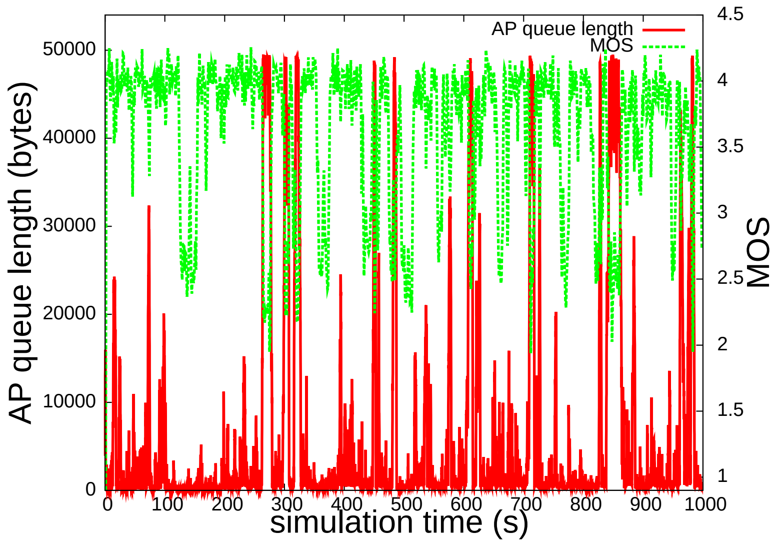

Figure 4.21. Variation of AP1 queue length of the extension method 2

transmission data rate according to its location. We employ a G.711 VoIP codec that sends a 160-byte packet at 20-ms intervals. We execute 200 simulations with different seeds, and each simulation time is 100 seconds. Thus, our simulation results indicate the average value of the simulations.

To show the effectiveness of estimation of AP queue length, we compare our basic method with our previous method based on only the number of data frame retries [27]. Figures 4.14 and 4.15 show MOS variation of MN and use rate of AP1 over the distance between the MN and the AP1. In the basic method, we can see that the MN can preserve VoIP quality because it detects congestion in the AP2 adequately and avoids handover to the AP2. On the other hand, although our previous method executes handover to AP2 when the MN reaches about 40 m, the quality after the handover degrades because of the congestion of AP2. That is, our previous method cannot detect the congestion in a WLAN. Therefore, our basic method can execute handovers considering the wireless network condition of both AP1 and AP2.

4.5.2 Evaluation of Extension Method 1

In the next evaluation, we show that our extension method 1 can avoid the ping-pong effect problem. Figure 4.13 (b) shows a simulation model where 20 MNs are randomly located within an overlap area between AP1 and AP2. Here, we compare three methods, i.e., the previous method, the basic method, and the extension method 1. At the start of simulation, every MN establishes a VoIP call with its peer CN through AP1.

Figures 4.16 and 4.17 show variation of MN's MOS and AP1 queue length, respectively. As can be seen from Fig. 4.17, the previous method cannot detect the congestion of AP1 and it continues to utilize the AP1 without executing handover to AP2, hence, AP queue length keeps high as well as adequate VoIP quality cannot be maintained during simulation time. Moreover, the basic method causes a significant fluctuation of AP queue length at initial duration due to simultaneous handovers of all MNs. The basic method can get adequate VoIP quality after gradually decreasing AP queue length. On the other hand, our extension method 1 can preserve VoIP call quality throughout the simulation time because MNs with extension method 1 execute handover in order of the transmission rate. Hence, extension method 1 can avoid the degradation due to the ping-pong effect.

4.5.3 Evaluation of Extension Method 2

We evaluate the effectiveness for the reduction of the redundant probe packets in extension method 2. We employ two simulation scenarios in Fig. 4.13 (b). In the first scenario, 42 MNs are randomly located within the overlap area. On the other hand, the second scenario has 48 MNs located randomly within the overlap area. Then, at first, all MNs establish VoIP calls with their CNs through AP1.

Figures 4.18 and 4.19 show variation of MN's MOS and AP1 queue length, respectively. From Fig. 4.18, we can see that, in the extension method 2, although the MOSs in both cases, 42 and 48 MNs, are degrading at the beginning of the simulation, the MOSs are recovered soon. On the other hand, in extension method 1, although the MOS of 42 MNs is recovered at one point, after that the MOS is gradually degraded again. In the 48 MNs, the MOS can not be maintained at all. From Fig. 4.19, we can see that in AP1 queue length, although both results of extension method 1 maintain higher queue length, that of extension method 2 gradually degrades. That is, in the extension method 2, as only one representative MN sends probe packets, the reduction of redundant probe packets brings an increase of acceptable VoIP calls. Therefore, extension method 2 can promptly and reliably execute handover considering the congestion of an AP, while avoiding the ping-pong effect and reducing redundant probe packets.

4.5.4 Random Movement Environment

Finally, we evaluate the performance of extension method 2 and our previous method in a random movement environment. As shown in Fig. 4.13 (c), the 15 MNs randomly move between two AP coverage areas at a speed of 1 m/s.

Figures 4.20 and 4.21 show the MOS and AP1 queue length for the previous method and extension method 2, respectively. From Fig. 4.20, the average of AP queue length of the previous method is extremely high and the MOS of MNs does not satisfy adequate VoIP quality at all. On the other hand, in Fig. 4.21, extension method 2 can almost maintain adequate VoIP quality. Also, though some degradation points exists, the MOS can be recovered very quickly. Therefore, MNs can promptly and reliably execute handover considering congestion of an AP by estimating the AP queue length using probe packets.

4.5.5 Limitation

Our proposed handover management can be applied in a situation that MN locate in overlap area of two APs. The distance between two APs must not execeeding 200 meters since the effective range for 802.11g is 150 meters. Otherwise, MN will experience ping-pong effect and multi-path transmission frequently leading to overload the networks and further loss the connectivity with one or both APs.

4.6. Summary

In a WLAN based on IEEE 802.11 specifications, MNs cannot detect the congestion of an AP. The congestion also contributes the degradation of VoIP quality during movement. In this paper, we thus proposed a handover strategy considering the congestion of an AP. Our proposed method employed probe packets to estimate an AP queue length from RTT between an MN and an AP. However, our proposed method leads to the degradation of VoIP quality due to (1) the ping-pong effect and (2) increase of redundant probe packets. To improve the degradation, we extended the basic method by adding two extension methods. In the first extension, to avoid the simultaneous handovers by many MNs (ping-pong effect), we provide a function to execute handover based on the transmission rate. In the second extension, only one representative MN sends probe packets and the other MNs estimate the AP queue length by capturing the probe packets. To show the effectiveness of the proposed methods, we evaluated them through simulation experiments. From the results, compared with the previous method, we showed the proposed methods can promptly and reliable execute handover considering the congestion of an AP while avoiding the ping-pong effect and reducing redundant probe packets.

Chapter 5

Handover Management for VoIP over Intermingled 802.11g and 802.16e

5.1. Introduction

Future wireless network will be consisting of various wireless network technologies and mobile users will traverse several different wireless access technologies while establishing VoIP communication. Recently, mobile WiMAX (IEEE 802.16e) has gained serious attention as a means of providing wireless broadband access to mobile users in a wide area and it provides QoS for various applications. On the other hand, IEEE 802.11g has already been widely used to provide high data rates in a limited area. Therefore, in the near future, since 802.11g and 802.16e networks are intermingled while complementing each other, an MN will be likely to execute many handovers between 802.11g and 802.16e networks with different IP subnets. In this study, to preserve VoIP quality during handover, we then propose handover management considering wireless link condition and congestion state for VoIP over the intermingled 802.11g and 802.16e networks.

5.2. Related Work

In order to detect the degradation of wireless link quality as well as congestion of wireless network in heterogeneous environment, we cannot only rely on one handover decision criterion such as CINR level of two different wireless technologies. Moreover, in vertical handover, we cannot compare directly the CINR level of different wireless technologies since each technology has different characteristics in terms of physical layer. For instance, they might have same CINR level but differ in data rate. Therefore, it is challenge to develop a vertical handover management considering suitable handover decision criteria from two different networks that give an indication of wireless link and congestion state in order to achieve optimal wireless resources utilization in heterogeneous environment. So far, several studies on the vertical handover scheme between 802.11 and 802.16 have been conducted. Reference [37] proposed a vertical handover algorithm in 802.11 and 802.16 hybrid networks exploiting data rate and channel occupancy as handover decision criteria and implemented on top of MAC layer but no explanation of whether seamless handover is supported. Authors in [38] introduce a mobile relay node as an intermediate node between 802.11 devices and 802.16 BS to convert 802.16 signal received from BS to 802.11 devices. This approach is inefficient because it requires additional intermediate node among 802.11 AP, BS and MN. The most recognized study is Media Independent Handover (MIH) that proposed by IEEE 802.21 working group [39]. MIH is intended to facilitate handover and interoperability of various other network technologies. However, deployment of MIH requires modification/additional layer to the existing protocol stack of mobile device and infrastructure networks leading to cost inefficiency. Therefore, in this study, we focus on an end-to-end handover management that only require some modification on MN side and then our proposed method aim to preserve VoIP quality during handover between the networks with different IP subnets.

5.3. Overview of 802.16e

The mobile WiMAX (802.16e) air interface utilizes OFDMA as radio access method in which data stream from multiple users are orthogonally multiplexed on downlink and uplink subchannels/subcarriers. 802.16e extend the 802.16-2004 standard which addressed Wireless Metropolitan Area Networks for broadband wireless access but support only fixed terminals. The 802.16e is designed to deliver service across many more subchannels than the OFDM 256-FFT as employed by 802.16-2004. The 802.16e standard employs OFDMA with a larger FFT space (2048-FFT, 1024-FFT, 512-FFT and 128-FFT) which is further divided into subchannels in both uplink and downlink. Different subchannels may be allocated to different users as a multiple access method.

Figure 5.1 shows an OFDMA frame when operating in TDD mode. The frame size is variable from 2 ms to 20 ms but most 802.16e equipment will support only 5 ms frames. The frame is divided into two sub-frames: downlink sub-frame and uplink sub-frame. The downlink subframe begins with a downlink preamble that is used for physical-layer procedures such as time and frequency synchronization and initial channel estimation. The downlink preamble is followed by a frame control header (FCH), which provides frame configuration information such as the MAP message length, the modulation and coding scheme and the usable subcarriers. Downlink and uplink MAP (DL-MAP and UL-MAP) are broadcast following the FCH in the downlink subframe. These MAP messages include burst profile that defines the modulation and coding scheme used in that link. 802.16e support adaptive modulation and coding scheme that enable the 802.16e to change the modulation and coding scheme on burst-by-burst basis per link, depending on channel conditions. As fact that the radio channel is better when MN closed to the BS and, in this case, BS can select modulation like 64QAM, which less robust link but offer higher bandwidth. But when MN far away from BS, BS can change the modulation scheme from less robust to more robust modulation e.q. 64QAM to 16QAM or QPSK, which reduces throughput and increases effective range. In order to implement the adaptive modulation and coding scheme, MN must periodically measure their own CINR which indicates current channel condition of MN and report to the BS for changing to the appropriate modulation and coding scheme for MN. This link adaptation scheme significantly increases overall the system capacity. On the other hand, the uplink sub-frame begins with a contention for initial ranging and bandwidth (BW) request. The initial ranging performs closed-loop frequency, time, and power adjustment during network entry as well as periodically afterward. The BW request is used by MN to make uplink BW request before transmission of uplink data. These portions of contention-based access is followed by several uplink bursts from different MNs.

MAC layer of 802.16e is a connection-oriented architecture and designed for handling applications with different QoS requirement. It consists of three sub-layers: Convergence Sublayer (CS), Common Part Sublayer (CPS) and Security Sublayer. CS is designed for classifying and mapping the MSDUs (MAC Service Data Unit) into appropriate CIDs (Connection IDentifier) which is a basic function of QoS management scheme of 802.16. CPS exists in the middle of the MAC layer and it responsible for bandwidth allocation, connection establishment and connection maintenance between the BS and MN. Security Sublayer provides authentication, secure key exchange and encryption.

Furthermore, 802.16e also defines five scheduling service to support a wide variety of applications including UGS (Unsolicited Grant Service), rtPS (real-time Polling Service), nrtPS (non-real-time Polling Service), BE (Best Effort) and ErtPS (Extended RealTime Polling Service). UGS is designed to support fixed-size data packets at a constant bit rate

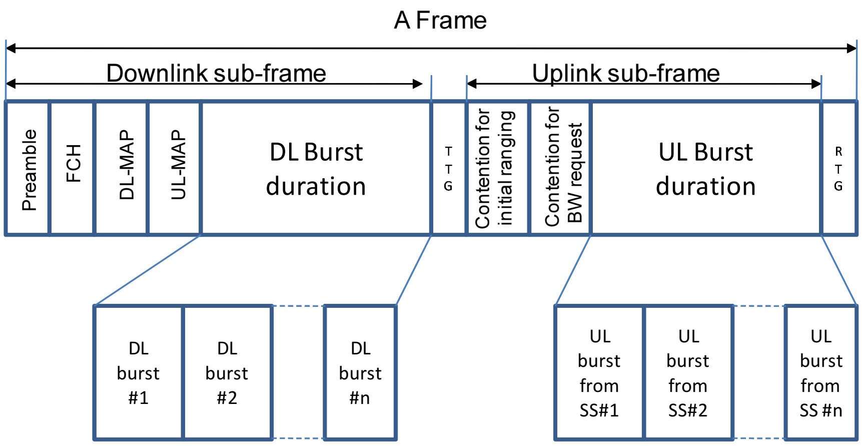

Figure 5.1. 802.16 MAC Frame

(CBR) such as VoIP without silence suppression. rtPS is designed to support real-time application with variable size data such as MPEG video streaming. nrtPS is designed to support delay tolerant data stream such as FTP and BE is designed to support data stream that do not require minimum service of guarantee such as web browsing. ErtPS is designed to support real-time application that variable data rates but required guaranteed data rate and delay. ErtPS is only defined in 802.16e and not in 802.16a.

All scheduling and QoS are controlled by BS. In addition, 802.16e adds power save mode, sleep and idle mode, to cope with power shortage of mobile terminal.

5.4. VoIP over Wireless Networks

In a mobile and wireless environment, typically, two main factors degrade VoIP quality over the wireless network: (1) degradation of wireless link quality and (2) wireless network congestion. First, because an MN freely moves, the communication quality degrades due to the fluctuation of the wireless link condition. Second, as VoIP is a bi-directional communication, an AP of 802.11g and a base station (BS) of 802.16e become a bottleneck with the increase of VoIP calls. The following subsections describe how 802.11g and 802.16e standard treat the VoIP application.

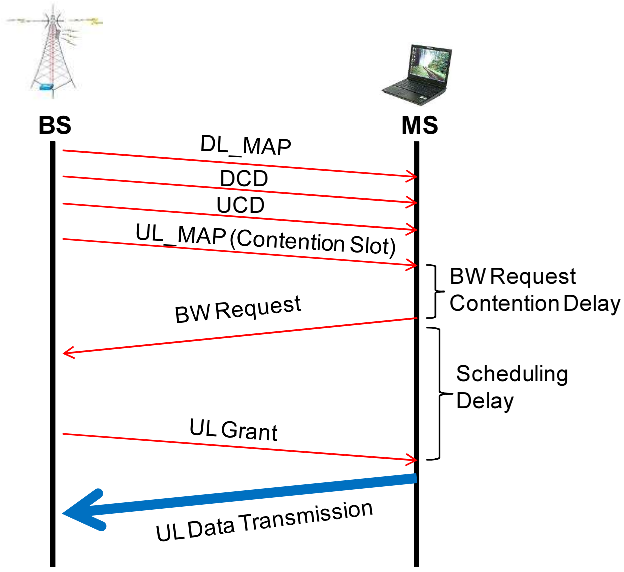

Figure 5.2. Uplink Data Transmission

5.4.1 VoIP over 802.11g

As VoIP is a bi-directional communication, AP becomes a bottleneck with the increase of VoIP calls. That is, VoIP packets to MNs are liable to experience large queuing delay or packet loss due to increase in queue length or buffer overflow in the AP buffer because each MN and AP has almost the same priority level of frame transmission by following the CSMA/CA scheme. In addition, in multi-rate WLANs, although a rate adaptation function changes the transmission rate in response to wireless link condition, a low transmission rate occupies a larger amount of wireless resources than that of a high transmission rate. Thus, compared with a high transmission rate, a low transmission rate tends to cause congestion at an AP. Therefore, to preserve quality of VoIP over 802.11g, we need to develop a handover management considering these two factors.

5.4.2 VoIP over 802.16e

802.16e supports high data rates and multi service types, hence, it is a strong contender for wireless broadband access technology to support real-time application such as VoIP over wireless networks. The 802.16e standard supports multiple service types but often only Best Effort (BE) service type is made available at the initial phase of deployment. The BE scheme has no classification and priority to delay and/or loss sensitive application, hence, the VoIP application must contend with various types of applications for obtaining transmission opportunity in 802.16e. Furthermore, delivering VoIP application over 802.16e networks involves many challenges. An MN is likely to traverse several 802.16e hotzones during a VoIP call, thereby leading to degradation of VoIP quality due to the movement of the MN and handover latency in 802.16e, which consists of latencies for neighbor BS scanning, handover message exchange, and network re-entry. Furthermore, as VoIP is a bi-directional communication, uplink VoIP communication will suffer from the increase of VoIP calls in a single BS due to uplink contention-based access in the 802.16 network. That is, VoIP packets from MN to BS are liable to experience a large queuing delay due to uplink BW request defined in the 802.16e standard and BS scheduling delay as shown in Fig. 5.2. Therefore, to preserve VoIP quality, we need to develop a handover management considering these factors.

5.5. Proposed Handover Decision Criteria

Since we intend to develop a multihomed MN-initiated handover scheme where MN has two different wireless interface, i.e., 802.11g and 802.16e Interfaces, we need to determine the handover decision criteria for each different interfaces that can be obtained at the MN side. Our proposed handover management aims to preserve VoIP quality during handover, hence, we need to consider several handover decision criteria indicating both wireless link condition and congestion state in both 802.11g and 802.16e networks. Since we have described handover decision criteria for 802.11g in chapter 4, we only describe the handover decision criteria for 802.16e in this chapter. The following subsections describe our proposed handover decision criteria for 802.16e and the motivation behind selecting them.

5.5.1 Handover Decision Criteria for 802.16e

We need to determine the handover decision criteria for MN-initiated handover that can be obtained at the MN side of 802.16e. Since we deal with a VoIP application which is bi-directional communication, we need to consider the criteria indicating both wireless link condition and congestion state in 802.16e. In this study, we propose a combined use of two handover decision criteria, i.e., CINR for indicating the wireless link quality and MN queue length for indicating the congestion state of the 802.16e network. We describe the motivation behind selecting these criteria as handover decision criteria for MN-initiated handover of the 802.16e.

Carrier-to-Interference-plus-Noise-Ratio (CINR)

The MN can detect the wireless link channel condition exploiting the received signal quality. Many handover technologies employ Received Signal Strength (RSS) as a handover decision criterion, which can indicate the wireless link condition. RSS provides a simple indication of how strong the signal is at the receiver front end. However, the received signal includes noise, interference, and other channel effects (e.g., multi-path fading and shadowing). Therefore, a high RSS does not always mean that the channel and signal quality are good. Instead, it gives an indication of whether a strong signal is present in the channel of interest. On the other hand, the CINR provides information on how strong the desired signal is compared to the interferer (or noise, or interference-plus-noise).

In the 802.16e, CINR level is used as a reference for changing burst profile transmission. The burst profile is a basic tool in the 802.16 standard MAC layer that is about physical transmission. The burst profile includes modulation and coding scheme (MCS) that determine the transmission rate and it used for the link adaptation procedure. The downlink burst profile is determined by BS according to quality of signal that received by each MN. While operating on a given burst profile, MN measures the instantaneous CINR and a series of instantaneous CINR values are used to derive the mean and standard deviation of the CINR. The mean CINR ($\mu_{CINR}[k]$) and standard deviation ($\sigma_{CINR}[k]$) of the CINR during the k-th measurement report are given by equation [50]:

$$\mu_{CINR}[k] = (1-\alpha)\mu_{CINR}[k-1] + [k] \tag{5.1}$$

$$x^2_{CINR}[k] = (1-\alpha)x^2_{CINR}[k] + \alpha|CINR[k]|^2 \tag{5.2}$$

$$\sigma_{CINR}[k] = \sqrt{x^2_{CINR}[k] - \mu^2_{CINR}[k]} \tag{5.3}$$

When the BS requests a CINR measurement report from the MN, the MN will answer back by including in the REP-RSP MAC message the estimates of the average and standard deviations of the CINR. CINR values are reported in a dB scale for the mean CINR as specified by

$$\mu_{CINR}[k] = 10\log(\mu_{CINR}[k]) \tag{5.4}$$

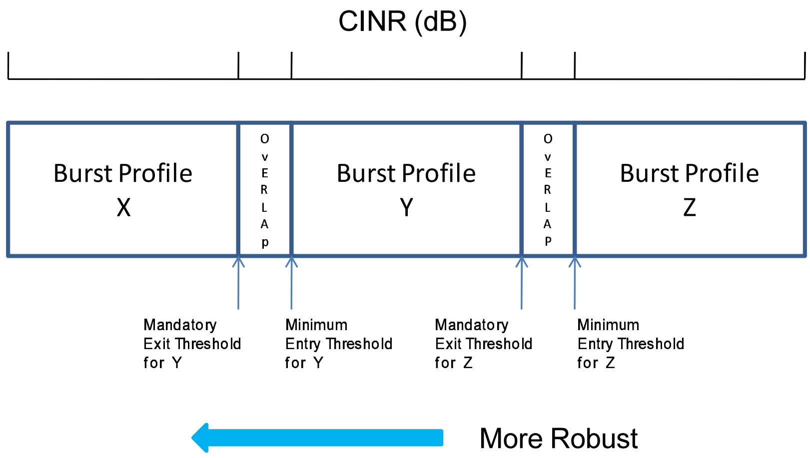

Figure 5.3. CINR as a reference for changing burst profile

The burst profile allocation can changes dynamically and possibly very fast depends on received CINR level. Figure 5.3 shows the possible CINR thresholds value for neighboring burst profile. There are two threshold are used for the selection of one of defined burst profile in Downlink Channel Descriptor (DCD) message that based on the received CINR: burst profile mandatory exit threshold and mandatory entry threshold. If CINR level is less than burst profile mandatory exit threshold, this burst profile can no longer be used and change to a more robust burst profile is required, otherwise, this burst profile can be used and change to a less robust burst profile.

CINR level plays an important role to determine the burst profile transmission in 802.16e. Therefore, in this study, we interest to exploit CINR as a handover decision criterion to initiate the handover in the 802.16e.

MN Queue Length in 802.16e

To recognize the congestion state of the 802.16e network, which means an uplink transmission condition, an MN receives a periodic report of BW allocated to each MN broadcast from a BS at a maximum periodic interval of 10 seconds through DL-MAP and UL-MAP. However, the MN cannot rely on this report to initiate the handover because it takes a long time to get the report. Hence, we need a criterion that can be directly and promptly obtained at the MN side without relying on the report from the BS. The only information indicating the congestion state of the 802.16e network that can be obtained directly at the MN side is an MN queue length. Since the MN queue length is affected by uplink

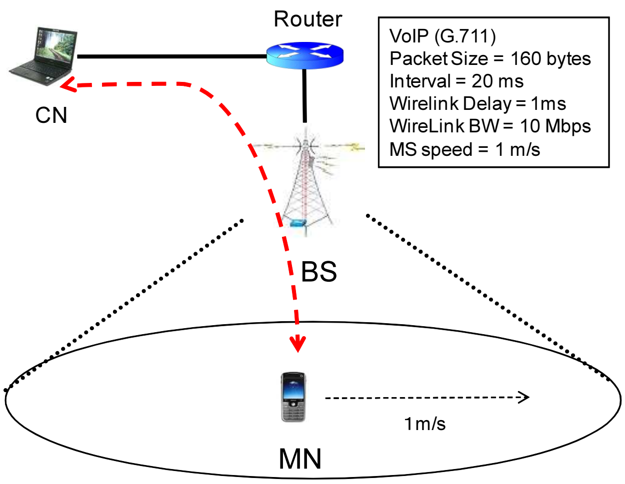

Figure 5.4. Simulation Model 3

latency, we first describe the uplink BW allocation, which leads to the uplink latency in 802.16e.

In the 802.16e standard [34], BW allocation is thoroughly managed by a BS, in which the downlink BW is managed by a BW scheduler at the BS. On the other hand, uplink BW is allocated by the BS through the BW request and uplink grant procedures. As shown in Fig. 5.2, the BS allocates BW to the MNs for the purpose of making BW requests by providing a BW request contention slot in UL-MAP. Then, the MN transmits the BW requests in response to the allocated time slot. After the MN obtains an Uplink Grant, it can then transmit its data with the allocated uplink grants. So, we can sum up the total end-to-end uplink delay (T_{UL-ETE}) as follows:

$$T_{UL-ETE} = T_{BWreq} + T_{Sch} + T_{queue} + T_o \tag{5.5}$$

where T_{BWreq} and T_{Sch} are BW request contention and BS scheduling delay, respectively. T_{queue} represents the queuing delay in the transmission buffer. T_o represent an additional delay, e.g., transmission delays over wireless link and VoIP codec delay. This BW request contention and scheduling delay may become longer with the increase in the number of MNs in a serving BS, which contends the BW request slot in the UL-MAP. The increase of BW request contention and scheduler delay cause the increase of the queuing length in the MN, thereby leading to a degradation of uplink communication quality.

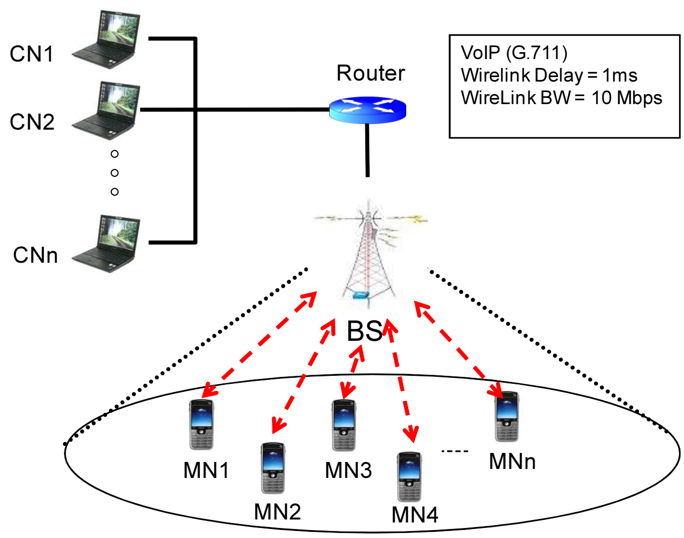

Figure 5.5. Simulation Model 4

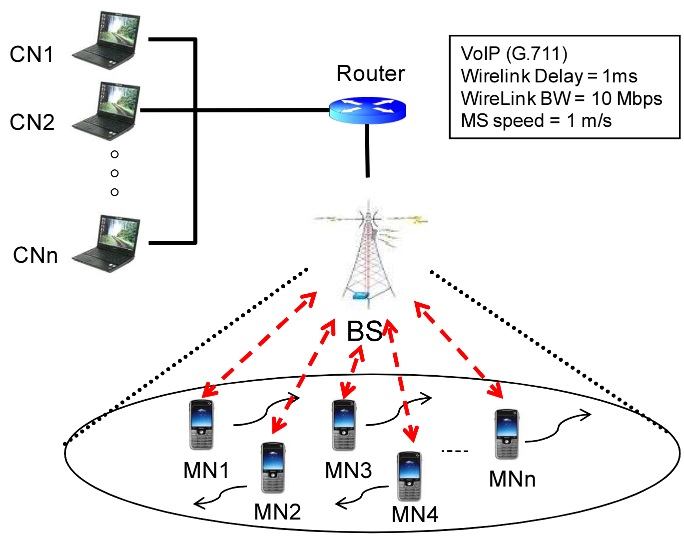

Figure 5.6. Simulation Model 5

Table 5.1. Simulation Parameters 2

Frequency Band	2.4GHz
Duplexing Mode	TDD
Channel Mode	10 MHz
FFT Size (of OFDMA sub-carriers)	1024
Frame Duration	5 ms
Transmit/Receive Time Gap (TTG)	10 us
Receive/Transmit Time Gap (RTG)	10 us
BS Tx Power	30 dBm
MN Tx Power	22 dBm
Fading Model	Rayleigh
VoIP Codec	G.711

5.6. Evaluation of Handover Decision Criteria for 802.16e

We conduct a simulation experiment to investigate the performance of VoIP over 802.16e and the behavior of downlink CINR and MN queue length using Qualnet 4.5 [55]. In a typical wireless environment, multiple propagation paths often exist from a transmitter to a receiver due to scattering by different objects. This leads to the so-called multipath fading. Fading can result in significant fluctuation of RSS, CINR, throughput, and more frequency packet loss and link failure. For realistic simulation of 802.16e, the effect of fading should be taken into account. We have introduced a Rayleigh fading model to meet an urban environment. The Rayleigh fading occurs when there is no line of sight between the BS and MN. The fading speed is affected by how fast the receiver and/or transmitter, or the surrounding objects are moving. We use QualNets Rayleigh fading model that uses pre-computed time series data sequence with different sample intervals to represent the different fading speeds or coherence times of the propagation channel [55]. In this simulation scenario, we set the maximum velocity of nodes and the surrounding objects to 10 meters per second and set the gaussian components including sampling rate of 1000, base doppler frequency of 30 Hz and the number of gaussian components of 16,384. In terms of VoIP application, we employ VoIP codec of G.711, which are commonly used in the VoIP system where G.711 sends a packet of 160 bytes every 20 ms. Through simulation experiments, we aim to determine a certain level of threshold for downlink CINR and MN queue length to meet the requirement of an adequate VoIP call quality.

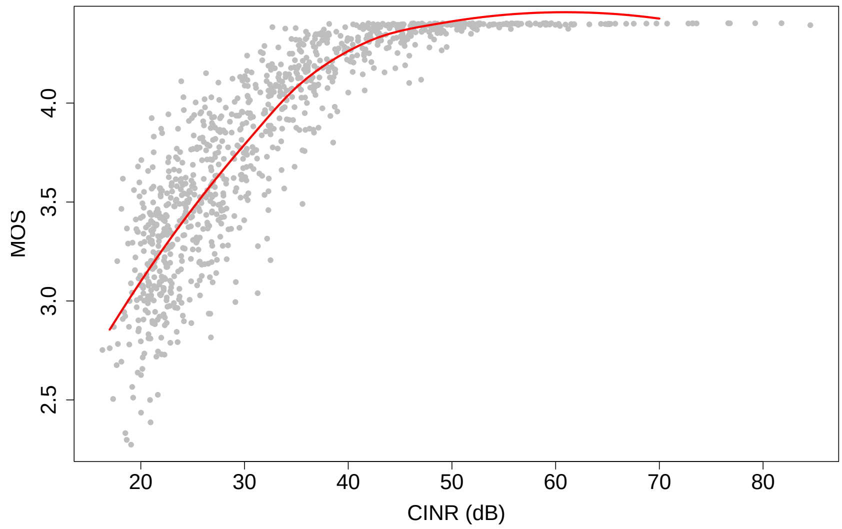

Figure 5.7. MOS vs. CINR

Figures 5.4, 5.5, and 5.6 show the simulation models and Table 5.1 indicates the simulation parameter. Although 802.16 standards have defined multiple service types in order to guarantee different levels of QoS, at the initial phase of deployment, often only the BE service is made available due to inexpensive implementation cost, therefore, In this study we limited our focus on employing BE service scheduling.

5.6.1 Evaluation of CINR

First of all, we conduct a simulation experiment to investigate the behavior of downlink CINR in 802.16e. Figure 5.4 and Table 5.1 show a simulation model and simulation parameters, respectively. In the simulation scenario (Fig. 5.4 and Table 5.1), an MN establishing VoIP call moves away from BS at the speed of 1 m/s. Figure 5.7 and 5.8 show the relationship between <Downlink MOS and CINR> as well as <packet loss ratio and CINR>, when an MN moves away from BS. Since the relationship starts to scatter when CINR level is below 50 dB due to the fluctuation of wireless link condition, we employ a local regression smoothing method (loess method) to grasp their trend and estimate the best fit of <CINR level and MOS> as well as <CINR level and packet loss > over the distance shown as a red curve. The red curve shows that CINR level above 27 dB satisfies an adequate VoIP call. Therefore, we can set the CINR level of 27 dB as a threshold for initiating handover in a mobile environment.

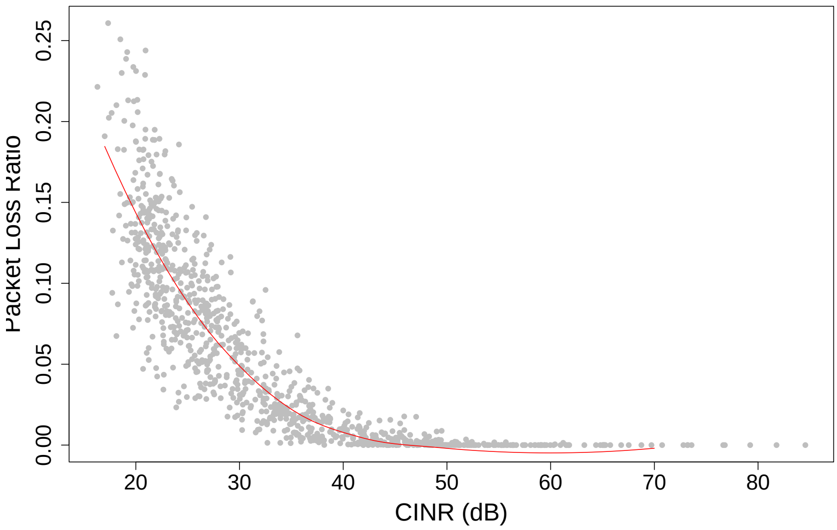

Figure 5.8. Packet Loss vs. CINR

5.6.2 Evaluation of MN Queue Length

We conduct a simulation experiment to investigate the relationship between the number of MNs establishing VoIP calls and MN queue length. In this simulation scenario, we randomly locate from one to 30 MNs within a coverage area of the single BS. Each MN communicates with a CN using VoIP. We have two scenarios for stationary MNs and moving MNs. Figures 5.5 and 5.6 show the simulation model for stationary MNs and moving MNs, respectively. For the moving scenario, every MN moves randomly at a speed of 1 m/s within a coverage area of the single BS.

Figures 5.9 and 5.12 show the average uplink and downlink VoIP quality as well as the average of MN queue length over the number of stationary and moving MNs in the single BS. These figures show that uplink VoIP quality decreases as the number of MNs in the single BS increases for both stationary and moving MNs scenario. On the other hand, downlink VoIP quality is kept at an adequate VoIP quality even if the number of VoIP calls increases. Moreover, the MN queue length increases as the number of MNs establishing VoIP calls in the single BS increases. From these figures, we can see that the bottleneck of the 802.16e network affects only flows from the MN to BS, i.e., uplink quality. This bottleneck causes the uplink latency, which is mainly raised from the uplink BW request contention and scheduling delay. This uplink latency causes an increase of the MN data queuing length, thereby leading to the degradation of uplink communication

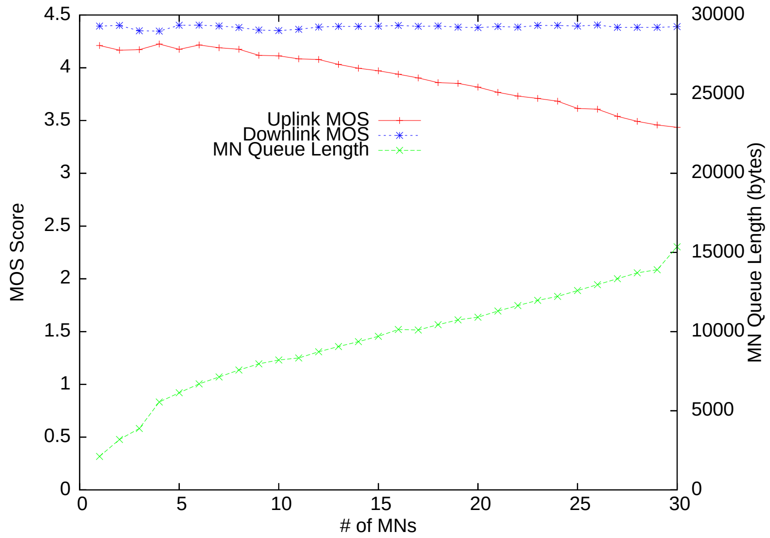

Figure 5.9. Relationship among Uplink MOS, MN Queue Length

quality.

In terms of accommodation of MNs in the single BS, we can see that a single BS can simultaneously accommodate 24 MNs VoIP calls. However, in a moving scenario, a BS can simultaneously only accommodate 21 MNs VoIP calls.

In this simulation experiment, we aim to determine the threshold level of MN queue length to meet the requirement of an adequate VoIP call quality condition in the 802.16e network. Therefore, we show the relationship among MN queue length, uplink VoIP quality and uplink latency in Figs. 5.15 and 5.16 for stationary and moving MNs scenarios. From these figures, we can see that VoIP quality decreased as increase of MN queue length. The MN queue length is increased due to increase of uplink latency. From these figures, we can see that the G.711 VoIP application has a tolerable uplink end-to-end delay up to 320 ms and 260 ms for the stationary and the moving scenarios, respectively.

Since we are interested in MN queue length, we can see from Fig. 5.15 that MN queue lengths of less than *13,000* meet the requirement for adequate VoIP quality for G.711 in the stationary scenario. In the moving MN scenario, from Fig. 5.16 MN queue lengths of less than *11,600* meet the requirement for adequate VoIP quality for G.711 VoIP calls.

VoIP quality in the moving scenario is degraded a bit faster than in the stationary scenario because of the movement of the MN. In terms of latency, there is no significant difference between the stationary and the moving MN scenario as shown in Fig. 5.10 and 5.6.2 for the G.711 VoIP codec. On the other hand, in terms of packet loss, there

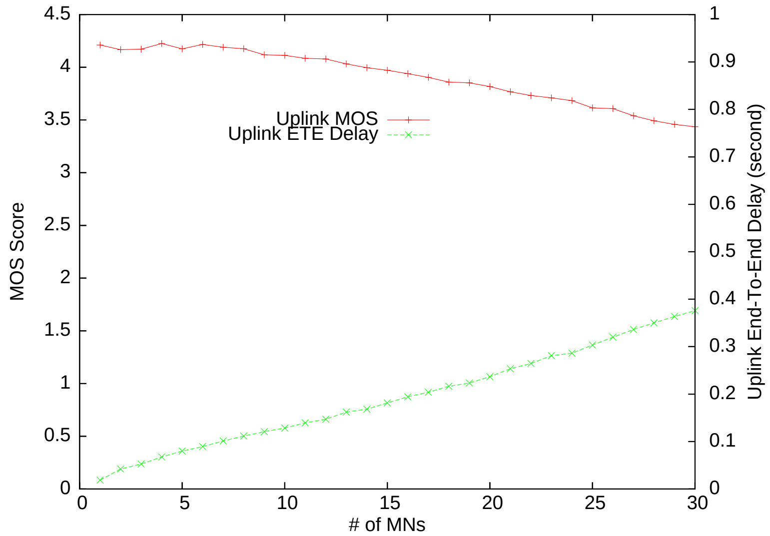

Figure 5.10. Relationship among Uplink MOS, Delay and ♯ of Stationary MN

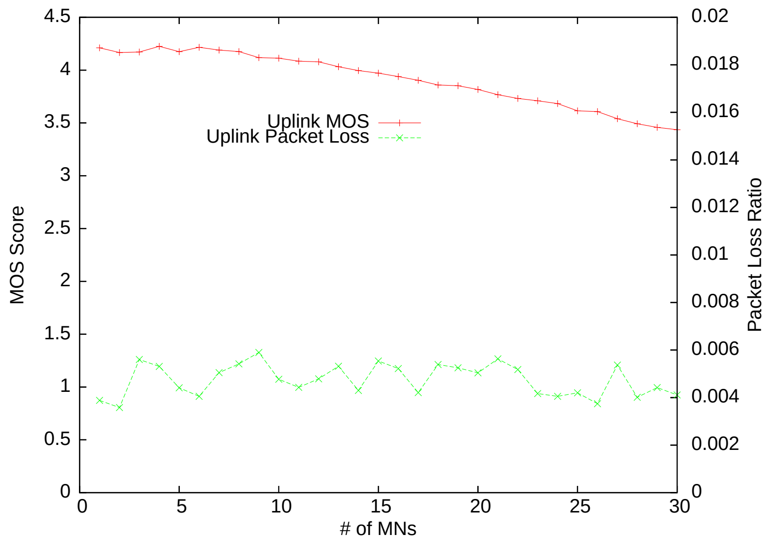

Figure 5.11. Relationship among Uplink MOS, Packet Loss and ♯ of Stationary MN

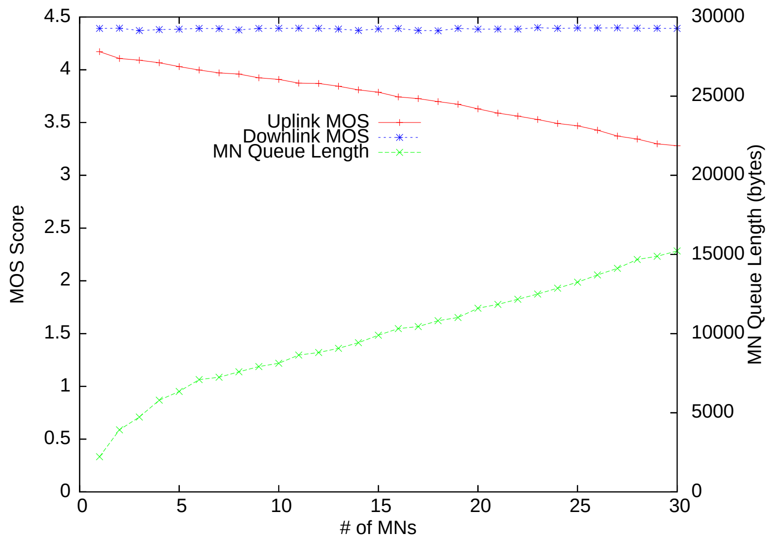

Figure 5.12. Relationship among Uplink MOS, MN Queue Length and ♯ of Moving MNs

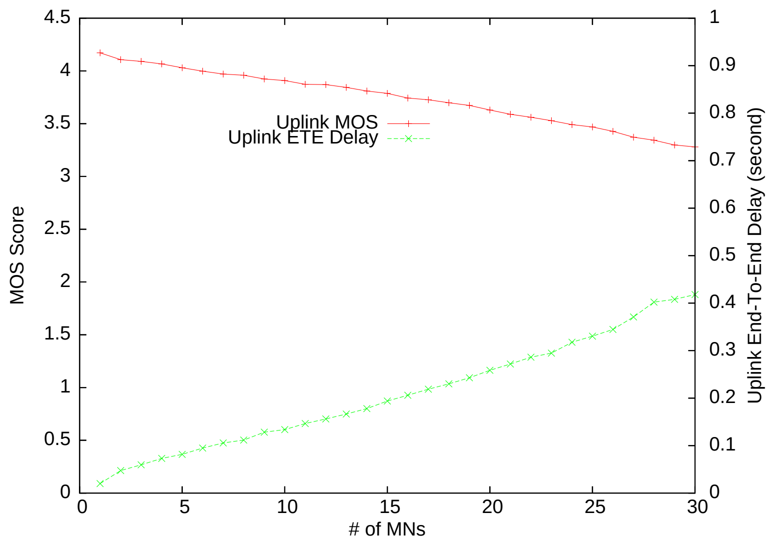

Figure 5.13. Relationship among Uplink MOS, Delay and ♯ of Moving MNs

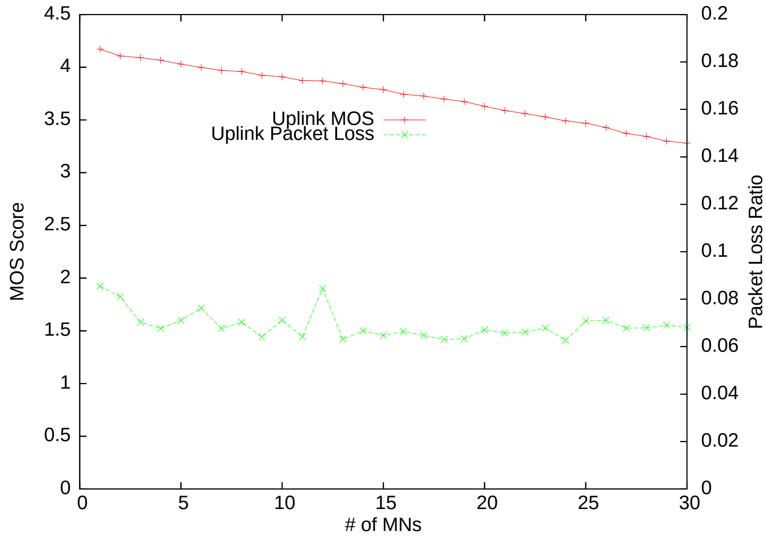

Figure 5.14. Relationship among Uplink MOS, Packet Loss and ♯ of Moving MNs

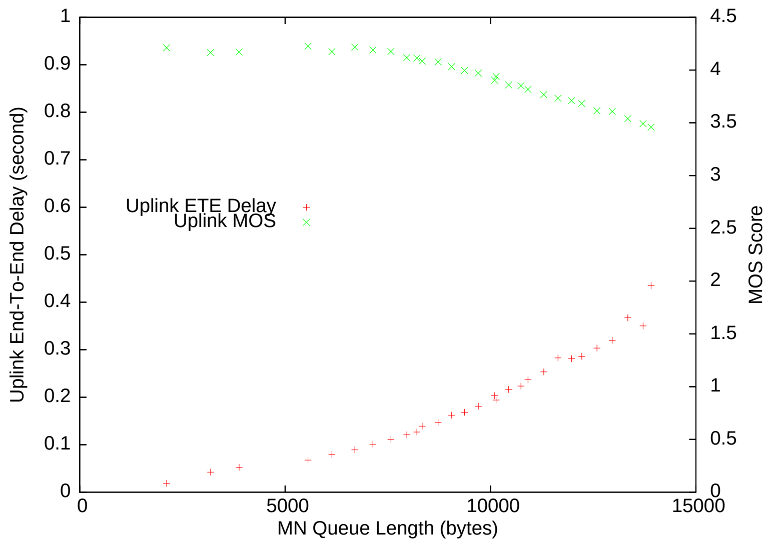

Figure 5.15. Ave. MN-QL, Uplink Delay, MOS of Stationary MNs

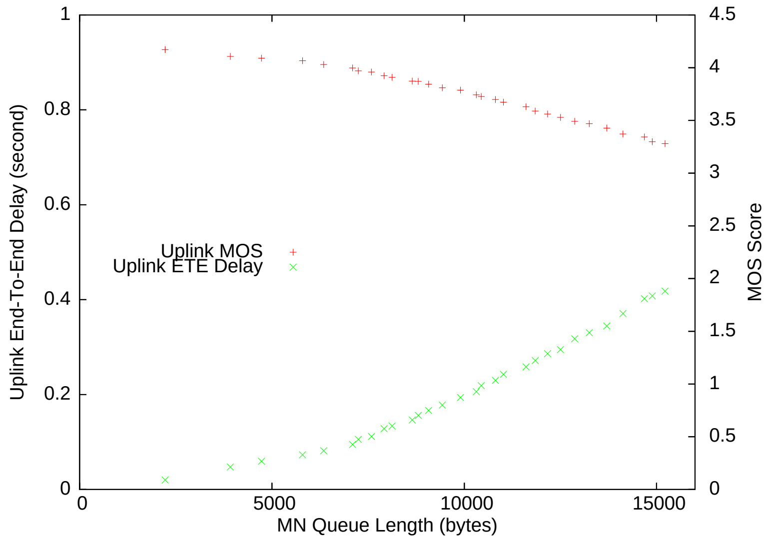

Figure 5.16. Ave. MN-QL, Uplink Delay, MOS of Moving MNs

is a significant increase of packet loss in the moving MN scenario with compare to the stationary scenario, as shown in Fig. 5.11 and 5.14. Therefore, MN queue length threshold level for the moving scenario is a bit shorter than the stationary scenario in order to anticipate the degradation due to the MN movement.

From the simulation results, we can see that the increase of the number of MNs causes the increase in uplink latency, thereby leading to an increase of MN queue length. Therefore, MN queue length can be a significant metric for detecting the congestion state of a wireless network and can be a reference for initiating the handover.

5.7. Proposed Handover Management

We propose an end-to-end multi-homed MN-initiated handover management for intermingled 802.11g and 802.16e networks. Our proposed handover management is implemented on transport layer of MN and obtains effective handover decision criteria from low layer using cross layer approach. The handover management exploit RTS retries and W-RTT of 802.11g interface as well as CINR level and MN queue length of 802.16e interface as handover decision criteria. The handover management switches between single-path and multi-path transmission modes in response to wireless network condition of both 802.11g and 802.16e. The handover management appropriately selects the path according to handover decision criteria.

5.7.1 Single-Path and Multi-Path Transmission

Our proposed handover method employs multi-homed MN in order to support soft-handover.The handover management properly switches between single-path and multi-path transmission modes in response to wireless network condition. Single-path transmission mode means that an MN communicates with a CN using only one interface. Multi-path transmission, on the other hand, means that an MN sends duplicated packets to a CN through two interfaces for supporting soft-handover.

Figure 5.17 shows an algorithm of switching to single/multi-path transmission when an MN located in an overlap area of 802.11g and 802.16e (AP and BS). When initial transmission path of MN is via 802.11g interface, the handover management monitors RTS retry ratio to detect the wireless link quality. If RTS retry ratio exceeds threshold RTS_thr, MN switches transmission path to multi-path mode, otherwise, stay in single-path mode via 802.11g interface. When initial transmission path of MN is via 802.16e interface, the handover management monitors downlink CINR level to detect the wireless link quality. If CINR level exceeds threshlod $CINR\_thr$, MN switches transmission path to multi-path mode, otherwise, stay in single-path mode via 802.16e interface.

When multi-path transmission is applied, the handover management monitors all handover decision criteria from both interfaces and compares them. As shown in Fig. 5.18, an algorithm of switching to single-path transmission works as follows. First of all, as shown in Fig. 5.18, an MN monitors and evaluates link state handover decision criteria, i.e., RTS retry ratio and CINR, of 802.11g and 802.16e, respectively. If both criteria still exceed the threshold, then MN stays in multi-path mode. If RTS retry ratio exceeds the threshold while CINR level are below threshold, then, MN switches to the single-path via 802.16e, otherwise, single-path via 802.11g. When both criteria are below threshold, MN evaluates the congestion state handover on criteria of both 802.11g and 802.16e, i.e., W-RTT and MN queue length. If both congestion state criteria still exceed the threshold, MN stays in multi-path mode. If W-RTT of 802.11g interface exceeds the threshold while MN queue length of 802.16e interface are below threshold, then, MN switches to the single-path via 802.16e, otherwise, single-path via 802.11g. When both criteria below threshold, MN switches back to initial transmission path as same path as when the MN establish a connection at first time.

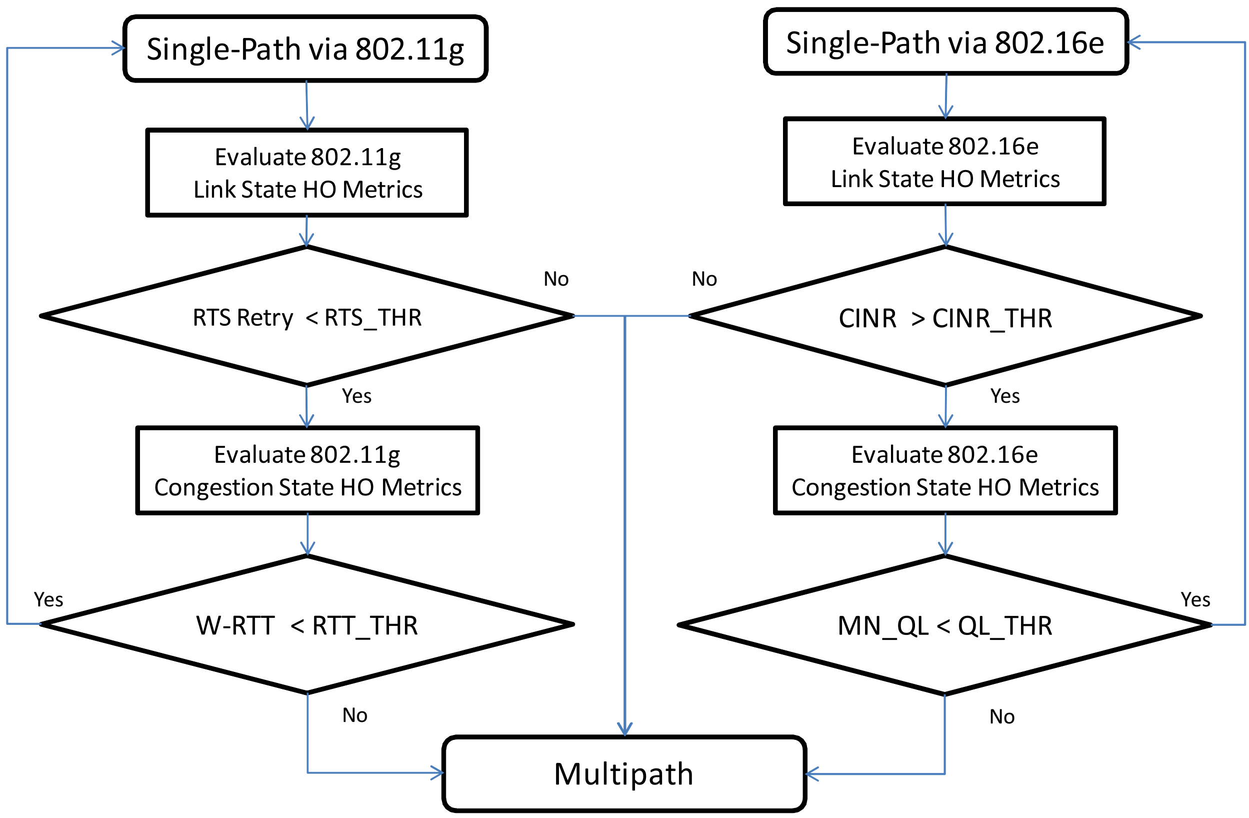

Figure 5.17. Algorithm of Switching from Singlepath to Singlepath/Multipath

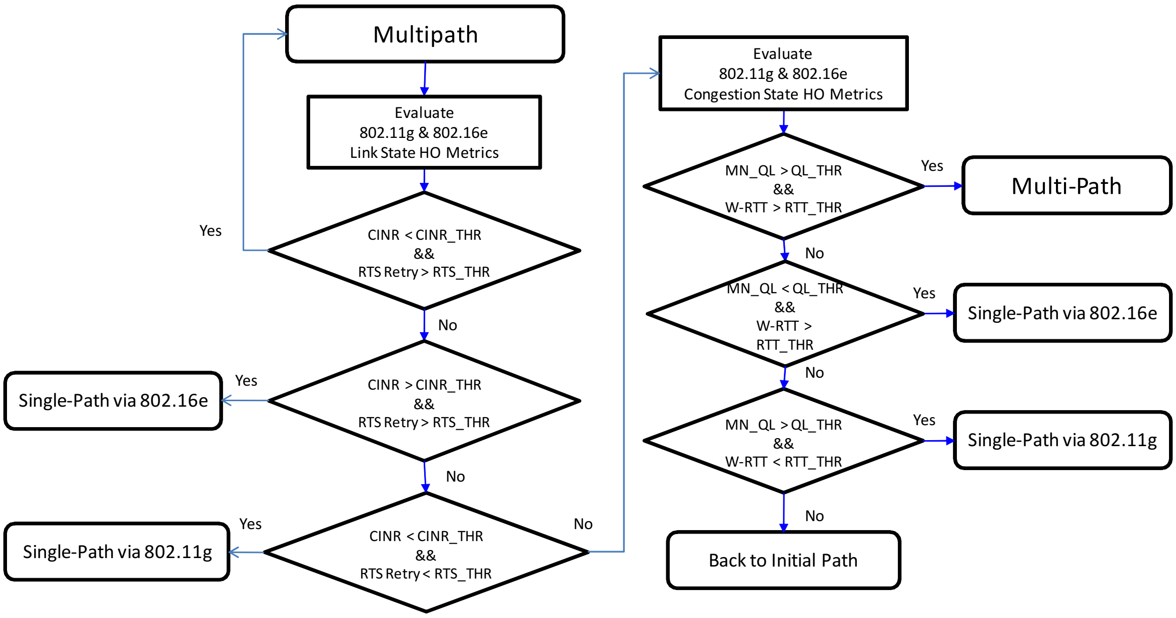

Figure 5.18. Algorithm of Switching from Multipath to Singlepath/Multipath

5.7.2 Assumption and Considered Environments

Our proposed handover management can be applied in a condition that 802.16e network cell overlays a 802.11g network cell, hence, an MN locate in 802.11g network cell can associate to 802.16e network. We assume that both interfaces are in standby mode. Therefore, MN does not need to perform network acquisition including network scanning and ranging for 802.11g and 802.16e, respectively. Therefore, we neglect a delay caused by network acquisition.

5.8. Simulation Experiments and Results

We conduct simulation experiments to investigate the effectiveness of our proposed vertical handover management using Qualnet 4.5. We set an urban environment by introducing a Rayleigh fading model to meet our scenario. The Rayleigh fading occurs when there is no line of sight between the BS and MN. The fading speed is affected by how fast the receiver and/or transmitter, or the surrounding objects are moving. In this simulation experiment, we employ VoIP codec of G.711, which are commonly used in the VoIP system where G.711 sends a packet of 160 bytes every 20 ms.

Table 5.2. Threshold Value

CINR_THR	26 dB
RTS_THR	0.6
QL_THR	12,000 bytes
RTT_THR	200 ms

5.8.1 Performance Evaluation of handover based on Link State Handover Decision Criteria

Figure.5.19 and 5.21 show the simulation model to evaluate the proposed handover algorithms. An MN, which is establishing a VoIP call with CN, moves from 802.11g area to 802.16e area at speed of 1 m/s, and vice versa.

From Fig.5.20 show the charachteristics of two handover decision criteria, i.e., RTS retry (802.11g) and CINR level(802.16e), when MN move from 802.11g to 802.16e. From this figure, RTS retry is fluctuated and exceeds the threshold at 171 seconds. MN totally switches to 802.16e networks at 207 seconds when RTS retry exceeds the threshold and CINR level of 802.16e is within acceptable level ($> 26dB$). On the other hand, Figure.5.22 show that charachteristics of two criteria when MN move from 802.16e to 802.11g. CINR level is fluctuated and start to exceeds the threshold at simulation time of 236 seconds. MN totally switch to 802.11g at 267 seconds when CINR level and RTS retry are less than the threshold. From these figures, we can see that the MN initiates and executes the handover before the handover decision criteria is degraded, leading to degradation of VoIP quality at a certain level.

In terms of VoIP quality, we can see that our proposed method can obtain average MOS values of 4.29 (Fig. 5.29 and 4.28 (Fig. 5.30) for uplink and downlink, respectively, when an MN moves from 802.11g to 802.16e. On the other hand, When MN move from 802.16e to 802.11g, Average MOS value are 4.057 (Fig. 5.32) and 4.286 (Fig. 5.33) for CN and MN, respectively. This shows that our proposed handover management can preserve VoIP quality during the MN's movement from 802.11g to 802.16e and vice versa.

5.8.2 Performance Evaluation of Handover based on Congestion State Handover Decision Criteria

We conducted simulation experiments to evaluate the effectiveness of our proposed handover in congested 802.11g network. We consider the scenario depicted in Fig. 5.27 to

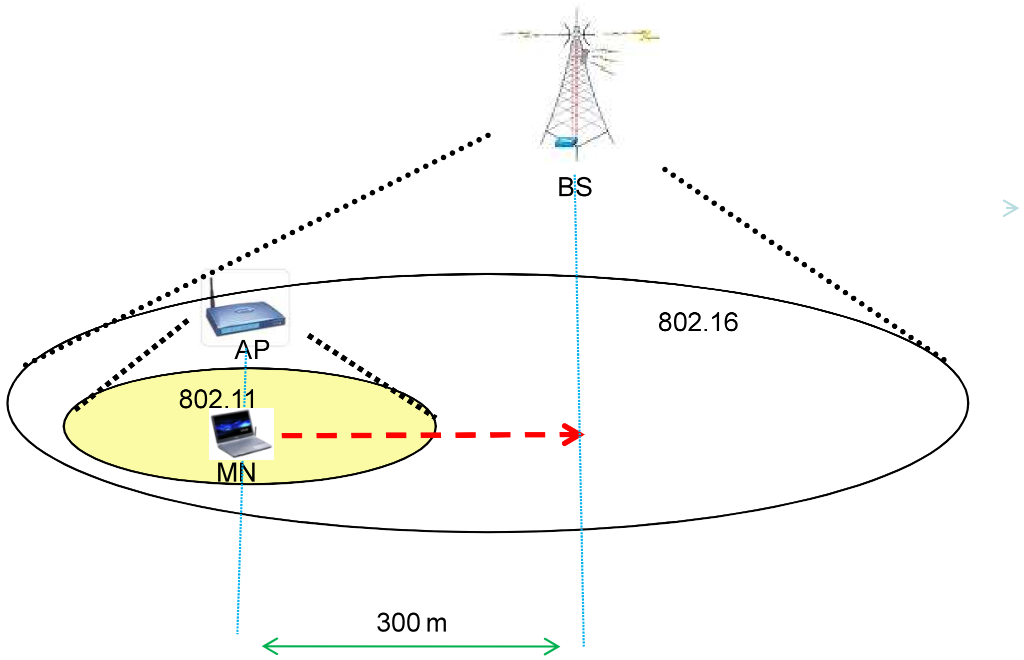

Figure 5.19. Simulation Model 6 (Move from 802.11g to 802.16e)

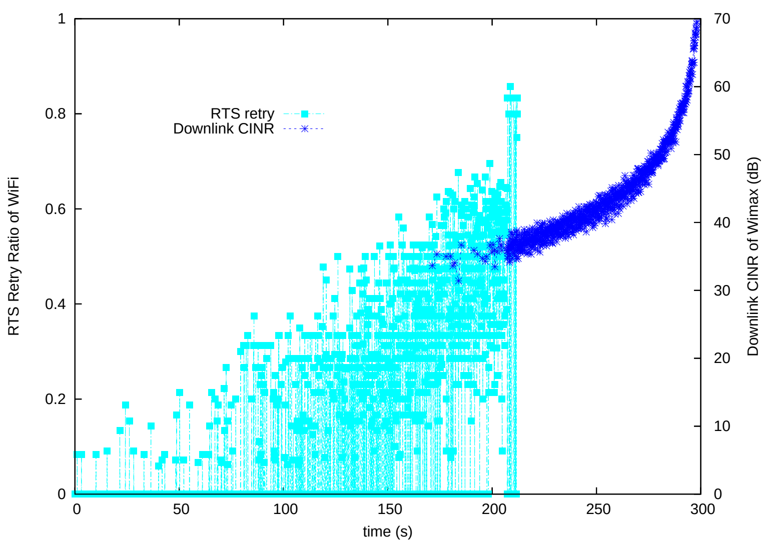

Figure 5.20. Link State Criteria Charachteristics (802.11g to 802.16e)

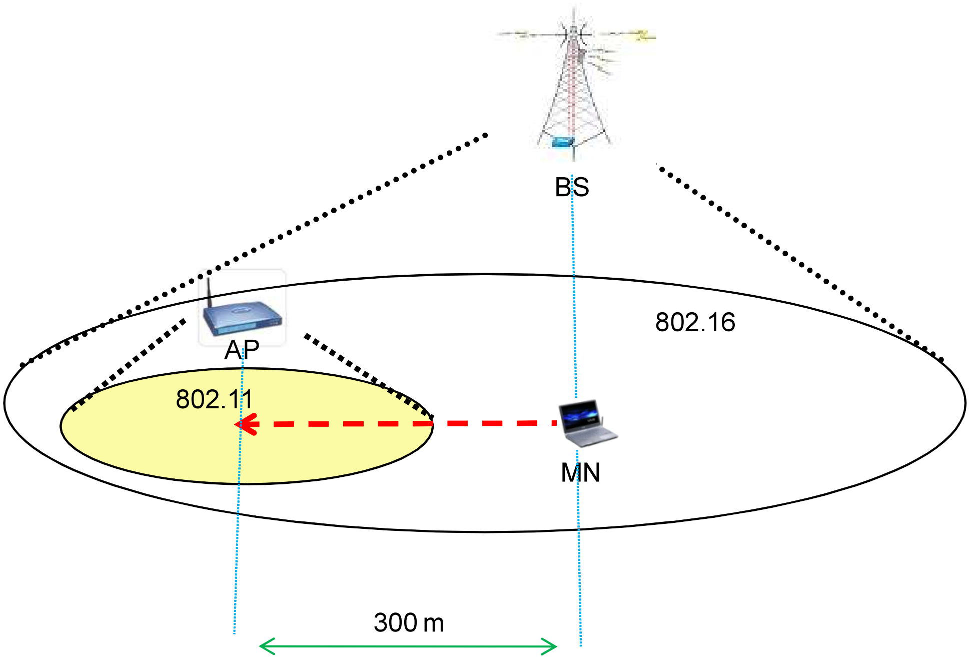

Figure 5.21. Simulation Model 7 (Move from 802.16e to 802.11g)

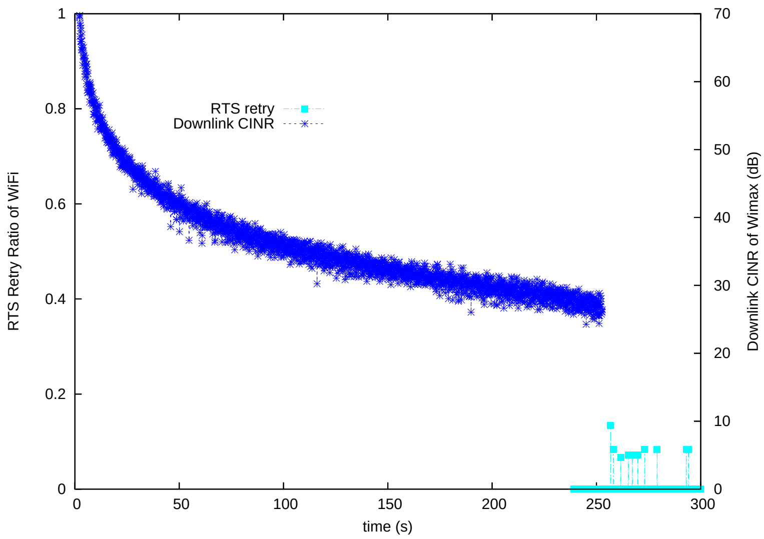

Figure 5.22. Link State Criteria Charachteristic (802.16e to 802.11g)

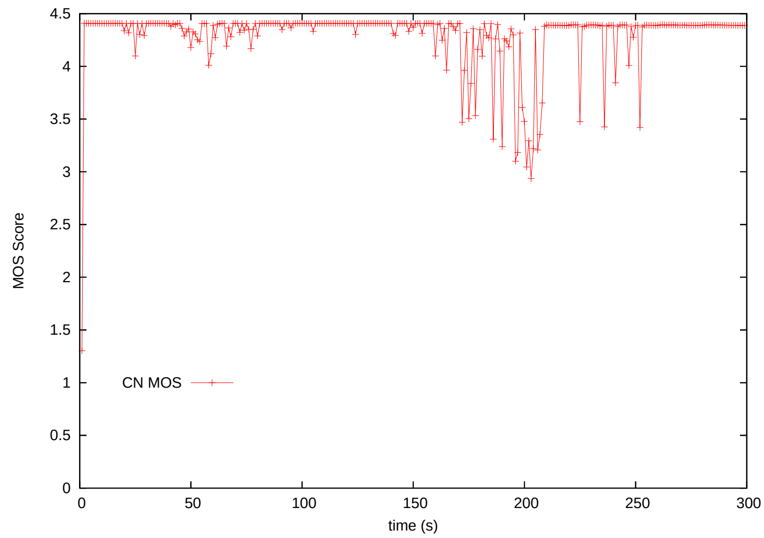

Figure 5.23. MOS of CN (802.11g to 802.16e)

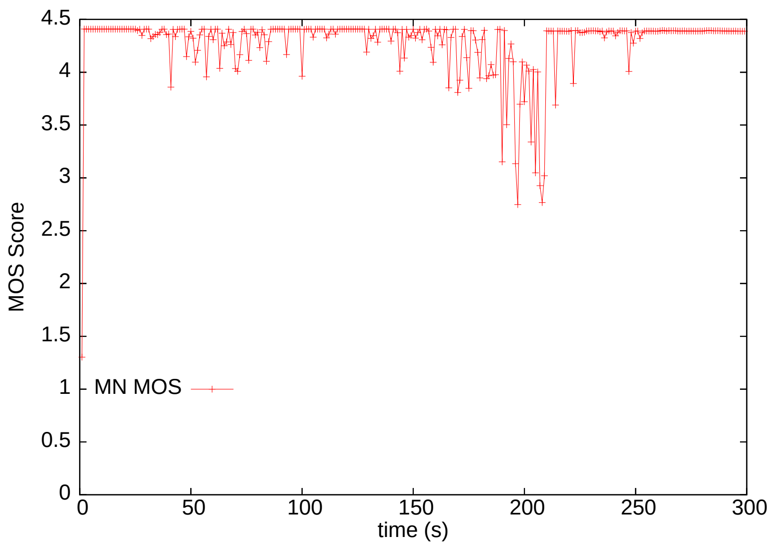

Figure 5.24. MOS of MN (802.11g to 802.16e)

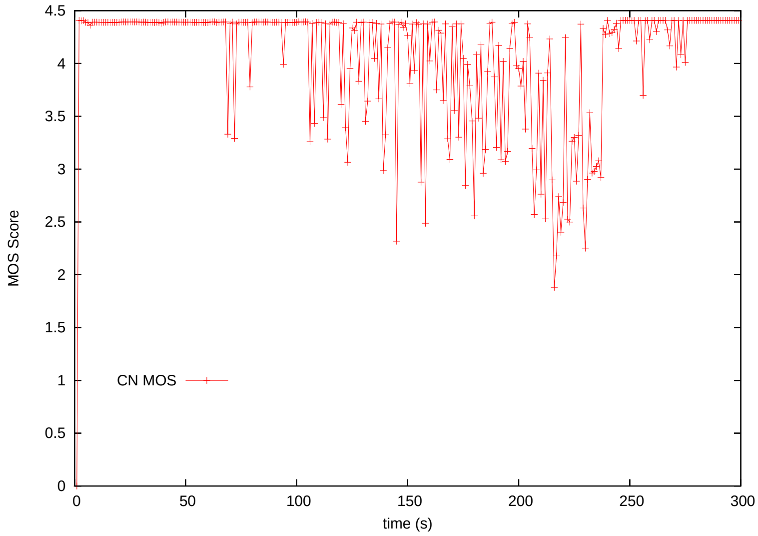

Figure 5.25. MOS of CN (802.16e to 802.11g)

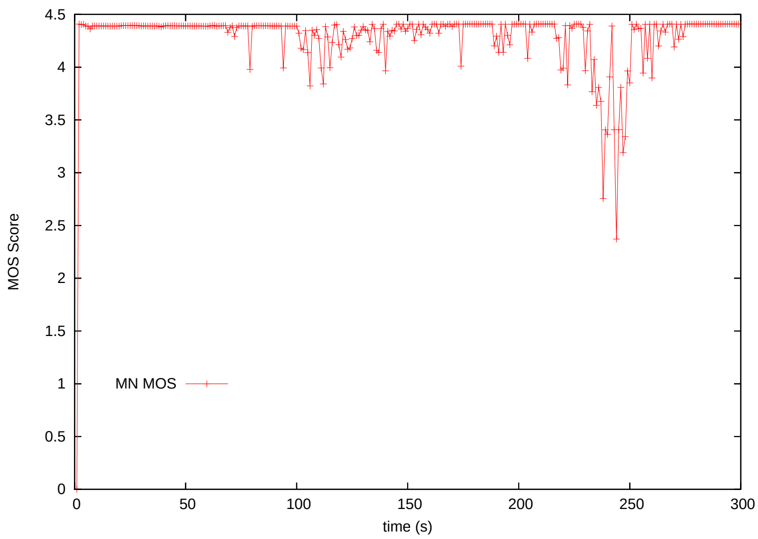

Figure 5.26. MOS of MN (802.16e to 802.11g)

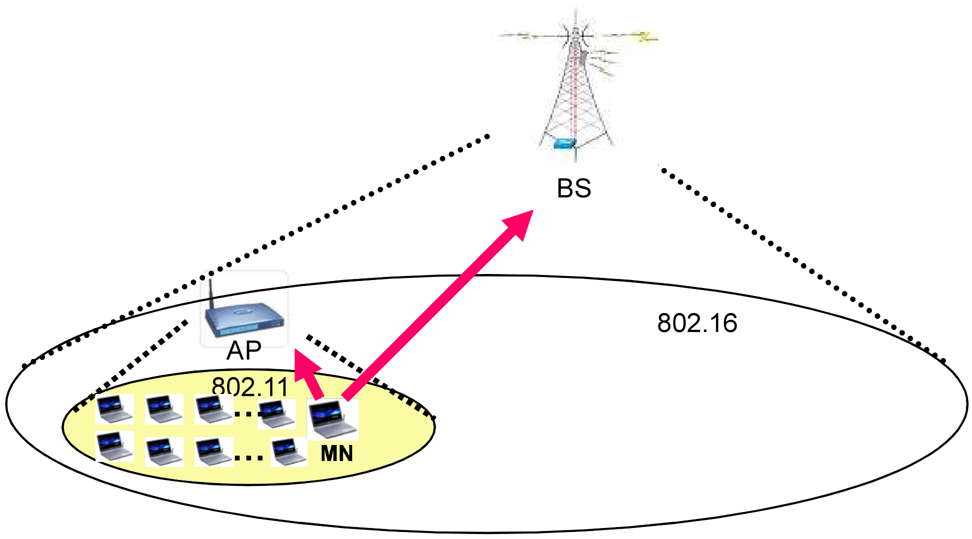

Figure 5.27. Simulation Model 8

validate our proposed handover scheme based on congestion state handover decision criteria, i.e., MN Queue Length (802.16e) and W-RTT (802.11g). At beginning of simulation, there are 13 MNs are randomly distributed in the vicinity of an AP (i.e., between 0 and 50m) where the access to 802.16e networks is granted. There is one of among MNs employing the proposed handover algorithms and it initially uses the 802.11g AP to established VoIP call. Every 5 second, a new MN establishes VoIP connection with their CN via 802.11g network, so that the traffic in the 802.11g network is gradually increases.

Simulation results show that average MOS of uplink and downlink are 4.258 and 4.249, respectively (Fig. 5.29 and 5.30). From Fig. 5.28 show that W-RTT exceeds the threshold of 200ms at simulation time of 47 seconds and then MN handover to 802.16e Network soon.

Moreover, we also evaluate when 802.16e network is congested. In this case, at beginning of simulation, there are 30 MNs are randomly distributed in the vicinity of AP and BS. There is one of among MNs employing the proposed handover algorithms and it initially uses the 802.16e to establish VoIP call. Every 3 second, a new MN establishes VoIP connection with their CN via 802.16e network, so that the traffic in the 802.16e network is gradually increases.

Simulation results show that average MOS of uplink and downlink are 3.882 and 4.336, respectively (Fig. 5.32 and 5.33). From Fig. 5.31 show that MN queue length exceeds the

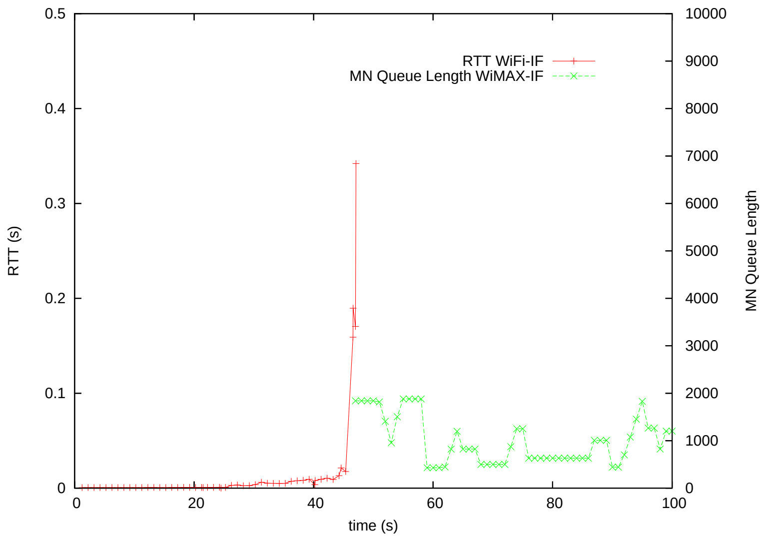

Figure 5.28. Congestion State Criteria Charachteristics (802.11g to 802.16e)

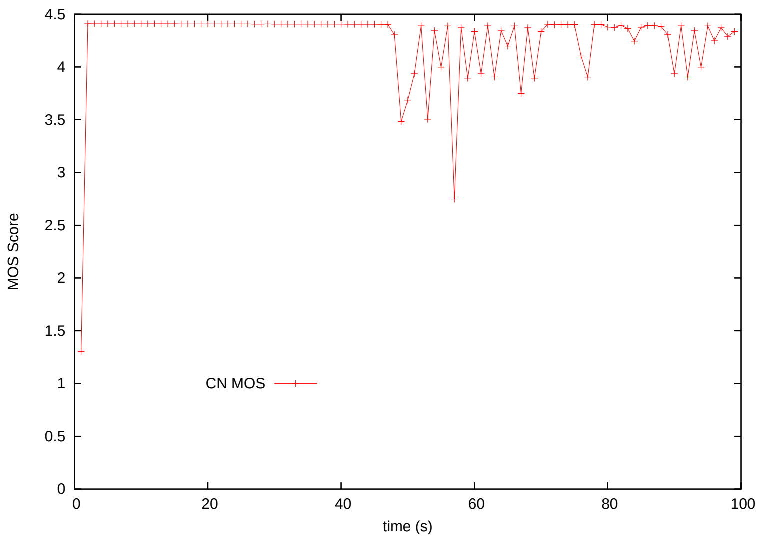

Figure 5.29. MOS of CN (802.11g to 802.16e)

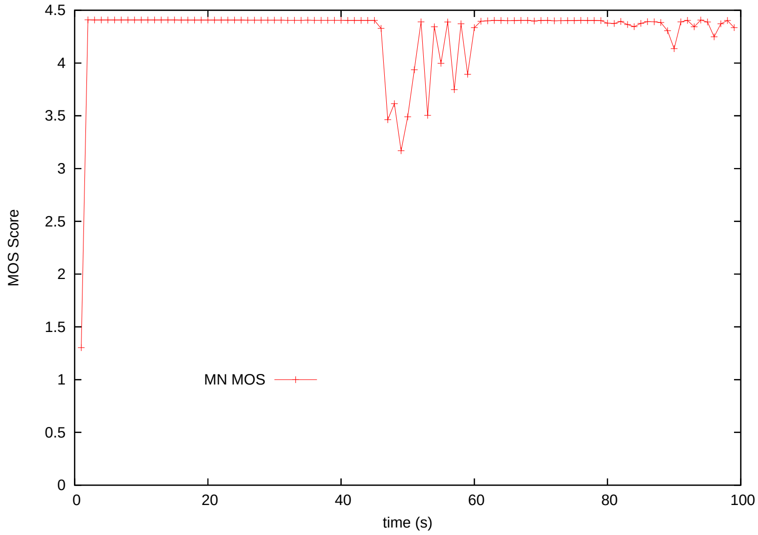

Figure 5.30. MOS of MN (802.11g to 802.16e)

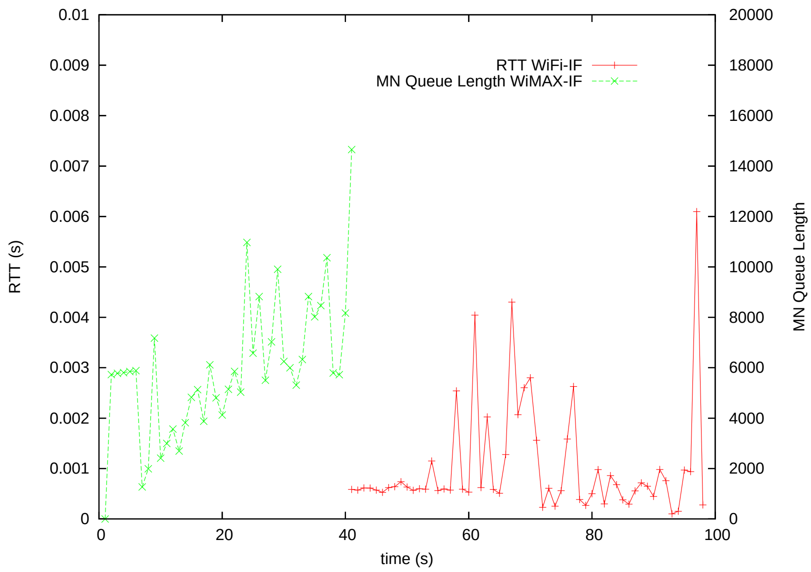

Figure 5.31. Congestion State Criteria Charachteristics (802.16e to 802.11g)

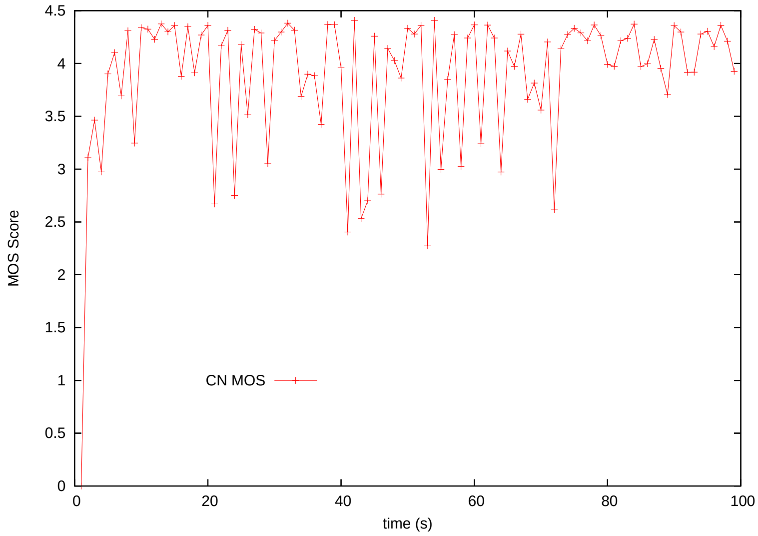

Figure 5.32. MOS of CN (802.16e to 802.11g)

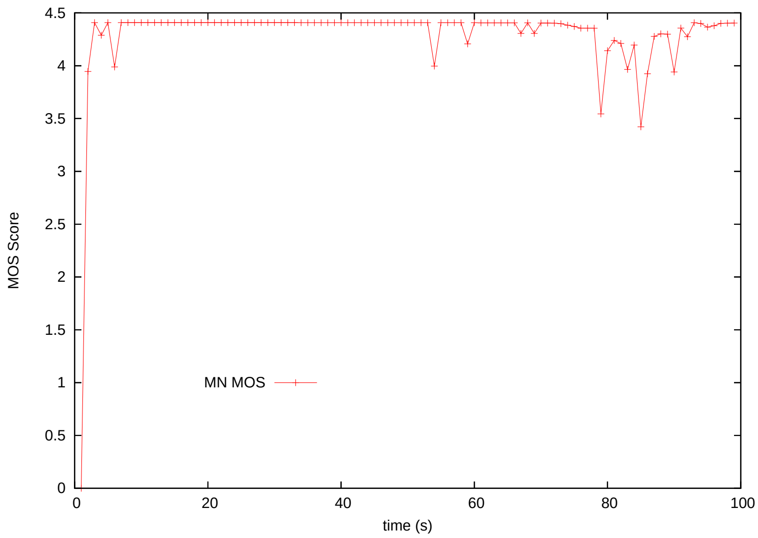

Figure 5.33. MOS of MN (802.16e to 802.11g)

threshold of 12,000 bytes at simulation time of 42 seconds and then the MN handover to 802.11g Network soon after. This shows that our proposed handover management can preserve VoIP quality when one of the wireless network is congested.

5.9. Summary

In this study, we have proposed an end-to-end seamless vertical handover management for VoIP over intermingled 802.11g and 802.16e networks. Our proposed handover management is implemented on transport layer of MN and obtains effective handover decision criteria from lower layer using cross layer approach. The handover management exploits RTS retries and W-RTT of 802.11g interface as well as CINR level and MN queue length of 802.16e interface as handover decision criteria. Our proposed method aim to preserve VoIP quality during handover between the networks with different IP subnets. We conducted simulation experiments to investigate the effectiveness of our proposed vertical handover management in moving and congested environment. Our simulation results show that our proposed handover management can preserve VoIP quality during MN's handovers from 802.11g to 802.16e and vise versa.

Chapter 6

Conclusion and Future Direction

6.1. Conclusion

The main theme of this dissertation is how to provide seamless handover for multi-homed MN that moves between 802.11g networks as well as intermingled 802.11g and 802.16e networks while preserving real-time session quality on an end-to-end basis. The 802.11g and 802.16e will become a key technology as means of an economically viable solution for providing wireless broadband access to mobile user. These two different wireless access technologies will co-exist while complementing each other in the near future, hence, a MN with dual interfaces will be likely to execute many handovers between 802.11g networks as well as between 802.11g and 802.16e networks with different IP subnets.

Meanwhile, there is a huge demand for VoIP service over wireless network. However, delivering VoIP over wireless network has many challenges because VoIP is a delay and packet loss sensitive application. An acceptable VoIP call must have a total E2E not exceeding 150-200 ms [2]. However, wireless access networks exhibit significant variations in packet delay. Enforcing a limit on the E2E delay is not easy when passing from one access network to another. VoIP applications must synchronize the speech playout. With a limited amount of buffering due to E2E delay limitations, a change in the network delay might cause a loss of synchronization, thus deteriorating the quality of a VoIP call also in the absence of packet loss. This is the case when packets arrive too late to be played back, or when they arrive too soon and the application cannot perform packet reordering. When moving to a network that offers a lower bit rate, a VoIP application may experience congestion situations with packet loss and service degradation. An acceptable VoIP call must have a packet loss ratio not exceeding 5 percent. In a mobile and wireless environment, typically, two main factors degrade VoIP quality over the wireless network:

(1) degradation of wireless link quality and (2) wireless network congestion. First, because an MN freely moves, the communication quality degrades due to the fluctuation of the wireless link condition. Second, as VoIP is a bi-directional communication, an access point (AP) of 802.11g and a BS of 802.16e become a bottleneck with the increase of VoIP calls.

In this dissertation, we have presented our proposed an end-to-end seamless handover management for VoIP over 802.11g networks as well as intermingled 802.11g and 802.16e networks considering wireless link condition and congestion state of both wireless network. We presented our proposed handover management for VoIP over 802.11g networks and intermingled 802.11g and 802.16e networks in chapter 4 and 5, respectively.

In chapter 4, we proposed handover management for VoIP over 802.11g networks. Our proposed method employed probe packets to estimate an AP queue length from RTT between an MN and an AP. However, our proposed method leads to the degradation of VoIP quality due to (1) the ping-pong effect and (2) increase of redundant probe packets. To improve the degradation, we extended the basic method by adding two extension methods. In the first extension, to avoid the simultaneous handovers by many MNs (ping-pong effect), we provide a function to execute handover based on the transmission rate. In the second extension, only one representative MN sends probe packets and the other MNs estimate the AP queue length by capturing the probe packets. To show the effectiveness of the proposed methods, we evaluated them through simulation experiments. From the results, compared with the previous method, we showed the proposed methods can promptly and reliably execute handover considering the congestion of an AP while avoiding the ping-pong effect and reducing redundant probe packets.

In chapter 5, we proposed handover management for VoIP over intermingled 802.11g and 802.16e networks. Our proposed handover management is implemented on transport layer of MN and obtains effective handover decision criteria from lower layer using cross layer approach. The handover management exploits RTS retries and RTT of 802.11g interface as well as CINR level and MN queue length of 802.16e interface as handover decision criteria. Our proposed method aim to preserve VoIP quality during handover between the networks with different IP subnets. We conducted simulation experiments to investigate the effectiveness of our proposed vertical handover management in moving and congested environment. Our simulation results show that our proposed handover management can preserve VoIP quality during such handovers.

In this dissertation, we show that end-to-end handover management using multihoming feature is an effective solution for achieving seamless handover for real-time application such as VoIP over 802.11g networks as well as intermingled 802.11g and 802.16e networks. Furthermore, since our proposed handover management is implemented on transport layer, the change of the physical layer is possible. Therefore, our proposed handover management can be also applied to any other IP-based broadband wireless access technology but they may differ in handover decision criteria due to nature different of physical charachteristics of each technology.

6.2. Future Direction

In this dissertation, we proposed end-to-end handover management for VoIP over multi-homed broadband wireless access network. Our proposed handover management focuses on handover initiation and decision to select the best interface as transmission path. In terms of implementation in the real network, we need to address several issues that must be considered. The following are several issues that we must consider in the future:

- In this study, we assigned a static IP address on each interface of MN, hence, in the case of IP address is changed or dynamically assigned by DHCP, MN does not have a mechanism to inform the IP changes to its peer or CN. Therefore, we need a location management such as HA in mobile IP that stores information about changes of MN's IP address or dynamic IP address reconfiguration in SCTP that can add/delete to the interface of an existing association and send a request to CN to set its primary destination address.

- When dynamic change of IP address is implemented, we must make sure that on-going VoIP session will not disturbed during the change of IP address. Therefore, we may adjust our proposed handover management in order to deal with this problem.

- When two networks, serving AP/BS and targeted BS/AP, are congested, our proposed handover management will experience extensive multipath transmission. This leads to overload the network and contribute to extensive congestion in two networks. Therefore, in the future, we need to extend our proposed handover management with congestion control mechanism in order to deal with the condition when two networks are congested.

Acknowledgments

All praise and thanks are due to Almighty Allah, Who in His Infinite Mercy and Grace enabled me to complete the present dissertation. I bow my head with all submission and humility by way of gratitude due to Almighty Allah.

First of all, I would like to express my gratitude to my supervisor, Professor Suguru Yamaguchi, who gave me a great opportunity to conduct doctoral course and research in NAIST. He gave me advice, support and encouragement to conduct this research as well as giving me extraordinary experiences through out the work. Without his kind support, I would not have been able to complete this dissertation.

I would like to express my appreciation to associate professor Youki Kadobayashi for his kind support and advice. He pointed out many important issues for this dissertation. His intellectual qualities and valuable advice help me to improve my dissertation.

I gratefully acknowledge assistance professor Shigeru Kashihara for setting me on this topic-path. He gave me guidance and support from the very early stage of this research and giving me extraordinary experiences through out the work. This dissertation could not have been accomplished without his assistance. I am grateful to him in every possible way and hope to keep up our collaboration in the future.

I would like to thank assistant professor Takeshi Okuda for providing me some nacessary tools/equipments that help me to proceed my research work. He also gave me valuable comments and advice in this research.

I would like to thank assistant professor Kazuya Tsukamoto at Kyushu Institute of Technology for his valuable advice and reviewing all my conference and journal papers as well as internet draft. Without his assistance, I could not publish papers as part of requirement for accomplishing Ph.D course. I am grateful to him in every possible way and hope to keep up our collaboration in the future.

I would like to thank assistant professor Teruaki Yokoyama at Cyber University for help me settle in IPlab and as a working partner in AI3 project.

I would like to thank assistant professor Hiroaki Hazeyama for useful advice and discussing with me regarding this research.

I would like to thank Yuzo Taenaka and Eigo Horiuchi for research collaboration. They gave a significant contribution in completion of this dissertation. I hope that we can keep up our collaboration in the future.

I would like to thank Kazuya Okada and Gregory Blanc for setting up the room for doctoral defense presentation and preparing the video documentation.

I would like to thank Mrs.Natsue Tanida and Ms.Emiko Okamoto for supporting me in school administration stuffs.

I would like to thank all IPlab members who were important to the successful realization of dissertation, as well as expressing my apology that I could not mention personally one by one.

Finally, I wish to express my gratitude to my family. The compilation of this dissertation manifests the desire of my father who has been the perennial source right from my childhood by giving me advice and inspiration, side by side. I wish to express my endearing affections to my mother who passed away during my Ph.D study in NAIST. She was my dependable source of comfort and supports me whenever I call. I pray her soul may rest in peace. I wish to express my gratitude to my wife, my son and my daughters support me in various way to enable me accomplished my Ph.D Degree.

Bibliography

[1] M. S. Kuran, T. Tugcu,“A survey on emerging broadband wireless access technology,” Computer Networks, vol.51, No.11, pp. 3013-3046, August 2007.

[2] M.Narbutt, A. Kelly, L. Murphy, P. Perry,“Adaptive VoIP playout scheduling: Assessing user satisfaction,” IEEE Internet Computing, pp. 28-34, July 2005.

[3] D. A. Maltz, P. Bhagwat, “MSOCKS: an architecture for transport layer mobility,” IEEE INFOCOM, San Francisco, CA, pp. 1037 . 1045, March 29 - April 2, 1998.

[4] D. Maltz, P. Bhagwat, “TCP splicing for application layer proxy performance,” IBM Research Report RC 21139, IBM T.J. Watson Research Center, March 1998.

[5] A. C. Snoeren, H. Balakrishnan, “An end-to-end approach to host mobility,” MOBICOM, Boston, MA, pp. 155-166, August 6-11, 2000.

[6] H. Matsuoka, T. Yoshimura, and T. Ohya, “End-to-end robust IP soft handover,” IEEE ICC, Anchorage, Alaska, pp. 532-536, May 11-15, 2003.

[7] A. Bakre, B. R. Badrinath, “I-TCP: indirect TCP for mobile hosts,” IEEE International Conference on Distributed Computing Systems, Vancouver, Canada, pp. 136-143, May 30 - June 2, 1995.

[8] Z. J. Haas, P. Agrawal, “Mobile-TCP: an asymmetric transport protocol design for mobile systems,” IEEE ICC, Montreal, Canada, pp. 1054-1058, June 8 - 12, 1997.

[9] K. Brown, S. Singh, “M-UDP: UDP for mobile cellular networks,” Computer Communication Review, vol. 26, no. 5, pp. 60-78, October 1996.

[10] D. Funato, K. Yasuda, and H. Tokuda, “TCP-R: TCP mobility support for continuous operation,” IEEE International Conference on Network Protocols, Atlanta, GA, pp. 229-236, October 28 - 31, 1997

[11] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg,"SIP: session initiation protocol," Request for Comments 2543, Internet Engineering Task Force, March 1999.

[12] H. Schulzrinne, E. Wedlund, " Application-layer mobility using sip," ACM Mobile Computing and Communications Review, pp. 47-57, 2000.

[13] C. Perkins (Ed.), "IP Mobility Support for IPv4," IETF RFC3344, August 2002.

[14] D. Johnson, C. Perkins, and J. Arkko, "Mobility Support in IPv6, " IETF RFC3775, June 2004.

[15] Stemm, M. and R. Katz, "Vertical Handoffs in Wireless Overlay Networks", Journal Mobile Networks and Applications, vol. 3, number 4, pages 335-350, 1998.

[16] Montavont, N., "Problem Statement for multihomed Mobile Nodes", draft-montavont-mobileip-multihoming-pb-statement-00.txt Internet draft, IETF, October 2003.

[17] E. K. Paik, et al., "Multihomed Mobile Networks Problem Statements", draft-paik-nemo-multihoming-problem-00.txt, Internet draft, IETF, October 2003.

[18] A. Mishra, et al., "An Empirical Analysis of the IEEE802.11 MAC Layer Handoff Process," ACM SIGCOMM Computer Communication Review, Vol.33, Issue 2, April 2003.

[19] R. Koodli, "Fast Handovers for Mobile IPv6, " IETF RFC4068, July 2005.

[20] H. Soliman, et al., "Hierarchical Mobile IPv6 Mobility Management (HMIPv6)," IETF RFC4140, August 2005.

[21] S. J. Koh, et al., "Mobile SCTP for Transport Layer Mobility," draft-reigel-sjkoh-sctp-mobility-05.txt, Internet draft, IETF, July 2005.

[22] John Fitzpatrick, et al., "An Approach to Transport Layer Handover of VoIP over WLAN," Proc. of IEEE CCNC, January 2006.

[23] Ernst, T., "Goals and Benefits of Multihoming,", draft-multihoming-generic-goals-and-benefits-00.txt, Internet draft, IETF, February 2004.

[24] ITU-T:"G.107", http://www.itu.int/rec/T-REC-G.107/en.

[25] K. Tsukamoto, et al., “Experimental Evaluation of Decision Criteria for WLAN handover: Signal Strength and Frame Retransmission,” IEICE Trans. on Communications, Vol. E90-B, No. 12, pp. 3579-3590, December 2007.

[26] H. Velayos and G. Karlsson, “Techniques to reduce the IEEE802.11b handoff time,” Proc. of IEEE ICC, vol. 7, pp. 3844-3848, June 2004.

[27] S. Kashihara and Y. Oie, “Handover Management based on the number of data frame retransmissions for VoWLAN,” Elsevier Computer Communications, vol. 30, no. 17, pp. 3257-3269, November 2007.

[28] S. Kashihara, et al.,“Service-oriented mobility management architecture for seamless handover in ubiquitous networks,” IEEE Wireless Communications, Vol. 14, No. 2, pp. 28-34, April 2007.

[29] Y. Taenaka, et al.,“Design and Implementation of Cross-layer Architecture for Seamless VoIP Handover,” Proc. of IEEE MHWMN, October 2007.

[30] Y.Fukuda and Y.Oie,“Decentralized Access Point Selection Architecture for Wireless LANs -Deployability and Robustness-,” Proc. of IEEE VTC2004-Fall, September 2004.

[31] Muhammad Niswar, Shigeru Kashihara, Kazuya Tsukamoto, Youki Kadobayashi, and Suguru Yamaguchi, “Handover Management for VoWLAN based on Estimation of AP Queue Length and Frame Retries”, IEICE Transactions on Information and System, Vol. E92-D, No. 10, pp.1847-1856, October 2009.

[32] M. Niswar, et al., “Seamless VoWLAN Handoff Management based on Estimation of AP Queue Length and Frame Retries,” in Proc. of IEEE PerCom, March 2009.

[33] Muhammad Niswar, et al., “ MS-initiated Handover Decision Criteria for VoIP over IEEE 802.16e,” In Proceedings of IEEE Pacific Rim Conference on Communications, Computers and Signal Processing(PACRIM’09), August 2009.

[34] IEEE Std. 802.16e-2005, IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, Amendment for Physical and Medium Access Layers for Combined Fixed and Mobile Operation in Licensed Bands, February 2006.

[35] IEEE Std 802.16-2004, IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems, October 2004.

[36] WiMAX Forum, “Mobile WiMAX-PartI: A Technical Overview and Performance Evaluation,” August 2006.

[37] Z.Dai et al., “Vertical handover criteria and algorithm in IEEE 802.11 and 802.16 hybrid networks,” Proc.of ICC 2008.

[38] Rashid A. S, et al, “WiFi/WiMAX Heterogeneous Seamless Handover,” Proc.of BroadCom, pp.169-174, 2008.

[39] V.Gupta et al., “IEEE 802.21 Tutorial,” July 2006.

[40] Dongmei Zhao, Xuemin Shen, ”Performance of Packet Voice Transmission using IEEE802.16 Protocol”, IEEE Wireless Communication Magazine, February 2007.

[41] Ching Yao Huang et all, ”Radio Resource Management of Heterogeneous Services in Mobile WiMAX Systems”, IEEE Wireless Communication Magazine, February 2007.

[42] N. Scalabrino, F. De Pellegrini, R. Riggio, A. Maestrini, C. Costa and I. Chlamtac, ”Measuring the Quality of VoIP Traffic on a WiMAX Testbed under VoIP traffic”, IEEE Tridentcom 2007, May 2007.

[43] S. Choi, et al., “Fast Handover Scheme for Real-Time Downlink Services in IEEE 802.16e BWA System,” in Proc. of IEEE VTC-Spring, vol.3, pp.2028-2032, May 2005.

[44] P. Li, et al., “A Seamless Handover Mechanism for IEEE802.16e Systems,” in Proc. of ICCT, November 2006.

[45] KA. Kim et al., “A Seamless Handover Mechanism for IEEE802.16e Broadband Wireless Access,” in LNCS 3515, pp.527-534, May 2005.

[46] S. Cho, et al., “Hard Handoff Scheme Exploiting Uplink and Downlink Signals in IEEE802.16e Systems,” in Proc. of IEEE VTC-Spring, pp.1236-1240, May 2006.

[47] DH. Lee, et al., “Fast Handover Algorithm for IEEE 802.16e Broadband Wireless Access System,” in Proc. of 1st International Symposium on Wireless Pervasive Computing, pp. 1- 6, January 2006.

[48] Gast, Matthew S., “802.11 Wireless Networks,” O’Reilly, April 2005.

[49] Nuaymi, L. “WiMAX: Technology for Broadband Wireless Access,” Wiley, July 2007.

[50] Andrews, Jeffrey G., et al., “Fundamentals of WiMAX: Understanding Broadband Wireless Networking,” Prentice Hall, June 2007.

[51] Chen, Kwang-Cheng, Marca, J.Roberto B.de., “Mobile WiMAX,” Wiley, 2008.

[52] ITU-T Recommendation G.107, “The E-Model: A Computational Model for Use in Transmission Planning,” March 2005.

[53] Cole, “ Voice over IP Performance Monitoring,” ACM SIGCOMM Computer Communication Review, Volume 31 , Issue 2, April 2001.

[54] Athina P.Markopoulo, Foud A.Tobagi, Mansour J.Karam, “Assessment of VoIP Quality over Internet Backbones,” In Proc. of IEEE INFOCOM 2002.

[55] Scalable Network Technologies, http://www.scalable-networks.com/

Publications

Journals

1 Muhammad Niswar, Shigeru Kashihara, Kazuya Tsukamoto, Youki Kadobayashi, and Suguru Yamaguchi, "Handover Management for VoWLAN based on Estimation of AP Queue Length and Frame Retries", *IEICE Transactions on Information and System*, Vol. E92-D, No. 10, pp.1847-1856, October 2009.

International Conferences

1 Muhammad Niswar, Eigo Horiuchi, Shigeru Kashihara, Tsukamoto Kazuya, Youki Kadobayashi, and Suguru Yamaguchi, "Seamless VoWLAN Handoff Management Based on Estimation of AP Queue Length and Frame Retries", *IEEE PerCom Workshop on Pervasive Wireless Networking (PWN 2009)*, Galveston Texas USA, March 2009.

2 Muhammad Niswar, Shigeru Kashihara, Yuzo Taenaka, Kazuya Tsukamoto, Youki Kadobayashi, and Suguru Yamaguchi, " MS-initiated Handover Decision Criteria for VoIP over IEEE 802.16e", *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing(PACRIM'09)*, Victoria Canada, August 2009.

3 Muhammad Niswar, Shigeru Kashihara, Yuzo Taenaka, Kazuya Tsukamoto, Youki Kadobayashi, and Suguru Yamaguchi,"Seamless Vertical Handover Management for VoIP over Intermingled 802.11g and 802.16e Networks", *8th Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT'10)*, Kuching Sarawak Malaysia, June 2010 (To Appear).

Technical Report

1 Muhammad Niswar, Shigeru Kashihara, Yuzo Taenaka, Kazuya Tsukamoto, Youki Kadobayashi, and Suguru Yamaguchi, " MS-initiated Handover Decision Criteria for VoIP over IEEE 802.16e", *Mobility Multimedia Communication (MoMuC2009-22)*, pp.65-70, July 2009.

2 Muhammad Niswar, Shigeru Kashihara, Yuzo Taenaka, Kazuya Tsukamoto, Youki Kadobayashi, and Suguru Yamaguchi, " Vertical Handover Management for VoIP over Multi-homed Heterogeneous Wireless Networks", *Mobility Multimedia Communication (MoMuC2010)*, March 2010 (To appear).

Internet Draft

1 Muhammad Niswar, Shigeru Kashihara, Kazuya Tsukamoto, Youki Kadobayashi, and Suguru Yamaguchi, "Inter-domain WLAN Handover Managmenet for Multi-homed Mobile Node", *draft-niswar-wlan-multihomed-handover-00.txt*, Internet Draft, IETF, December 2009.