

CAPACITY DEFICIT AND LINK LOSS IN WLAN TO
CELLULAR VERTICAL HANDOFF

CAPACITY DEFICIT AND LINK LOSS IN WLAN TO
CELLULAR VERTICAL HANDOFF

BY

SEYED VAHID AZHARI,

B.SC. ENG. (COMPUTER ENGINEERING, IRAN UNIVERSITY OF SCIENCE
& TECHNOLOGY),

M.SC., (COMPUTER ENGINEERING, UNIVERSITY OF TEHRAN, IRAN)

DECEMBER 2007

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

© Copyright 2008 by Seyed Vahid Azhari,

B.Sc. Eng. (Computer Engineering, Iran University of Science &
Technology),

M.Sc., (Computer Engineering, University of Tehran, Iran)

All Rights Reserved

DOCTOR OF PHILOSOPHY (2008)
(Electrical and Computer Engineering)

McMaster University
Hamilton, Ontario

TITLE: Capacity Deficit and Link Loss in WLAN to Cellular Vertical Handoff

AUTHOR: Seyed Vahid Azhari,
B.Sc. Eng. (Computer Engineering, Iran University of Science & Technology),
M.Sc., (Computer Engineering, University of Tehran, Iran)

SUPERVISOR: Dr. Terence D. Todd, Professor of Electrical and Computer Engineering, B.A.Sc., M.A.Sc., Ph.D. (University of Waterloo)

NUMBER OF PAGES: xv,128

Abstract

Mobile handset manufacturers have begun to include wireless LAN (WLAN) interfaces in their cellular handsets. This allows users to access WLAN networks when they are available and to revert to conventional cellular communications otherwise. In this way the handset can dynamically use the “best” available network, by switching connections between the two network interfaces. This switching is referred to as a vertical handoff (VHO).

When handling real-time connections, handsets must be capable of performing a *seamless* vertical handoff. This occurs when the interface switching does not disrupt the quality of service requirements of the active connections. Vertical handoffs are generally time consuming, and this delay creates a difficult problem since WLAN coverage can be lost very abruptly.

In this thesis, we propose and investigate several methods of mitigating this problem. A solution based on using a Vertical Handoff Support Node (VHSN) is proposed. When the WLAN link is lost, the VHSN is able to quickly redirect packets through the local cellular base-station during the time that handoff is taking place. This approach can eliminate VHO link loss. It is shown that the act of WLAN-to-cellular handoff can result in a severe bandwidth deficit problem on the WLAN. A novel bandwidth reservation and securing mechanism is proposed which overcomes this problem and performs significantly better than schemes based on modified versions of the static guard channel scheme used in cellular networks. The work is characterized by simulation and analytic models which investigate the key performance aspects of this type of system.

Acknowledgements

I praise ALLAH for the knowledge he bestowed upon me to benefit from the opportunities he put forward for me. I praise him for putting me among such unique people without whom this work would not have been done the same way. I am grateful for the efforts and guidance of my supervisor, Dr. Terence D. Todd, who helped me through my research in a unique way. His insight and constructive criticism played a major role in my thesis. I also thank the members of my supervisory committee Dr. Dongmei Zhao and Dr. Steve Hranilovic for pointing out matters that contributed to the enrichment of this work. The many helpful discussions on various queueing theory issues with Dr. Douglas Down also contributed to this work. I would also like to thank my dear colleagues, in particular my friend Mohammad Smadi with whom I spent such a joyful time working on joint projects in the lab and benefited from the many detailed discussions that we had. Also many thanks to the staff at our department which help me and many more students every day with our matters. Last but not least, behind every man's success is a woman's love, care, patience and sacrifice. I would like to thank my wife for all she did during my period of study here in Canada. I honor her for all the home sickness, loneliness and work she put up with just to make my life easier. This work would not have been possible without her understanding and support. Finally, to my parents, I would like to express my humble gratitude for making me what I am through years of constant caring for me and doing their best to teach me the values on which I have based my life.

Contents

Abstract	iii
Acknowledgements	iv
1 Introduction	1
1.1 Overview	1
1.2 Motivation	2
1.3 Scope of Thesis	2
1.4 Organization of Thesis	3
2 Background	5
2.1 Wireless Cellular Networks	5
2.2 Wireless Local Area Networks	10
2.3 Vertical Handoff in Wireless Networks	16
2.3.1 Basic VHO Problem	16
2.3.2 Hard vs. Soft Vertical Handoff	18
2.3.3 Vertical Handoffs in the Context of the Internet Protocol Stack	19
2.3.4 WLANs and Cellular Network Integration Architectures	24
2.4 The Effect of WLAN Deployment on WLAN to Cellular VHO Performance	27
2.5 Vertical Handoff Triggering Techniques	29
2.6 Summary	31

3	The Vertical Handoff Support Node	32
3.1	System Architecture	32
3.2	The VHSN Solution	35
3.3	VHSN Performance Analysis	40
3.3.1	Packet Loss Performance	43
3.3.2	VHSN Overhead and Scalability	44
3.3.3	Call Blocking and Dropping Performance	46
3.4	Simulation Results	48
3.5	Summary	58
4	The WLAN Capacity Deficit Problem	60
4.1	System Model	61
4.2	Limited Data-rate Usage	64
4.3	WLAN Capacity Deficit	64
4.4	The Cellular Guard Channel Scheme	67
4.5	Performance Evaluation	68
4.6	Modeling and Analysis	70
4.7	Summary	77
5	Transient Bandwidth Reservation	79
5.1	Performance Evaluation	83
5.2	Modeling and Analysis	86
5.3	Effect of Vertical Handoff Latency and Link-Dropping Rate	98
5.4	Effect of Boundary AP Load	101
5.5	Effect of Vertical Handoff Rate	103
5.6	Effect of MFH Signal Level Threshold	107
5.7	Effect of Wireless Propagation Parameters	110
5.8	Effect of Call Bitrate	114
6	Conclusion	119
	Bibliography	122

List of Tables

3.1	Vertical Handoff Delay Parameters	42
3.2	Percentage of Additional VHSN Bandwidth Overhead Normalized to Cellular Call Bandwidth	45
3.3	VHSN Scalability, Max λ_{VHO} (1/sec) at 0.1% Rejection Probability . . .	46
3.4	Parameter Values	51
4.1	Relationship Between a Single G.711 VoIP Call Bitrate, RSSI, and its Bandwidth Consumption on an IEEE 802.11g AP.	61
4.2	Simulation Parameters	66
4.3	Performance Evaluation of the LDU scheme.	67
5.1	6Mbps WLAN Deployment Parameters	111

List of Figures

2.1	<i>A Mobile Station Communicates to Others Through a Central Radio Tower.</i>	6
2.2	<i>The Concept of Cellular Networks and Frequency Reuse. Each Cell is labelled With its Frequency Set.</i>	7
2.3	<i>Handoff in a Cellular Network.</i>	9
2.4	<i>The Configuration of an IEEE 802.11 Independent Basic Service Set (IBSS).</i>	12
2.5	<i>The configuration of two IEEE 802.11 Basic Service Sets (BSS) and the Resulting Extended Service Set (ESS).</i>	14
2.6	<i>Hybrid Wireless Networks.</i>	15
2.7	<i>Dynamics of a WLAN to Cellular Vertical Handoff.</i>	17
2.8	<i>Dynamics of a Soft Handoff.</i>	19
2.9	<i>Basic Mobile IP Architecture.</i>	21
2.10	<i>Loosely Coupled WLAN/Cellular Integration Architecture</i>	26
3.1	<i>Loosely Coupled WLAN/Cellular Integration with Enterprise Anchor Node</i>	33
3.2	<i>Vertical Handoff Success Probability as a Function of its Latency, for Some Typical WLAN Deployments [SAT06].</i>	34
3.3	<i>GSM Call Establishment Time Results</i>	36
3.4	<i>VHSN Fast Link Recovery Mechanism</i>	37
3.5	<i>VHSN Handoff Timing.</i>	38
3.6	<i>Reservation Based TDMA Radio MAC Layer</i>	41
3.7	<i>Markov Process for Cellular Base Station Channel Occupancy (VHSN Case).</i>	48
3.8	<i>Diagram Showing the System Model Used for Simulation.</i>	50
3.9	<i>Simulation Results of New Call Blocking and VHO Dropping Probabilities. Comparison of VHSN-Assisted and Conventional Soft Handoff.</i>	52

3.10	<i>Analytic Results for New Call Blocking and VHO Dropping Probabilities. Comparison of VHSN and Conventional Soft Handoff.</i>	53
3.11	<i>Comparison of Analysis and Simulation.</i>	54
3.12	<i>Mean and Standard Deviation For Packet Loss During a VHO for (a) the Non-VHSN and (b) the VHSN-Assisted Vertical Handoffs vs. the VHO Arrival Rate.</i>	56
3.13	<i>Pdf of VHO Packet Loss</i>	57
3.14	<i>Mean (a) and pdf (Simulation), (b) of Initial Delay vs. VHO Arrival Rate.</i>	58
3.15	<i>Series of Events and Packet Sequence Numbers Received by the MS During (a) Conventional Soft and (b) VHSN-Assisted Vertical Handoff.</i>	59
4.1	<i>Floor Plan of the Coverage Area Considered. There Are Four APs One of Which Covers the Exit and is Named BAP. The Coverage Area of Each AP Corresponding to S_{min} is Shown. Also Active MSs Holding Calls, Which are Distributed Throughout the Coverage Area are Shown. In Addition a Single MS is Performing a VHO While Exiting the Building.</i>	62
4.2	<i>BAP Blocking and Dropping Probabilities vs. Fraction of Static Guard Bandwidth Reservation.</i>	69
4.3	<i>Markov Chain Modeling the State of an AP With Queueing of HHO Requests.</i>	72
4.4	<i>Steady State pmf for the Markov Chain in Figure 4.3 and Conditional State pmf Given a VHO Occurrence.</i>	75
4.5	<i>Comparison Between the Analytical and Simulation Results for the Static Reservation Scheme.</i>	77
5.1	<i>Block Diagram for the TR-MFH Scheme.</i>	80
5.2	<i>The Extended MFH-Region in the TR-MFH Scheme.</i>	82
5.3	<i>Boundary AP New Call Blocking, HHO Dropping and VHO Dropping versus Transient Guard Bandwidth Reservation Using Momentary Forced HHO (TR-MFH).</i>	83
5.4	<i>The WLAN Deployment Showing the BAP Coverage Region and the MFH Region for Other APs. The Shaded Area Indicates the Valid MFH Region.</i>	90

5.5	<i>The Shaded Area is Where Valid MFH-Users May Return to the BAP After Being Forced to Perform a HHO to a Neighbouring AP.</i>	92
5.6	<i>Comparison Between the Analytical and Simulation Results for the TR-MFH Scheme.</i>	95
5.7	<i>Analytical (a), and Simulation (b) Results for Mean and Standard Deviation of the Eventual Number of MFH-Calls, N_{mfh}, vs. the Fraction of Transient Bandwidth Reservation at the BAP. Also Included on the Right Side Axis is the Probability of the No-Deficit Case.</i>	97
5.8	<i>Boundary AP New Call Blocking, HHO Dropping and VHO Dropping vs. Transient Guard Bandwidth Reservation Using Momentary Forced HHO (TR-MFH) When VHO Latency is 7 Seconds.</i>	99
5.9	<i>Probability of Minimum Bitrate During a VHO for Various Values of VHO Latency.</i>	100
5.10	<i>VHO Capacity-Dropping and New Call Blocking Rates for the BAP vs. per AP New Call Arrival Rate With the Other Parameters as in Table 4.2, When No Reservation Scheme is Employed.</i>	102
5.11	<i>Performance of the SR Scheme Under a high VHO Arrival Rate, $\lambda_{vho} = 0.168/sec$.</i>	104
5.12	<i>Performance Evaluation of the Transient Bandwidth Reservation Approach for Increasing Values of λ_{vho}, When the Reservation Fraction is (a) 15% and (b) 30%.</i>	105
5.13	<i>The Effect of S_{mfh} (in dBm) on the Likelihood of Having a Candidate-Constrained Case (a) and the Average Free Call Capacity at the BAP After Running the MFH Procedure (b).</i>	108
5.14	<i>Probability That a Given Call Associated With the BAP is a Valid MFH Candidate for Different Values of S_{mfh}.</i>	110
5.15	<i>Influence of Static (a) and Transient (b) Bandwidth Reservation Schemes When Different Deployment and Wireless Propagation Parameters are Used.</i>	113
5.16	<i>Fraction of an IEEE 802.11g AP Bandwidth Consumed by One Call as a Function of Its Bitrate for Some Well-Known Voice Codecs.</i>	115

5.17 *Performance Comparison of SR and TR-MFH Schemes for Different Voice
Codecs: (a) G.711 With Header Compression and (b) G.723.1.* 116

5.18 *Performance of SR and TR-MFH Schemes for the G.711 Codec* 117

List of Acronyms

2.5G Second and a Half Generation Wireless Telephone Technology

3G Third Generation Wireless Telephone Technology

3GPP Third Generation Partnership Project

4G Fourth Generation Wireless Telephone Technology

AN Anchor Node

AP Access Point

ARP Address Resolution Protocol

BAP Boundary Access Point

BER Bit Error Rate

BS Base Station

BW Bandwidth

CDC Candidate Constrained Case

CN Called Node

CoA Care-of Address

CPC Capacity Constrained Case

EDGE Enhanced Data rates for GSM Evolution

EXC Exact Case

FA Foreign Agent

GAN Generic Access Network

GGSN Gateway GPRS Support Node
GPRS General Packet Radio Service
GSM Global System for Mobile Communications
GW Gateway
HA Home Agent
HHO Horizontal Handoff
HO_SIG Handoff Signalling Packet
IAPP Inter Access Point Protocol
IP Internet Protocol
ISM Industrial, Scientific, Medical band
L2 Layer Two
L3 Layer Three
LAN Local Area Network
LDU Limited Data-rate Use
MAC Medium Access Control
MFH Momentary Forced Horizontal Handoff
MIP Mobile IP
MOS Mean Opinion Score
MS Mobile Station
MSR Mobile Specific Routers
NDF No Deficit Case

nFA New Foreign Agent

NR No Reservation

oFA Old Foreign Agent

PBX Private Branch Exchange

PDP Packet Data Protocol

PHY Physical Layer

PLMN Public Land Mobile Network

PSTN Public Switched Telephone Network

RSSI Received Signal Strength Indicator

SGSN Serving GPRS Support Node

SIP Session Initiation Protocol

SR Static Reservation

TCP Transport Control Protocol

TDMA Time Division Multiple Access

TR Transient Reservation

TR-MFH Transient Reservation with Momentary Forced Horizontal Handoff

UDP User Datagram Protocol

UMA Unlicensed Mobile Access

UMTS Universal Mobile Telecommunications System

VHO Vertical Handoff

VHSN Vertical Handoff Support Node

VoIP Voice over IP

WLAN Wireless Local Area Network

Chapter 1

Introduction

1.1 Overview

Cellular networks are among the most significant technologies developed in the communications field. These types of networks were originally designed to carry voice traffic using a circuit-switched core network that is connected to the public switched telephone network (PSTN). In recent years, cellular networks have evolved so that they now offer both voice and data services, which includes packet-switched connectivity to the Internet. Radio coverage in cellular networks is provided by basestations whose coverage area is typically very large (several kilometres), and for this reason their maximum per user throughput is relatively low (e.g., <1Mbps).

More recently, wireless local area network (WLAN) technology has been commercialized. WLANs are an outgrowth of local area networks (LANs) such as Ethernet, which first appeared in the 1980s. WLANs are capable of providing much higher user throughput than that in cellular networks (>1Mbps), but over far shorter distances (e.g., <150m). These types of networks are suitable for coverage situations where user mobility is relatively low.

Cellular handset manufacturers such as Nokia and Motorola have now started to include WLAN interfaces in their cellular handsets. This type of device is referred to as a *dual-mode handset*. When WLAN coverage is available, dual-mode handsets can be used to make voice calls through the WLAN using voice-over-IP (VoIP). When

a user moves out of WLAN coverage during an active voice connection, the handset must switch the connection to its cellular network interface. This switching of radio access networks is called a vertical handoff (VHO), and is ideally done in a seamless fashion without packet loss or noticeable delay. This is a difficult objective due to the time consuming processes involved in network switching, and because of the speed with which WLAN coverage can be lost.

1.2 Motivation

In future wireless packet switched networks, real-time VHO is a necessity. In the cellular-to-WLAN direction, existing VHO techniques which are based on received signal strength monitoring are suitable due to the ubiquity of cellular coverage in most situations. The focus of this thesis is on vertical handoffs that occur in the WLAN-to-cellular direction. Due to the abrupt behaviour of WLAN signal propagation at its coverage boundaries, voice calls are highly susceptible to link-loss during a VHO. Therefore, it is important to understand the effect of the WLAN on VHO performance for real-time applications. The focus is restricted to voice as it is one of the predominant network applications with real-time requirements.

1.3 Scope of Thesis

A soft VHO scheme is adopted, based on the loosely-coupled integration architecture [ETS01]. In this type of setup, the WLAN and cellular access networks are independently connected to the public Internet or PSTN. The problem of WLAN-to-cellular VHO is considered from the WLAN link viewpoint as a mobile station moves away from the WLAN during vertical handoff.

A mobile station (MS) associated in the WLAN is likely to experience abrupt coverage loss as it transitions to the cellular network during a VHO [SAT06]. This is due to the non-graceful WLAN signal deterioration at the boundaries of its coverage area. An independent cellular client station is proposed which connects to the WLAN and assists the MS in performing a VHO in the face of WLAN link loss.

An additional solution is developed to prevent WLAN link loss and it is shown that this solution can cause capacity deficit problems for the WLAN access points. A solution is provided for this problem, which involves bandwidth reservation and bandwidth securing.

1.4 Organization of Thesis

In the next chapter a brief review of the literature is given related to the basics of vertical handoff. Soft handoff techniques and the effect of VHO on the IP protocol stack are covered as well as integration models for WLANs and cellular networks. An overview of WLAN deployment techniques is then provided, and how they relate to VHO performance. The chapter ends by surveying some well-known handoff triggering techniques.

Chapter 3 proposes the use of a Vertical Handoff Support Node (VHSN), which improves the VHO link loss rate. This chapter provides a detailed and illustrative explanation of the operation of the VHSN along with analysis and simulation results which evaluate its performance.

Chapter 4 introduces the capacity deficit issue in WLANs which arises as a VHO call reduces its bitrate while still connected to the WLAN. It is shown that VHO calls can be dropped in the WLAN due to a lack of bandwidth. In addition, the Static Reservation (SR) approach, which is a modified version of the well-known guard channel reservation scheme used in cellular networks is adopted to improve the handoff performance. Using simulation and analysis it is shown that this scheme eliminates the capacity deficit problem but at the cost of an unacceptable new call blocking rate in the WLAN.

In Chapter 5 the Transient Reservation with Momentary Forced Handover, TR-MFH, mechanism is proposed. TR-MFH performs bandwidth reservation in a novel way and does not cause the new call blocking rate to be significantly increased. TR-MFH is used with a novel bandwidth securing mechanism which enables it to eliminate the WLAN capacity deficit problem in many cases at the cost of a small increase in the new call blocking rate. Simulations and analysis are used to show the

improvements gained using this approach. The effect of various system parameters on the performance of the TR-MFH scheme are also studied in this chapter. We then conclude the thesis in Chapter 6.

Chapter 2

Background

This chapter provides an overview of the concepts relevant to a vertical handoff. The chapter starts by giving a short background on cellular and wireless local area networks (WLAN). The concept of a vertical handoff is then explained outlining soft and hard vertical handoff approaches. In addition, the effect of vertical handoff on the Internet protocol stack is addressed. This is followed by introducing two well-known architectures for inter-networking between WLANs and cellular networks and how this affects a vertical handoff. The influence of the WLAN deployment on vertical handoff performance is addressed and finally a survey of various handoff triggering schemes is included.

2.1 Wireless Cellular Networks

Early wireless mobile communication networks were mainly used by the police and various private organizations. In these networks, mobile stations (MS) communicated through a single radio tower (base station) which provided coverage for a very large geographical area (Figure 2.1). In order to support multiple calls, frequency division multiple access (FDMA) was used and each call was assigned a separate frequency channel. Due to the limited number of frequency channels allocated to these networks, they were incapable of supporting many simultaneous calls.

According to the physics of wave propagation, in free-space the power of a wireless

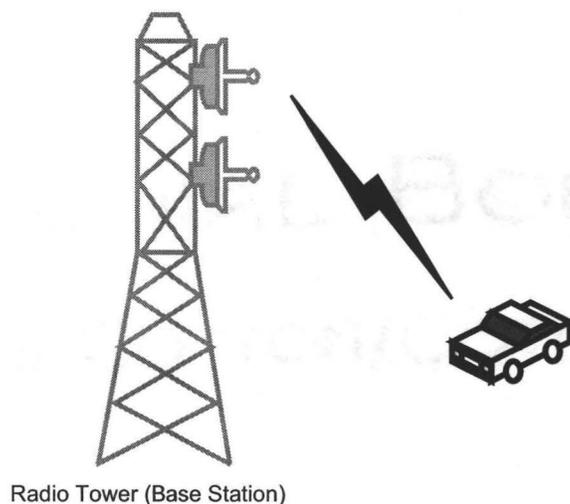


Figure 2.1: A Mobile Station Communicates to Others Through a Central Radio Tower.

signal decays proportionally to the square of the distance between the transmitter and the receiver [Rap96]. Therefore, a transmission at a given location can not be heard by a receiver that is sufficiently distant. This property allows a frequency channel to be used independently at several distant locations and is called the frequency reuse technique [Rap96], illustrated in Figure 2.2. Here, many base stations (BS) with lower transmission power and smaller coverage area are deployed over a large geographical region. The coverage area of each BS is called a cell and such networks are called *Cellular Networks*. It should be noted that cell coverage boundaries are not perfect and may have irregular shapes as well as small overlaps, however, they are commonly shown as regular hexagons for convenience. The existing frequencies are divided into a number of disjoint frequency sets and each cell is allocated one of these sets. To reduce interference, the frequency allocation is done such that no two adjacent cells use the same frequency set. Therefore, with the same number of frequency channels it is possible to cover a much larger area and increase the number of simultaneous calls.

Since cellular networks use different frequencies in adjacent cells, a MS bearing a call has to switch its frequency channel once it moves from the coverage area of

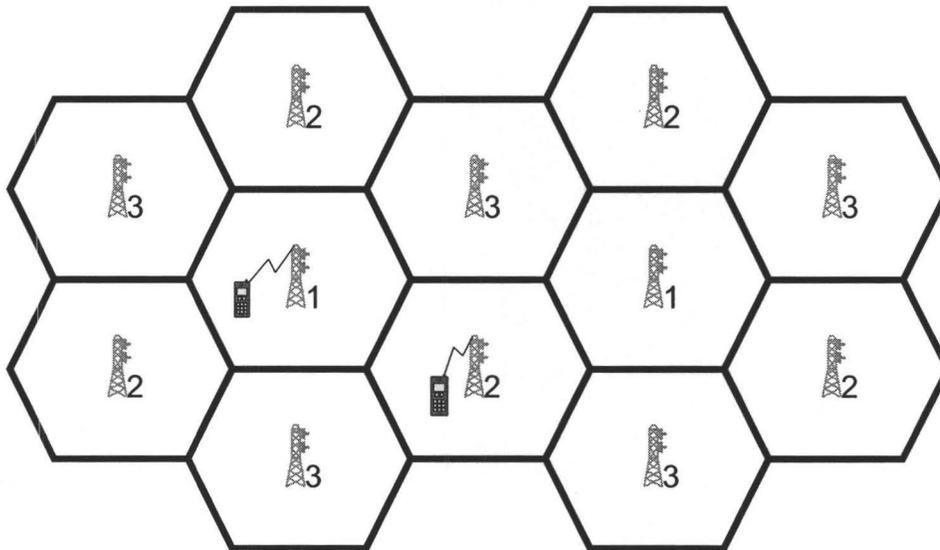


Figure 2.2: The Concept of Cellular Networks and Frequency Reuse. Each Cell is labelled With its Frequency Set.

one cell to the other. This event is called a handoff and is dealt with using mobility management schemes. Cellular networks tend to use proprietary mobility management schemes. A few standard mobility management schemes such as Mobile IP and SIP will be discussed in Section 2.3.3. A handoff is generally triggered based on the value of signal strengths received from adjacent and current base stations. A mobile station constantly calculates the power of the received signal also known as the received signal strength indicator (RSSI) from its current BS. It also periodically checks the RSSI value for all adjacent BSs. Whenever the current RSSI value drops below a certain threshold, S_{tp} , the MS will attempt to do a handoff by connecting to one of the available BSs with a stronger RSSI. Due to the random nature of signal behaviour and the unpredictable movement of the MS, the MS may switch back and forth between two adjacent BSs when their RSSI values are very close. This is called the *ping-pong* effect [PKH⁺00]. To overcome this, a hysteresis mechanism is used, whereby the MS will only handoff to a new BS when its current RSSI is below S_{tp} and the RSSI from the new BS is larger than S_H . Having S_H sufficiently larger than

S_{tp} significantly reduces the possibility of the ping-pong effect. Figure 2.3 illustrates the concept of handoff in cellular systems including the hysteresis margin.

The *cellular core network* interconnects base stations and is responsible for delivering user traffic to the BS. When a MS performs a handoff, its call is redirected to the new BS. This is the responsibility of the core network. The core network also takes care of other functions such as user authentication and performance monitoring. User authentication is the mechanism by which a user's identity is verified by the network and performance monitoring concerns the collection of statistics for various performance metrics such as packet delay and loss.

Today's cellular networks have evolved through many generations of technological advances. First generation (1G) cellular networks are characterized as being purely analog [Rap96]. These systems use FDMA to communicate with different MSs. For example, the Advanced Mobile Phone System (AMPS) is a 1G cellular phone system developed by Bell Laboratories and was introduced in North America at the beginning of the 1980s. AMPS operates in the 800 MHz band.

Second generation (2G) cellular technology is mostly based on digital communication and signal processing techniques that improve call quality and network capacity. Many second generation systems are based on the Time Division Multiple Access (TDMA) scheme. In TDMA a periodic time frame is divided into a number of slots each of which is assigned to a MS for communicating to the BS. Some 2G networks combine TDMA and FDMA. D-AMPS is an example of a 2G cellular network which was developed based on AMPS [Rap96].

Another example of a 2G network is the Global System for Mobile Communications (GSM) which is currently one of the most popular cellular networks [Rap96]. According to a recent estimate, nearly two billion of the current 2.4 billion cellular subscribers use GSM technology [Cel]. GSM is based on a combination of TDMA and FDMA and usually operates in the 1800 MHz or 900 MHz frequency bands. The GSM network provides circuit-switched voice services to users, but by 1997, the General Packet Radio Service (GPRS) was integrated into the GSM standards, which allows for the transport of packet-switched data traffic over existing GSM infrastructure. In order to support GPRS, additional functional entities are introduced into the GSM

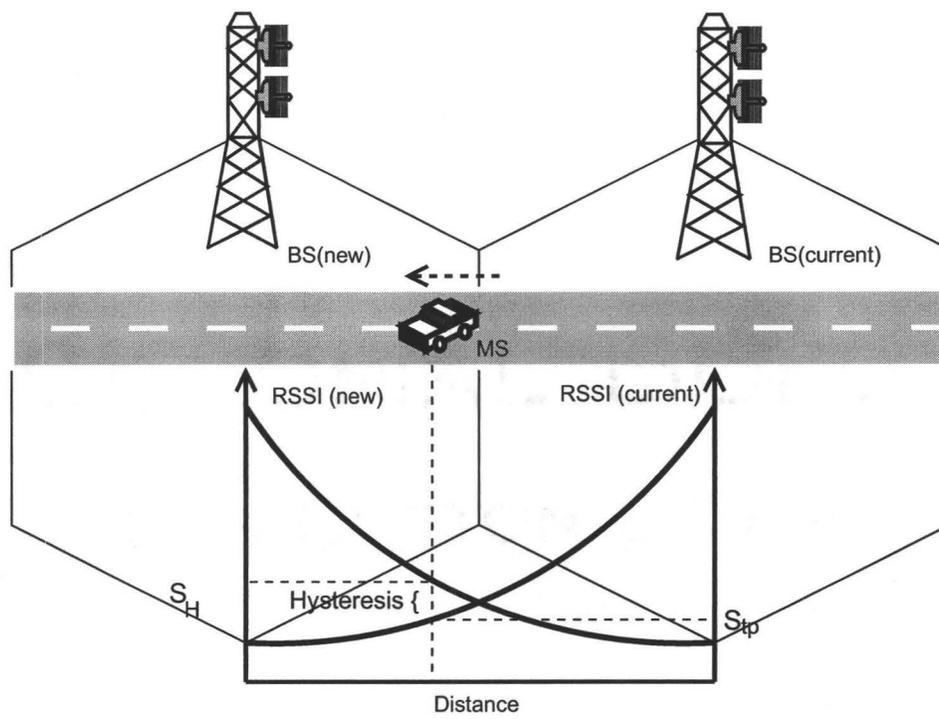


Figure 2.3: Handoff in a Cellular Network.

core network to deal with packet-switched services. GPRS uses the same physical layer as the GSM network but has its own protocol stack above the link layer.

The main goal of third generation (3G) cellular networks is to provide users with more bandwidth [Rap96]. This enables cellular operators to provide users with rich packet switched services such as music download and live TV. The third generation of cellular mobile networks became commercially available in 2001 when NTT DoCoMo, Japan, launched its network. Universal Mobile Telecommunications System (UMTS) is also another 3G cellular mobile technology which is the successor to GSM and GPRS. Cellular networks are still evolving and the deployment of 3.5G technologies and 4G networks are currently under development.

Although various cellular systems are different in terms of bandwidth and access technology, they all share similar attributes. Since cellular networks provide coverage over a relatively large geographical area (several kilometres), they are suitable for highly mobile users such as car and train passengers. In addition, due to sophisticated coding and error correction schemes used at the physical layer, they tend to provide very reliable links. However, these links operate at fairly low bitrates. For example, a GSM link can typically operate at a bitrate of 9.6Kbps and a UMTS link can provide a typical bitrate of 3.6Mbps on its downlink. These values are small compared to other technologies such as wireless local area networks which can provide bitrates of up to 54Mbps.

2.2 Wireless Local Area Networks

The use of wireless technology for carrying packet switched data traffic dates back to the ALOHAnet project at the University of Hawaii in 1970 [Abr85]. In that project, several computers distributed over four islands were linked wirelessly. Due to the high cost of radio communication networks, deploying them for a small area was not cost effective at the time. However, with the advances in electronic chip manufacturing technology and the development of more efficient wireless devices, the cost of such networks decreased.

The first wireless local area network (WLAN) standard was ratified in 1997 by

working group 11 of the IEEE LAN/MAN Standards Committee and is known as the IEEE 802.11 standard. The IEEE 802.11 standards define the physical (PHY) and medium access control (MAC) layers for an IEEE 802.11 compliant WLAN. The PHY layer can either use the 2.4 GHz unlicensed Industrial Scientific Medical (ISM) radio frequency (RF) band or infra-red (IR) for communication. The 1997 IEEE 802.11 standard provided a maximum bitrate of 2 Mbps.

IEEE 802.11 was amended by the IEEE 802.11b standard released in 1999, which operates in the 2.4 GHz ISM band and introduced higher bitrates of 5.5 and 11 Mbps by using more efficient signal modulation techniques. IEEE 802.11b operates in the same frequency band as many other wireless devices such as microwave ovens, Bluetooth, and cordless phones, and can experience interference from these devices. In addition, the 2.4 GHz ISM band only allows for three non-overlapping channels, which limits the capacity of a WLAN deployment based on IEEE 802.11b. As a result, in 1999 the IEEE 802.11a standard was also ratified which operates in the 5 GHz unlicensed ISM band. An IEEE 802.11a compliant device supports bitrates of up to 54 Mbps. In addition, the 5 GHz band allows for up to 12 non-overlapping channels. These two features allow for a much higher capacity WLAN deployment. However, the IEEE 802.11a standard is not compatible with IEEE 802.11b due to the use of different frequency bands. Therefore, in June 2003 the IEEE 802.11g standard was ratified which operates in the 2.4 GHz band and is backward compatible with IEEE 802.11b. This technology supports IEEE 802.11b bitrates of 1, 2, 5.5 and 11 Mbps in addition to bitrates of up to 54 Mbps.

The Medium Access Control (MAC) layer is responsible for arbitrating access to the medium of a group of IEEE 802.11 compliant wireless devices. The MAC layer for all IEEE 802.11a/b/g standards is the same. The IEEE 802.11 standard has defined two access schemes within the MAC layer, a centralized contention-free polling based Point Coordination Function (PCF) and a distributed contention based Distributed Coordination Function (DCF). The DCF uses the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) method [KT75]. The DCF is currently implemented in all IEEE 802.11 products while PCF is not commonly used.

Although IEEE 802.11a/b/g standards are adopted and used in many devices, the

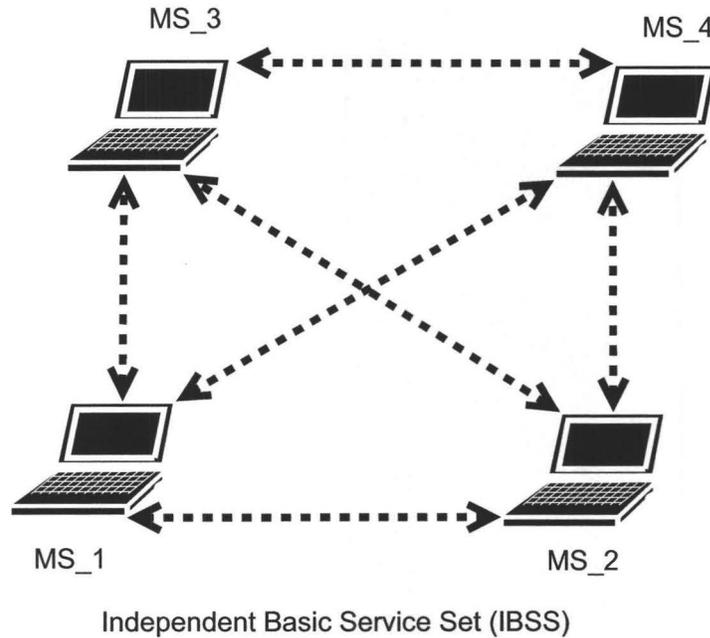


Figure 2.4: *The Configuration of an IEEE 802.11 Independent Basic Service Set (IBSS).*

widely used DCF access scheme does not provide any Quality of Service (QoS) and deals with flows as best effort traffic. This puts limitations on the capability of IEEE 802.11 devices to support real-time applications such as packet voice. To address this issue, the IEEE 802.11 working group introduced the IEEE 802.11e standard in 2005. This standard defines a new enhanced coordination function: the Hybrid Coordination Function (HCF). Within HCF, two channel access methods are defined, namely, the HCF Controlled Channel Access (HCCA) and Enhanced Distributed Channel Access (EDCA). Both EDCA and HCCA provide differentiated service to flows by defining traffic classes and assigning them different priority levels. The MAC layer is designed such that higher priority traffic classes have a higher chance of seizing the channel. Real-time applications can send their data as high priority traffic and benefit from the better performance experienced by this class of traffic.

The IEEE 802.11 standard defines two operating modes for a network, the Independent Basic Service Set (IBSS) and the Basic Service Set (BSS). Figure 2.4 shows

the configuration of an IBSS in which four mobile stations form a network by directly communicating. Such a system is called an ad hoc network. By comparison, a BSS is identified by an IEEE 802.11 device called an access point (AP). Mobile stations connect to an AP and use it to communicate to other connected mobile stations or nodes beyond the AP. Such MSs are said to be associated with the AP. In a BSS, all MSs first send their packets to the AP. If the packet destination is a MS currently associated to the same AP, the AP delivers the packet directly, otherwise the AP sends the packet to the Distribution System (DS). The DS is responsible for connecting different BSSs together and to the local network (e.g., a LAN). Figure 2.5 shows two BSSs which are connected through a wired distribution system. When multiple BSSs are connected in this way, the configuration is referred to as an Extended Service Set (ESS).

Each IEEE 802.11 AP functions as a wireless bridge and together with the DS, the entire IEEE 802.11 WLAN operates as a bridged LAN. This simplifies integrating WLANs into existing LAN infrastructure. In addition, IEEE 802.11 WLANs have low deployment costs and provide high capacity. These features have led to the widespread adoption of IEEE 802.11 WLANs.

Many companies are starting to use Voice-over-IP (VoIP) to place phone calls over their WLAN. For example, Cisco has introduced its wireless IP phone 7920 [Cisa] which is an IEEE 802.11b handset that supports VoIP. In 2005 the WLAN VoIP phone market reached nearly \$72 million with Cisco leading the revenue market [WiF]. Although much less than cellular phone market sales, the WLAN phone market showed a 327% sales increase in 2006 compared to only a 13% increase in cellular mobile phone sales [WiF].

When an MS moves between the coverage area of two adjacent WLAN APs, it has to perform a handoff before losing its link. To support real-time applications such as VoIP, fast handoff mechanisms that do not interrupt the flow of packets have been developed [MSA04] [SMA04] [RS05] [VK04]. Cisco Centralized Key Management (CCKM) is a proprietary technology which enables fast roaming of IEEE 802.11 clients with support for security based on the Remote Authentication Dial-In User Service (RADIUS).

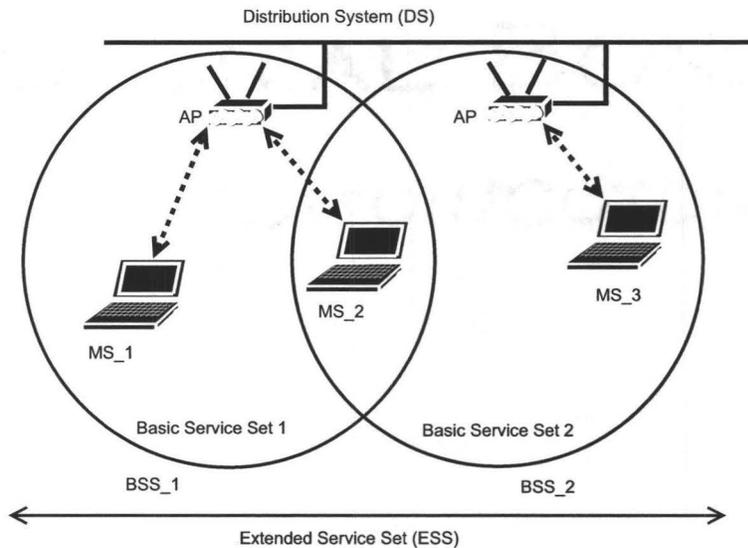
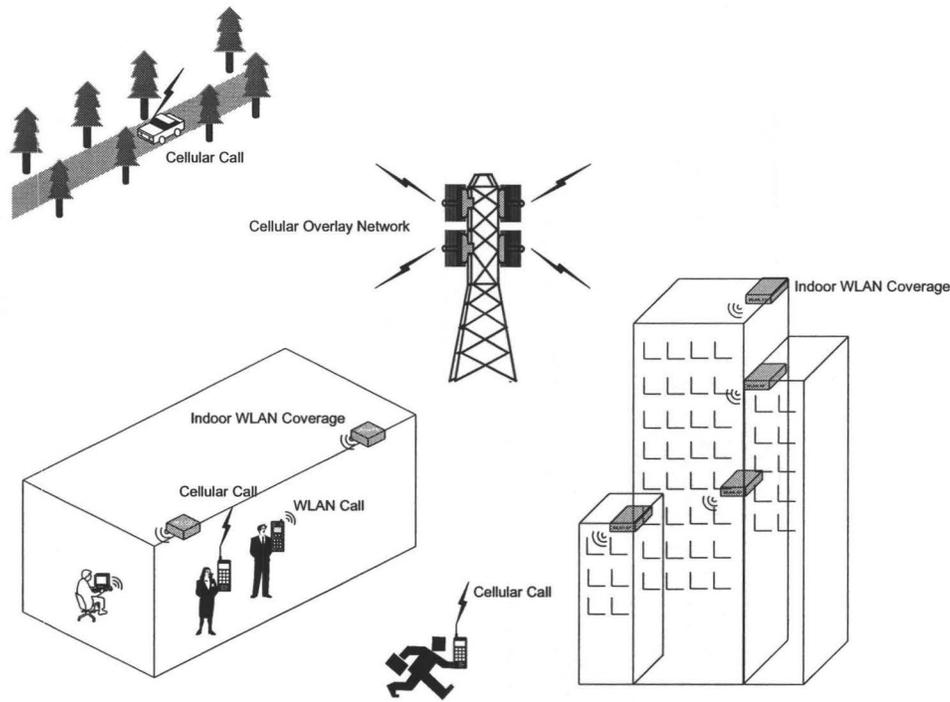


Figure 2.5: The configuration of two IEEE 802.11 Basic Service Sets (BSS) and the Resulting Extended Service Set (ESS).

Wireless LAN coverage is generally not as pervasive compared with that of cellular. The coverage range of a WLAN AP is only tens of meters, much smaller than the several kilometre range which is typical of a cellular base station. Also, WLAN signal strength at its coverage boundaries can be highly variable and a link can be lost abruptly. Therefore, to effectively use handoff mechanisms, a WLAN should be carefully deployed such that coverage is not abruptly lost at its periphery. Because of their low cost, high capacity, and local coverage, WLANs are suitable for small densely populated areas with relatively stationary users. Cellular networks on the other hand, can be used to provide ubiquitous coverage over large areas, but cellular service is more expensive and offers lower data rates compared to WLAN. Due to the complementary features of WLANs and cellular networks, future wireless networks will be hybrid combinations of the two, i.e., ubiquitous cellular coverage which overlays islands of WLAN coverage such as the example shown in Figure 2.6.

In a hybrid overlay network, users can always use the cellular network. They can also use WLAN coverage wherever it exists, taking advantage of its high data rate

Figure 2.6: *Hybrid Wireless Networks.*

at a low cost. In order to do this, dual-mode handsets are required which integrate access to both networks in a single device. A 2006 study estimates that the number of WLAN deployments will double by 2011 [Hot]. Also, according to a recent report, the fastest growing segment of the mobile phone market belongs to dual-mode handset sales [Dua]. From 2005 to 2006 the number of WLAN handset shipments had a five fold increase, with nearly 70% of the revenue from dual-mode handsets and the rest due to single-mode WLAN-only handsets.

Dual-mode handsets can connect to the “best” network that is available by switching their connections between different network interfaces. The act of switching a connection between two different wireless networks is called a vertical handoff (VHO). This is in contrast to a horizontal handoff (HHO) that occurs when a MS switches its connection between two access points of the same network. When a handset is

supporting a real-time voice call, a VHO has to be achieved without incurring excessive packet loss and outage. This may be difficult to achieve since a vertical handoff is generally time consuming and involves the exchange of numerous signalling messages between the handset and the wireless network. This will be further discussed in Section 2.3.3.

2.3 Vertical Handoff in Wireless Networks

In this section, the basic vertical handoff (VHO) process and its variations are described. The effect of a VHO on different layers of the network protocol stack and solutions that address these effects will also be discussed.

2.3.1 Basic VHO Problem

The switching of a mobile stations's wireless connection from one access network to the other is called vertical handoff ¹ (VHO). A *seamless* vertical handoff is defined as one that provides non-interrupted end-to-end communications during the VHO, which is vital for real-time applications.

Normally, a handoff consists of three phases, namely, handoff detection, candidate network selection or scanning, and handoff execution. Vertical handoff detection is the task of triggering a notification that the mobile station's connection to the current network is not adequate and may soon be dropped. A commonly used and effective indication for a VHO event is the received signal strength indicator (RSSI) of the mobile station, where a VHO will be triggered whenever the RSSI drops below a certain threshold. VHO triggering will be further addressed in Section 2.5.

A VHO can be detected and initiated by the network or the MS, or it can be detected by one and initiated by the other. Traditionally in IEEE 802.11 WLANs, handoff detection and initiation are both handled by the MS. It is important that a handoff be detected early enough such that the scanning and execution phases can complete before the MS loses its link. On the other hand, triggering too early is not

¹For the sake of brevity we will also refer to a vertical handoff as simply a handoff or handover.

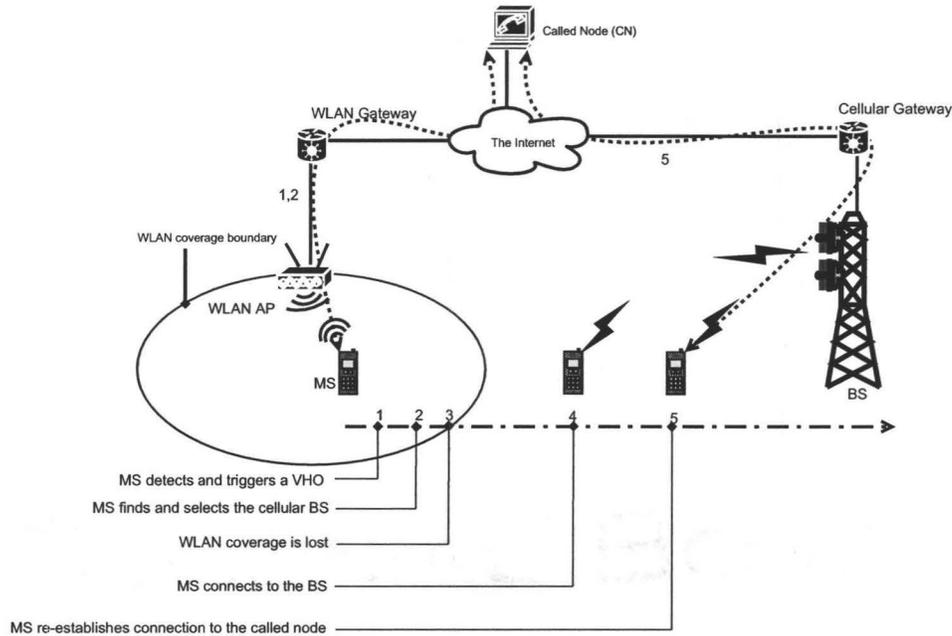


Figure 2.7: Dynamics of a WLAN to Cellular Vertical Handoff.

suitable either, as it may result in a VHO when a user is temporarily experiencing a low signal level. Such a situation is referred to as *false triggering*. Figure 2.7 illustrates an example of a vertical handoff that occurs while a MS is transitioning out of a WLAN coverage area. Here, the MS is communicating with a node which is referred to as the *Called Node (CN)*. The MS initially uses its WLAN interface for communication, but at some point, labelled '1', a VHO is detected and triggered by the MS.

After a VHO has been triggered and during the candidate network selection phase, a new *point of attachment* is chosen for the MS. This can be done by the network or the MS. For vertical handoffs where the current access network is unaware of neighbouring networks, it is common that the MS perform the candidate selection phase. In Figure 2.7, the MS scans for cellular coverage and selects a BS as its next point of attachment.

Once the new access network is selected, the VHO execution phase starts by

associating and authenticating the MS with the new access network. In Figure 2.7 this takes place between points '2' and '4' when the MS is establishing a link to the cellular BS. This is followed by re-establishing its current connection through the new access network ('4'-'5').

2.3.2 Hard vs. Soft Vertical Handoff

A handoff is referred to as *hard*, when the old connection is removed before the new one is established. This can happen for example, in mobile stations with a single radio interface that can only be connected to one access network at a time an example of which are regular cellular handsets. Hard handoff is generally easier to implement but suffers from packet loss and outage. The outage occurs from the time that the old connection is dropped until the new connection is established and as a result, hard handoff approaches are less suitable for real-time applications such as VoIP.

On the other hand, a *soft* handoff is such that the old connection is removed *after* the new one has been created². If carefully orchestrated, a soft handoff can provide zero packet loss and no outage, making it an attractive candidate for real-time applications. This however, requires the MS to maintain two radio links at the same time. In addition, separate traffic flows that arrive from the old and the new networks should be merged.

One way to implement a soft handoff is to have the other end of the connection, the called node (CN), duplicate the incoming packet flow through the new access network. Alternatively, a device with a connection to both the MS and CN can duplicate the packet flow without involving the CN. Such a device is commonly referred to as an *anchor node* (AN). Figure 2.8 illustrates an example of this in the case of WLAN to cellular VHO, where the MS to CN connection is realized by two subconnections, i.e., MS-AN and AN-CN. During a VHO, the AN creates an additional subconnection to the MS through the new access network while maintaining the other subconnections. The WLAN connection is operational before the VHO initiation and during VHO execution. It is then removed after the VHO is completed. The cellular connection

²This type of handoff is also referred to as seamless handoff in the literature.

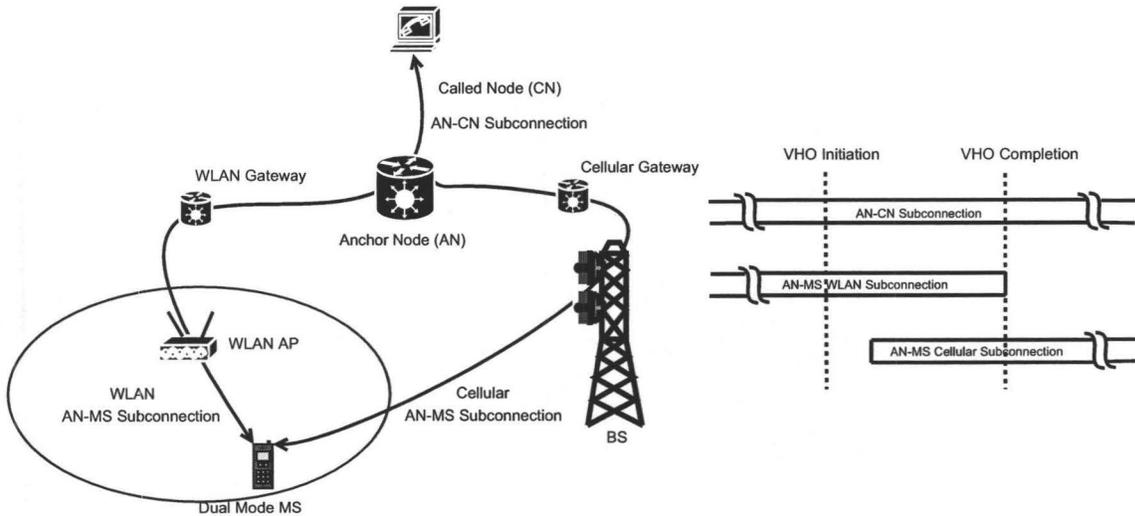


Figure 2.8: *Dynamics of a Soft Handoff.*

becomes operational some time during the VHO execution and remains so after the VHO is completed. A number of soft handoff schemes have been proposed within this framework [STZK06] and will be discussed in the next subsection.

2.3.3 Vertical Handoffs in the Context of the Internet Protocol Stack

Vertical handoffs can affect the entire TCP/IP protocol stack. At the physical and link layers the new radio interface has to be configured and a link to the new access point or base station has to be established. In addition, the IP address of the MS may also change after a VHO, in which case the called node should be notified of the new IP address.

The most prominent IP layer solution to mobility is Mobile IP (MIP) [Per02], which was proposed by the Internet Engineering Task Force (IETF). MIP hides IP address changes due to a handoff by requiring a MS to communicate with other nodes through a dedicated home agent (HA) on the MS's home network. In MIP, the MS is assigned two IP addresses, a home address and a care-of address (CoA). The home

address identifies the MS in its home network and is used to identify the MS to other nodes regardless of the MS's current location.

The CoA is assigned to the MS by the foreign network to which it is currently connected. The CoA is also used by the HA to communicate with the MS while the MS is visiting a foreign network. While in a foreign network, an entity called the Foreign Agent (FA) is responsible for receiving packets that are generated by the MS or destined to it. These packets are in turn sent to or received from the home agent. In practice, the FA may be implemented in the MS.

Figure 2.9 is a simple illustration of the Mobile IP architecture. Any node that wishes to send a packet to the MS will send it to its home address. These packets are then received by the MS's home agent and are forwarded to the MS's care-of address. This way, the rest of the network will not notice the change in the IP address of the MS. When the MS has a packet to send, it can encapsulate it in another packet destined to its HA and have the HA send the original encapsulated packet to its intended destination.

When a vertical handoff occurs, the MS will associate with a new Foreign Agent (nFA) and receives a new CoA from that nFA. It then informs its HA of the new CoA by a *Registration Request and Response* packet interaction. Following the registration process, a new path shown as a dashed line in Figure 2.9 will be used for communication between the MS/nFA and its HA.

As seen in Figure 2.9, these interactions cause triangular routing in the path between MS and CN, which leads to inefficient communication. In other words, the routing path between MS and CN is no longer a shortest path since it goes through the HA. Moreover, if many paths are redirected this way, the home agent will become a bottleneck. The route optimization extension to MIP [ea97] allows the MS to directly communicate with the CN using its CoA. After completion of the handoff procedure, the HA sends a *binding* update to the CN informing it of the new CoA for the MS. The CN then sends subsequent packets directly to the MS's new IP address.

A binding update can also be used to enable soft handoff. Before the flow of packets through the new foreign agent (nFA) starts, the MS requests that the nFA send a binding update to the old FA (oFA). The binding update contains the MS's

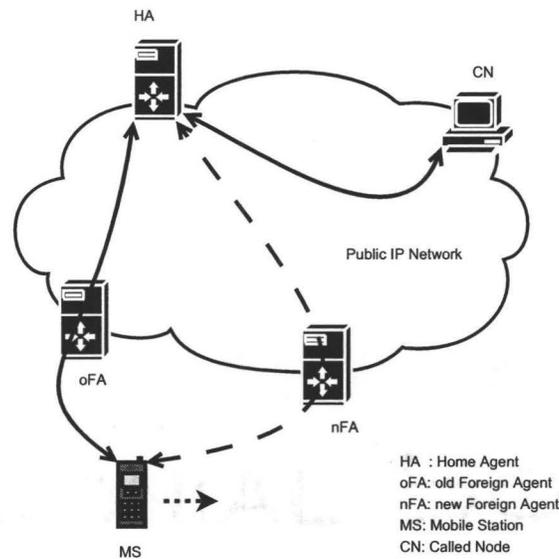


Figure 2.9: *Basic Mobile IP Architecture.*

new CoA, which is then used by the oFA to forward packets arriving on the old path to the MS. This reduces packet loss and outage. In another improvement to Mobile IP [ea99], a buffering and forwarding scheme is used at the foreign agents to reduce data loss during a handoff. The FA buffers all packets that it forwards to the MS. When a MS initiates a handoff, its new FA requests buffered packets from the old FA that belong to the MS. The new FA in turn forwards these packets to the MS. The idea of FA forwarding has also been presented in the context of a post-registration handoff scheme, which is an extension to Mobile IP [ea01].

To take advantage of these schemes, the foreign agents and the MIP protocol need to be changed. Another drawback of using such forwarding schemes for WLAN to cellular handoff is that the forwarding path is very lengthy. This increases packet delay and is not a suitable solution for real-time applications.

The authors in [BAA04] propose a mobility management solution based on mobile IP, which reduces the latency of MIP signalling by handling mobility events locally. This is achieved by adopting a hierarchical mobility management architecture to hide the movement of mobile stations within a foreign domain. As long as a MS is changing

foreign agents within the same domain, the home agent need not be notified. The home agent contacts the MS through the gateway FA responsible for the entire foreign domain. In addition, soft handoff can be achieved by having the gateway foreign agent multicast packets to the old and new foreign agents of a MS.

There are also network layer mobility solutions which are not based on MIP. The scheme in [MS98], uses a multicast address in the MS which receives advertisements from potential access points. Handoff initiation relies on the detection of periodic beacons from the different networks. A combination of analytic models and testbed experiments are used to evaluate VHO performance. It is shown that handoff latencies can easily be as high as 3 seconds and by that fast beaconing and packet/header double-casting, can be reduced to 800msec [MS98]. However, this value is still unacceptable given the 150-200 msec end-to-end delay requirement for real-time voice.

Transport and application layers also need to cope with handoff. During a VHO, the TCP entity in the MS may experience outage, which will be interpreted as network congestion. This will cause TCP to send retransmissions from the MS, which may also fail due to this outage. A prolonged outage period will cause the TCP entity in the MS to significantly reduce its window size. As a result, once the VHO is completed and the connection is restored, packet flow from the MS will resume with a small rate, taking a long time to send the backlogged packets to the CN. For a detailed discussion of TCP problems in wireless networks one can refer to [BSK95].

In [BCI04] an experimental testbed is used to evaluate the performance of vertical handoffs between GPRS, WLAN, and an Ethernet LAN. Mobile IPv6 [ea04a] is adopted as the mobility management scheme and the performance of UDP and TCP flows have been evaluated. This paper highlights how differences in network link characteristics can produce severe performance problems for TCP during a vertical handoff. Large vertical handoff delays of 4-5 seconds were experienced when moving between WLAN and GPRS networks. Using layer 2 events such as packet transmission failure to detect network coverage loss, VHO delay was reduced to about 2 seconds. This work illustrates the performance effects caused by VHO and the importance of using L2 triggers to provide earlier handoff triggering.

In [KCC05], seamless handover is achieved using a TCP migration scheme. However, this approach is only suitable for TCP applications and does not apply to real-time traffic which operates over UDP.

There are also schemes which tackle vertical handoff within the application layer. Such approaches require minimal assistance from the underlying network, which makes them more attractive than MIP-based schemes. One such mechanism has been presented in [ea02b] using proxy servers and mobile specific routers (MSR). All traffic to/from mobile stations goes through proxy servers which are collocated with MSRs. The MSRs maintain a mobile specific IP address table and are responsible for packet delivery. Since all mobile station traffic goes through the proxy servers and the corresponding MSR, mobility can be hidden from the rest of the network elements. This approach is not end-to-end but the proxy servers are application layer entities. However, this approach is proposed for web based traffic and it may not be suitable for real-time flows such as voice. Moreover, the approach does not provide soft handoff and may suffer from outage and packet loss.

The Session Initiation Protocol (SIP), developed by the IETF [ea02a], is an application layer protocol for establishing and maintaining sessions. It also includes mechanisms for dealing with terminal as well as user mobility, which has promoted the use of SIP as a mobility management scheme. An important property of SIP is that control (signalling) and data traffic paths can be independent. Therefore, the triangular routing problem of mobile IP is not an issue. Unlike mobile IP, SIP does not use IP encapsulation which increases its communications efficiency. In addition, since SIP is an application layer solution it is independent of the network technology. SIP has also been adopted by the Third Generation Partnership Project (3GPP) for setting up real-time multimedia sessions. However, SIP sometimes results in large vertical handoff latencies due to heavy signalling message processing. In [BBD03] VHO latencies of up to six seconds have been reported. In [BDA05] a solution is proposed to provide SIP based soft handoff, which eliminates packet loss and outage in the face of high handoff latency.

2.3.4 WLANs and Cellular Network Integration Architectures

In an early work, five architectures were proposed for supporting WLAN-GPRS vertical handoff [PKH⁺00]. Two of the architectures integrate the WLAN as part of the GPRS network by connecting it to the GPRS core network. In the other approach, a logical entity called a virtual AP connected to the WLAN is used to interwork between GPRS and WLAN. Here, the WLAN views the GPRS network as another WLAN access point represented by the virtual AP. Mobility is managed by the WLAN according to IEEE 802.11 and the Inter-Access-Point Protocol (IAPP) specifications. The other alternative is a proxy/mobility gateway approach which introduces a mobility gateway as a connection anchor point between WLAN and GPRS networks. The mobility gateway acts as a proxy for all traffic originating or terminating at the MSs. The other communication party connects to the MS through this proxy and is unaware of the actual location of the MS. Mobile IP is presented as the last alternative.

The quoted studies do not address the issue of seamless vertical handover and the type of traffic considered is confined to non-real-time data. The distinction between handoff from GPRS to WLAN and from WLAN to GPRS is also recognized in this work. While GPRS to WLAN handoff can be dealt with in a timely fashion, WLAN to GPRS handover may occur with very little warning since WLAN coverage can experience severe degradation very quickly. A neural network handoff triggering approach is employed to provide faster handoff initiation.

In an effort to standardize WLAN-cellular network integration models, The European Telecommunications Standards Institute (ETSI) has specified two approaches namely, loose coupling and tight coupling [ETS01]. In the tight coupling architecture a WLAN coverage area is viewed by the cellular core network as an additional coverage cell. An interworking gateway is connected between the WLAN's infrastructure network and the cellular core network which makes the WLAN appear as a cellular base station (BS). In this model the WLAN is used as a simple "transport pipe" to carry native cellular signalling and services between the MS and the cellular

network. Tight coupling generally benefits from lower vertical handoff latencies since the handoff is internal to the core network and uses native cellular mobility management algorithms. This approach is currently being standardized and deployed under the UMA/3GPP-GAN activities [UMA]. Tight coupling is very cellular centric and allows the cellular operator to maintain control over services offered to the MSs.

In [ea04b], a seamless VHO scheme using pre-authentication and pre-registration was proposed for WLANs tightly coupled to a UMTS network. Pre-registration is a mobile IP-based fast handoff scheme that triggers MIP handoff before link layer handoff, thereby limiting packet loss and handoff delay to that caused by link layer handoff [ea01]. Moreover, by having the old AP buffer packets during handoff and forwarding them to the new AP after handoff completion, packet loss is reduced. Such a forwarding mechanism is reasonable since the APs involved are typically separated by a small number of hops, resulting in a relatively small forwarding latency. This is usually not the case in loosely coupled WLAN/cellular networks.

Loosely coupled architectures connect a WLAN coverage area to the cellular network through the cellular operator's PSTN and IP network connections, which usually consists of many hops. This type of arrangement is increasingly used in business/enterprise networks, and an example is shown in Figure 2.10. Loose coupling is less proprietary than tight coupling since each of the WLAN and cellular networks are independent IP wireless access domains. This approach is also potentially more scalable since the WLAN traffic does not have to go through the cellular core network. In fact, as shown in Figure 2.10, two independent call legs are used to connect the MS to the other end of the connection (CN). However, real-time vertical handoff may be challenging to achieve since a new connection has to be established to the cellular access network and this tends to be a time consuming process. This is due to the significant amount of signalling that is required for the VHO. In [Net00], for example, mean cellular call setup times for an EDGE Radio Access Network were found to be approximately 7.9 seconds. Much larger values were quoted in [NKK05], where numbers in the 28-31 second range were found for GSM call establishment. The authors of [STZK06] have collected data for link establishment latency by making GSM calls

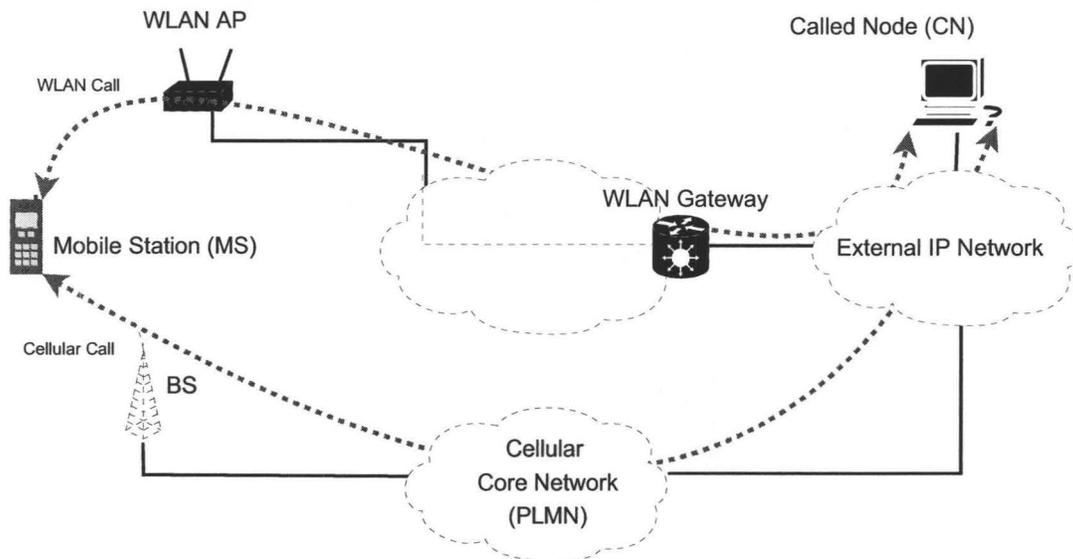


Figure 2.10: *Loosely Coupled WLAN/Cellular Integration Architecture*

every 15 minutes for two continuous weeks. This was done using an Option Globetrotter GPRS/GSM PC-Radio Card connected to the (Canadian) Rogers network and calling a PSTN destination in a lab at McMaster University. It was found that the call distribution can be approximated as a truncated Gaussian distribution with an average and standard deviation of 6.68 and 0.123 seconds, respectively. Interestingly, the standard deviation of GSM call latency is quite low for this particular cellular operator. The study could not however, obtain a breakdown of the various cellular/PSTN components of the latency since it is internal to the cellular core network and the PSTN.

The authors in [WBBD05] have analyzed the delay associated with vertical handoff using SIP in a loosely coupled WLAN-UMTS setup. Analytical results show that WLAN-to-UMTS VHO incurs unacceptable delay, (i.e., 3 to 4 seconds), for supporting real-time multimedia services, and is mainly due to the transmission of SIP signalling messages. On the other hand, UMTS-to-WLAN handoff experiences much less delay, mainly contributed by the processing delay of signalling messages at the WLAN gateways and servers. While the WLAN-to-UMTS case requires the deployment of soft handoff to reduce the delay, faster servers and more efficient host configuration

mechanisms are sufficient to perform a lossless handoff in the UMTS-to-WLAN case.

Another example of a loosely coupled architecture is the hybrid coupling with radio access system (HCRAS), which is a WLAN/UMTS internetworking architecture that has been recently proposed in [LZ05]. The system uses an IEEE 802.16 link installed between the UMTS network and the local WLAN. HCRAS can dynamically distribute traffic among the WLAN and UMTS networks and can significantly reduce signalling and handoff latency. This approach is based on Mobile IP Fast Handoff which establishes a bidirectional tunnel between the old and new foreign agents for data forwarding purposes during handoff. This approach is, however, not transparent to the WLAN and cellular networks.

Using MIPv6 has also been considered in [ea04c], which uses a loosely coupled GPRS-WLAN experimental testbed. The impact of VHO on TCP connections was considered and experiments showed a 3.8 sec VHO. By using fast router advertisements (RA), RA Caching, and binding update simulcasting, this can be reduced to about 1.36 sec which is still much higher than the 150 to 200 msec end-to-end delay which is required for real-time voice.

2.4 The Effect of WLAN Deployment on WLAN to Cellular VHO Performance

Signal coverage behaviour at the peripheries of a wireless access network may have a significant influence on VHO performance. If a coverage area has abrupt signal decrease at its boundaries, then performing seamless VHO may be difficult. In this situation, radio coverage can be lost without sufficient warning to initiate and complete a VHO.

Modeling signal propagation is a rich area of research. The most widely used propagation models assume exponential path loss with lognormal shadowing [Rap96]. According to this model, the signal strength decreases inversely proportional to a certain power of the distance between the transmitter and receiver. The local shadowing

component which contributes to the signal decay is modeled by a zero mean lognormally distributed random variable.

Traditionally, propagation models were developed for outdoor environments. However, a significant fraction of WLAN deployments cover indoor areas. In [Rap96] many empirical and semi-empirical models have been reviewed for predicting indoor signal propagation. These models usually describe the mean signal value and are based on the exponential path loss model with a path loss exponent which depends on the environment and other added loss parameters which model wall and floor effects. In [ZVP03] an indoor measurement campaign in the 2.45 GHz ISM frequency band has been described and used to find optimal parameters for two standard indoor propagation models. These references, however, do not consider outdoor signal propagation as a result of the indoor coverage. This information is very useful in predicting WLAN link status for a VHO call which is exiting an indoor coverage region.

The other factor that affects indoor to outdoor WLAN received signal strength is the characteristics of the WLAN deployment, which dictates the indoor coverage distribution due to the positioning of the access points. When the APs are deployed densely, the WLAN is said to be a *capacity deployment*, whereas locating APs sparsely results in a *coverage deployment*. Capacity deployments provide a higher RSSI across the deployment region, which enables users to operate at a high bitrate. On the other hand, a coverage deployment can cover a larger area with the same number of APs at the cost of lower average RSSI values and hence lower bitrates.

Reference [Hil01] describes approaches developed for the design of campus-wide WLAN networks. WLAN design schemes are proposed for capacity and coverage based deployments by using the proper combination of access point locations, frequency assignments, and receiver threshold settings. Reference [Pra00] also recognizes capacity and coverage deployments and discusses the relationship between receiver sensitivity, bit rate, and coverage range for Orinoco-based WLAN systems using well-known indoor propagation models. Capacity versus coverage deployment was also investigated in [SKK01] and it was found that as users move further from an IEEE 802.11 AP the throughput and hence capacity is reduced even before falling back to lower data rates.

In [KTZ04], the possibility of abrupt WLAN coverage loss was investigated and an explicit VHO trigger node was used. However, it was not shown how WLAN coverage changes as a user passes through the transition region.

A very recent study, based on the architecture proposed in [STZK06], reports on measurements taken from indoor-to-outdoor WLAN coverage loss and its effect on vertical handoff outage [SAT06]. It has been shown that even if the cellular interface is active at the time of VHO triggering, considerable call disruption will typically occur before the new call leg is established. The time frame for finishing the VHO can be worse in high capacity WLAN deployments where lower data rate use is restricted through various information element advertisements [SAT06]. In such deployments, user links are disconnected by the AP if they go below a certain data rate to preserve capacity. Therefore, a MS has less time to finish a VHO before losing its WLAN link.

2.5 Vertical Handoff Triggering Techniques

Vertical handoff triggering in hybrid wireless networks has been investigated in a comprehensive manner [PKH⁺00]. Many classical handoff decision algorithms based on signal strength have been summarized in conjunction with using thresholds, hysteresis, and dwell timers. A neural network approach to detect the proper moment for VHO initiation has also been proposed and shown to outperform other classical handoff decision algorithms.

Classical handover detection mechanisms are based on the received signal strength. However, one could imagine more sophisticated decision variables for triggering a VHO. In general, a handoff may be triggered if the performance of the link will be superior on a different wireless network. For example, a MS which has connections to both a cellular network and a WLAN may choose the WLAN for its inexpensive service. Such handover triggering schemes can use various parameters. In [BHJ03] a combined signal strength and distance criteria is used to trigger handover between a UMTS cell and a WLAN in a tightly coupled architecture. For the case of UMTS to WLAN handoff, only the distance and signal strength to WLAN AP are considered.

In the other direction, handoff is initiated if the signal strength to the WLAN AP is lower than some threshold and signal strength (and distance) to the UMTS BS is higher (smaller) than a given threshold. This policy prioritizes the use of WLAN over UMTS. The authors, however, do not comment on the choice of different thresholds used. It has been analytically shown that VHO probability is decreased thus signalling overhead due to excessive VHO on both networks is reduced. However, this approach leads to a more populated UMTS cell and a less populated WLAN compared to conventional RSSI based triggering. It is more desirable to relieve UMTS load by having users handoff to WLAN as soon as possible; this also gives users the opportunity to enjoy higher bit rates and less packet delay.

In [LSB⁺04], a location-aided measurement technique for handoff triggering has been proposed. This uses measurements collected by MSs communicating with the target wireless network to assist a MS with making a VHO decision. Given the current location of a MS, this data can be used to predict the quality of the connection (RSSI, etc.) between the MS and the target access network without having to perform actual measurements. In [BSL05] movement and location prediction is also used to estimate the time of VHO initiation.

In [LSB⁺04] a number of other handoff triggers are also presented which are categorized into physical and link layer based. Physical layer based triggers include signal strength, interference levels, and Bit Error Rate (BER). Link layer based triggers consist of QoS violation triggers, connection admission control triggers, location based triggers which take into account distance to various BSs and APs, and a priori knowledge-based triggers such as maps of the coverage area. A weighted combination of these triggers can be used to achieve a single VHO trigger.

A smart VHO decision model has been proposed in [LJCG04] to decide the best network interface for handoff. A score function is utilized to make a smart decision based on various factors, such as the properties of available network interfaces and user preferences. The score function calculates a score for every network interface based on a weighted sum of usage expense, link capacity, and power consumption. This approach, however, does not consider loading and signal strength at the target network. It also does not take into account VHO execution latency and the possibility

of coverage loss on the old network. A similar study has been done in [ZM04] where new optimizations for vertical handoff decision algorithms have been developed to maximize the benefit of the handoff. The optimizations incorporate a network elimination feature to reduce the delay and processing required in the evaluation of the cost function, and a multi-network optimization is introduced to improve throughput for mobile terminals with multiple active sessions. A performance analysis demonstrates significant gains in quality of service and a more efficient use of resources.

2.6 Summary

Real-time network applications such as voice-over-IP (VoIP) require soft vertical handoff. However, WLAN to cellular vertical handoffs experience outage due to the abruptness of WLAN link loss, the severity of which depends on the WLAN deployment. For this reason, the VHO triggering mechanism is very important so that a VHO is initiated as early as possible while preventing unacceptable false triggering. In the next chapter, a solution is proposed that overcomes the outage and packet loss occurring during a WLAN-to-cellular vertical handoff. The solution is applied to the loosely coupled architecture and uses SIP as the mobility management scheme.

Chapter 3

The Vertical Handoff Support Node

3.1 System Architecture

In a loosely coupled architecture such as the one illustrated in Figure 3.1, the WLAN and cellular (e.g., GSM) networks both connect through the PSTN (or possibly the external IP network). In this architecture the enterprise LAN is connected to the public switched telephone network (PSTN) through an IP-PBX or some other such PSTN gateway. The IP-PBX functions as an anchor node (AN) for voice connections as shown in the figure. In the example shown, the dual mode mobile station (MS) initiates a voice call with a PSTN called node, CN. This call consists of two concatenated real-time call legs. The first (i.e., Call Leg 1) runs from the MS to the AN through the enterprise IP network, typically using SIP/RTP/IP protocols. The rest of the call (i.e., Call Leg 2) is a PSTN call that runs from the AN to the CN. In this case the AN functions as a gateway between the IP and PSTN world. In some implementations the gateway could also place external calls through an IP network such as the Internet using VoIP protocols rather than the PSTN.

When the MS moves out of WLAN coverage, a vertical handoff is required to replace the WLAN call leg (1) with one through the cellular network (3). When a VHO occurs the connection path must be transferred from the MS's WLAN radio

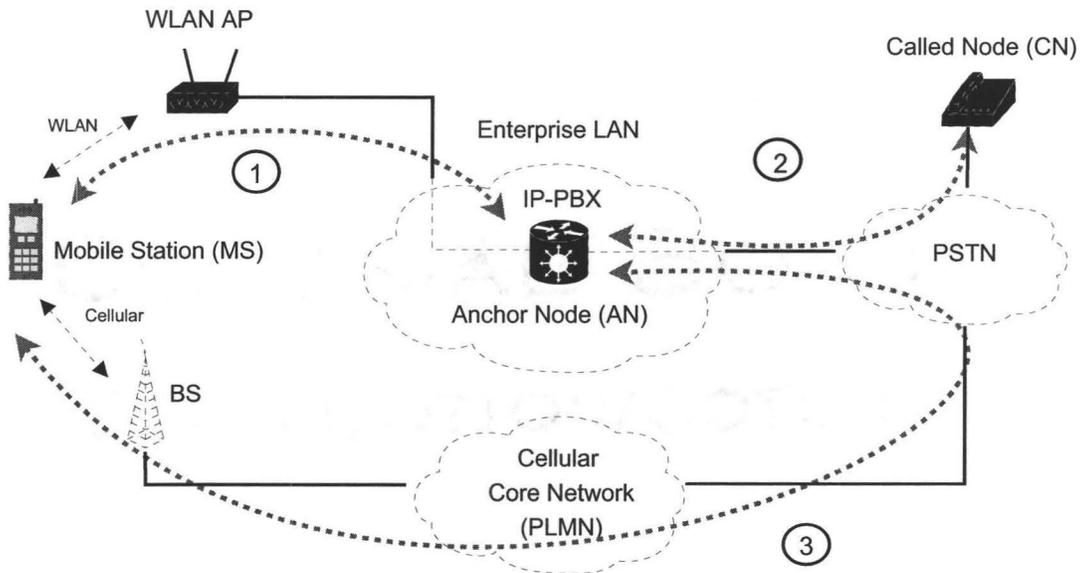


Figure 3.1: *Loosely Coupled WLAN/Cellular Integration with Enterprise Anchor Node*

interface to its cellular radio interface. The result is shown in Figure 3.1 as call Leg 3. This call leg is essentially a cellular call which runs from the MS to the AN passing through both the cellular core network and the PSTN. Once call leg 3 is established, call leg 1 can then be removed and the voice call now consists of the concatenation of call legs 2 and 3. If the MS moves back within WLAN coverage, a handoff can be initiated in the reverse direction. In this work we focus on WLAN-to-cellular VHO since handoff in the cellular-to-WLAN case is typically much easier due to the ubiquity of cellular coverage. Note that the original call leg (2) to the CN remains fixed and handoff is transparent to the CN.

Soft handoff occurs when call leg 3 is established while call leg 1 is still operational. Therefore, when call leg 1 terminates, the media stream is already flowing through the MS, leading to a seamless handoff [STZK06]. In current IP-PBX's, soft handoff can be accomplished by using an AN with voice conferencing capabilities. In this case the original call is established through a "conference room" at the AN [STZK06]. When the vertical handoff is required, call leg 3 is generated by the MS establishing a new call to the conference room. When this has been established the first call leg can be dropped. The reader is referred to [STZK06] for a detailed description of the

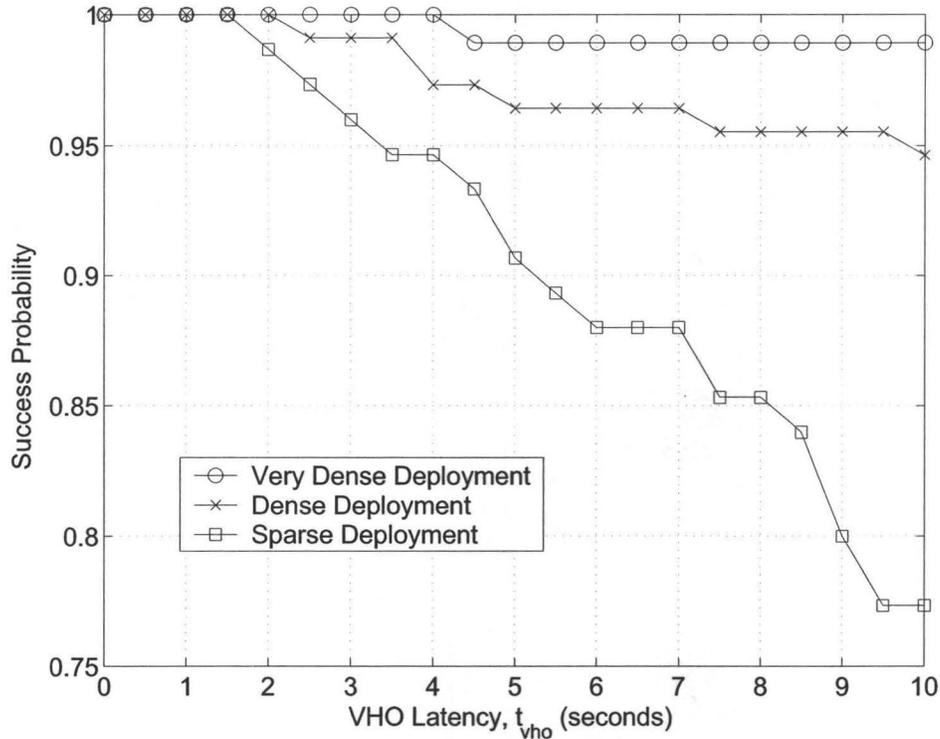


Figure 3.2: Vertical Handoff Success Probability as a Function of its Latency, for Some Typical WLAN Deployments [SAT06].

signalling required for this type of handoff.

It should be noted that we adopt this architecture and the anchor-node based soft handoff approach without tying it to any specific cellular technology or vertical handoff signalling. Essentially, we require that a soft handoff mechanism which works for a loosely coupled architecture be present. Although uninterrupted soft vertical handoff is possible, in practice the soft handoff can fail if WLAN signal degradation happens too quickly. In many common situations such as when exiting a building, WLAN coverage loss can occur with very little warning (e.g., $< 2sec$ [SAT06]). Hence, the soft VHO procedure will experience outage and result in packet loss if a loss of WLAN coverage occurs earlier than the time needed to properly trigger and execute the vertical handoff, i.e., establish the cellular leg of the call to the anchor-node.

A study performed in [SAT06] reports on measurements taken of indoor-to-outdoor WLAN coverage loss and its effect on vertical handoff success rate. Even if the cellular interface is active at the time, considerable voice call disruption will typically occur before the new call leg can be established. Figure 3.2 taken from the results in [SAT06] shows VHO success probability which is complementary to VHO dropping rate for typical WLAN deployment scenarios. The very dense deployment scenario corresponds to one in which a minimum signal strength of -55dBm is maintained throughout the deployment region. The dense and sparse deployments correspond to a minimum signal strength of -65dBm and -75dBm, respectively. A VHO event is called successful if it completes without the WLAN link being lost. In [SAT06], a VHO is initiated using conventional RSSI threshold based triggering. The figure illustrates how VHO success probability decreases as VHO latency increases. Since VHO latencies in loosely coupled settings tend to be large, one can see that providing a high success probability close to what is expected in local cellular handoffs (e.g., > 99.9%) is almost impossible.

The work done in [STZK06] also reports on data collected for cellular link establishment latency, which is effectively the VHO latency, by making GSM calls every 15 minutes for two continuous weeks. Their results are shown in Figure 3.3 and do not include the initial base station attachment delay. Matching the average vertical handoff latency of 6.6 seconds from their results with Figure 3.2, one can only expect a VHO success rate of up to 98.5% meaning that at best 1.5% of the VHO attempts will end up experiencing outage and packet loss.

3.2 The VHSN Solution

We now propose a mechanism for mitigating the VHO link loss problem using a fast recovery technique. Figure 3.4 shows a modified version of Figure 2.10 with a Vertical Handoff Support Node (VHSN) added locally to the wired LAN infrastructure. The VHSN contains both a wired (Ethernet) LAN interface and a cellular air interface. The VHSN does not extend any wireless coverage as in multi-hop forwarding mechanisms [HWT⁺05] [WCM00] [QW00], and does not communicate directly with any

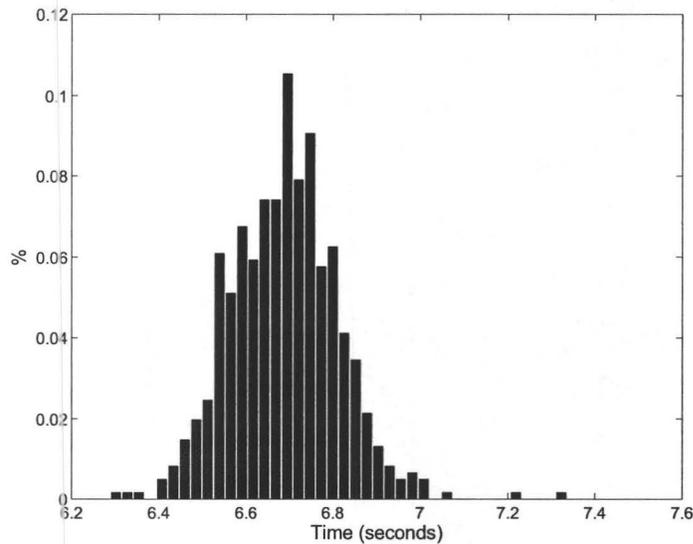


Figure 3.3: *GSM Call Establishment Time Results*

mobile station. When it is installed, the VHSN connects to the local cellular base station (the BS in Figure 3.4) as a conventional client mobile station. This is an important distinction since it operates as a simple Ethernet LAN and cellular client end station.

When a VHO is triggered by the mobile station, it will send a VHO notification packet (HO_SIG) to the VHSN through its WLAN or cellular interface. Upon receiving this notification from the MS, the VHSN obtains access to all incoming packets destined to the MS. This can be done by having the VHSN issue a gratuitous ARP in a manner similar to that performed by a home agent in standard mobile IP. The VHSN then quickly redirects the connection through the local cellular base station (BS) which then forwards it to the MS which is already attached to the cellular network. This redirection path is shown in Figure 3.4 as “(R)”.

In this figure, the initial voice call is labelled “(1)” which passes from the MS, through the enterprise WLAN to the anchor node (AN) and out to the CN via leg 2. While the third leg of the call is being established, the VHSN intercepts the incoming media stream on the enterprise LAN and redirects it over the recovery path through

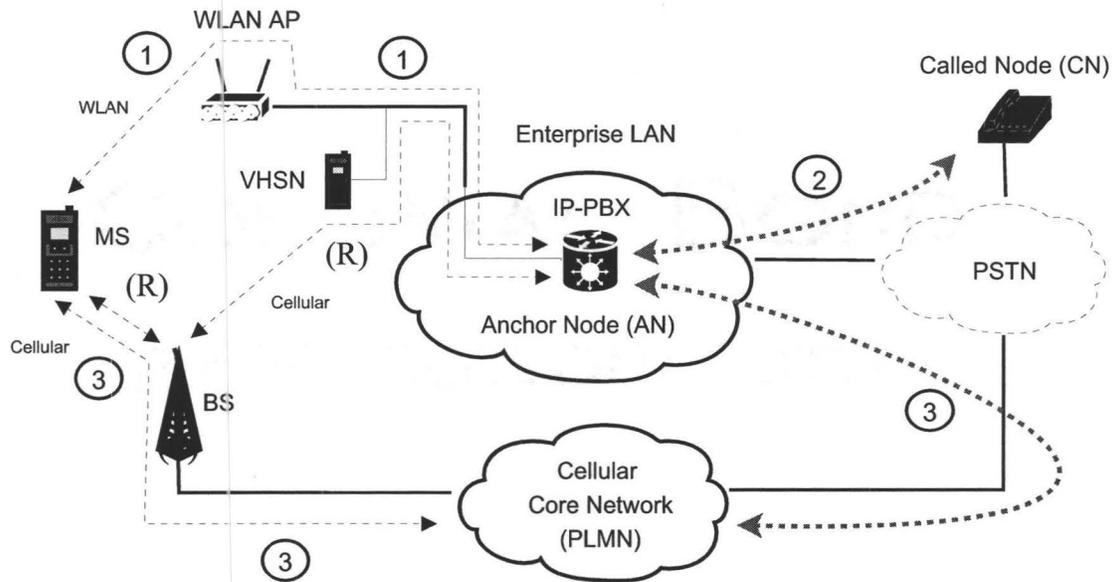


Figure 3.4: *VHSN Fast Link Recovery Mechanism*

the cellular base station (BS) using its cellular interface. The BS forwards this media stream to the MS, which is also a cellular client station. Note that the forwarding procedure through the cellular BS is only possible using a packet oriented cellular air interface. Now that the recovery media path is being established the MS can initiate a full soft handoff by creating the third leg of the call.

Once the soft handoff has been executed, the path labelled “(3)” is used and the recovery path is no longer needed. At this point the new cellular call leg is available from the MS, through the BS and cellular core network and back to the AN through the PSTN as shown by call leg 3 in Figure 3.4. It is worth mentioning that packets originating from the MS would also be sent through the redirection link to the VHSN and from there delivered to their destination until the MS establishes the cellular leg (i.e., leg 3) of the connection to the anchor node (AN).

We assume that a mobile station always attaches to the cellular network upon power up, even if WLAN connectivity is being used. The MS will then obtain information about its serving VHSN when it first associates to the WLAN. This information includes what is needed to make cellular contact with the VHSN. Figure 3.5 illustrates the VHO process which is also described below in more detail, assuming

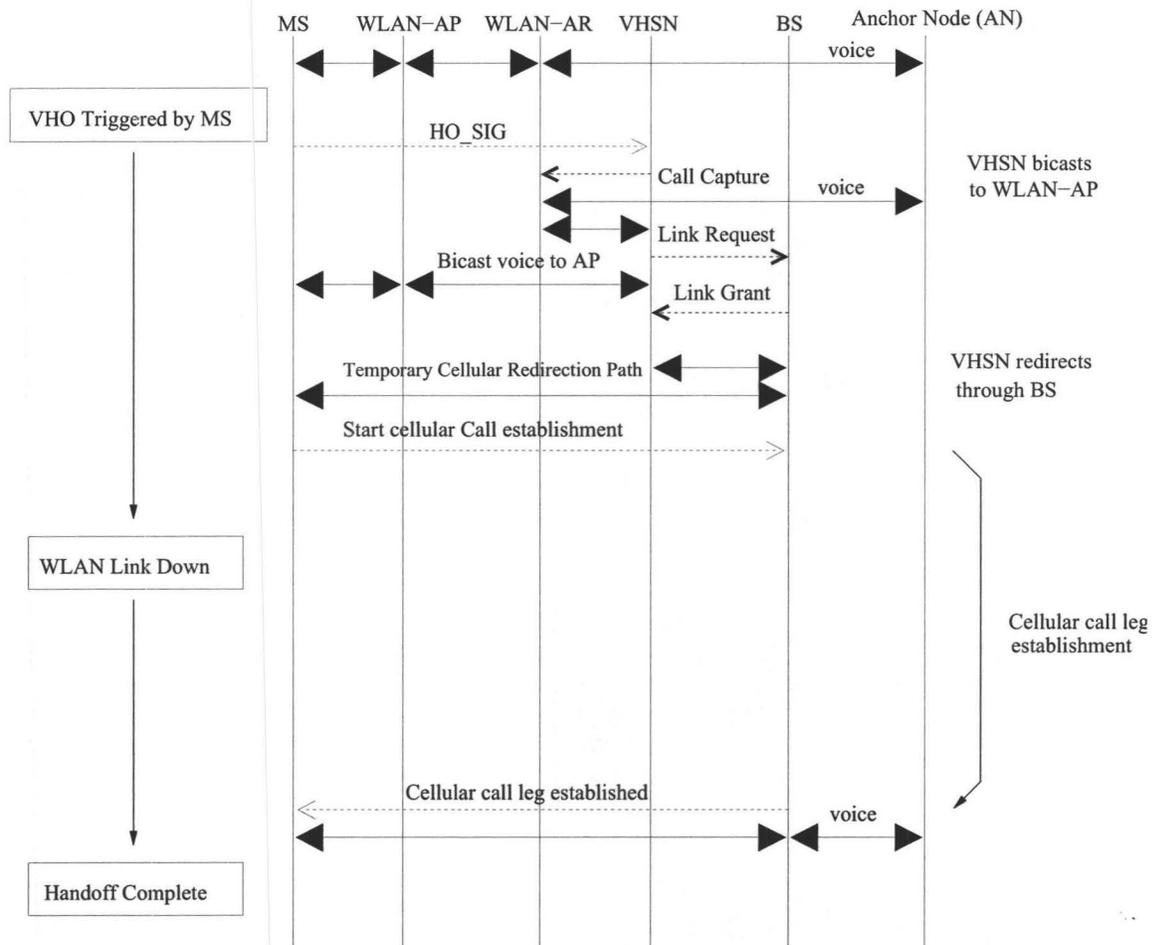


Figure 3.5: VHSN Handoff Timing.

that the vertical handoff is MS-initiated:

1. The MS's RSSI for the WLAN AP link decreases until it determines that a handoff is required. In most systems this is triggered when the RSSI reaches a predetermined threshold, S_{tp} , however more sophisticated schemes are possible. At this point the MS must determine if there is a suitable WLAN AP candidate to which it can perform a conventional WLAN-to-WLAN handoff. Once this phase is exhausted the MS triggers the initiation of a vertical handoff. The MS sends a handoff signalling packet, HO_SIG, to the VHSN through its WLAN or cellular connection which prepares the VHSN for connection redirection.

2. At this point there are two processes running, i.e., (a) The VHSN is receiving packets destined to the MS through the local WLAN-AR which is the access router responsible for routing packets to the WLAN (not shown in Figure 3.4) and is bi-casting them to the MS through both the cellular BS and the WLAN AP, and, (b) The MS has initiated a cellular connection to the AN (i.e., leg 3). The delay incurred before the MS starts receiving packets from the VHSN through its cellular interface can be calculated as,

$$T_{initial} = T_{connect} + T_{mac}^u + T^u + T_{mac}^d + T^d, \quad (3.1)$$

which is the summation of the time it takes for the VHSN to, access the cellular radio, T_{mac}^u , send the first uplink packets towards the BS, T^u , and for the BS to access the downlink radio, T_{mac}^d , and send the first downlink packets to the MS, T^d , in addition to the redirection link establishment time, $T_{connect}$. Once a link is set up between the VHSN and the MS, the additional packet delay between them will be $T^u + T^d$ (i.e., uplink and downlink latencies).

3. The third leg of the call is created and the soft handoff mechanism described in Section 3.1 (i.e., using the anchor-node) will now disconnect the WLAN leg of the call. However, packets previously sent along the WLAN path will still be arriving at the WLAN which will be collected and forwarded by the VHSN. When all packets destined to the WLAN have been delivered, the VHSN functions are completed.

During the recovery period, additional resources are being used in the cellular system compared with what would normally be required for a call. However, these are only needed for the limited time required to generate the new cellular call leg (i.e., VHO latency). Since the cellular base station is local and only radio resources in the local cell are needed, this mechanism has the potential for re-establishing the redirection connection very quickly before the rest of the vertical handoff occurs.

Unlike previous ad hoc or infrastructure-based relaying approaches [HWT⁺05] [WCM00] [QW00], the proposed mechanism does not use any frequency resources that have not already been deployed by the cellular network, and does not require

any infrastructure changes. A single VHSN can also service an entire WLAN subnet, subject to capacity constraints and air interface capabilities. This scalability is discussed later. A requirement for the VHSN is that the WLAN hotspot be entirely contained within the cellular base station's coverage area, which is typically the case.

From a practical point of view the VHSN can be deployed by either the cellular network operator or the WLAN owner since the proposed mechanism does not necessarily require cooperation between the two. Also, since the VHSN is nothing more than a cellular client node with an Ethernet connection, it could be made very inexpensively. A strength of the scheme is that it can easily be integrated into a suitable cellular network, which is particularly advantageous in loosely coupled systems where outages are expected to be common.

The VHSN must maintain a continuous association with the cellular base station. The VHSN connection cost could be an issue if it is not owned by the cellular operator and is subject to continuous air-time charges. It is not expected that this would be the case since when idle, the VHSN would remain attached in the same way that existing GPRS/GSM clients do so. This is typical even in existing 2G networks, where for example, a GPRS/GSM PDP context is maintained despite the use of a second air interface [UMA]. In this case additional charging only happens when traffic is passed through the BS.

An important assumed requirement for supporting the VHSN is the ability for the cellular base station to permit local station-to-station packet redirection paths as required for the MS to VHSN connection. Although this functionality would be very straightforward in 4G air interfaces, it may not always be provided, and when this is the case the VHSN functionality may not be feasible. The results in this chapter can be viewed as representing the improvements possible when this capability is provided.

3.3 VHSN Performance Analysis

In this section we discuss the performance obtained using VHSN. We consider packet loss performance, VHSN overhead and scalability, and the call blocking/dropping behaviour of the cellular system compared with conventional soft handoff. The results

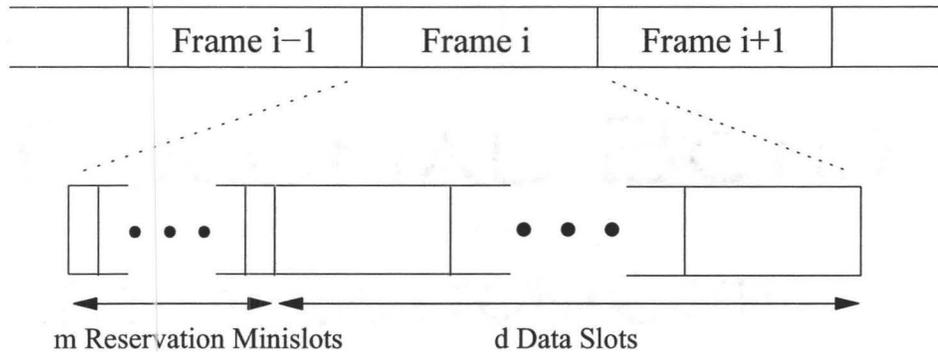


Figure 3.6: Reservation Based TDMA Radio MAC Layer

obtained are compared with simulation results in Section 3.4.

A demand-assignment TDMA air interface is assumed for the cellular network as would be the case in a 4G cellular network. Figure 3.6 shows an example of the type of access scheme assumed. The timeline consists of recurring frames that are collected into a superframe (not shown). Each frame consists of a set of m reservation minislots followed by d data slots. A slotted-ALOHA based access mechanism is employed in which mobile stations make resource requests from the BS during the reservation minislots. An unsuccessful request must be retransmitted in a future contention slot after a random number of access minislots. Upon receiving a resource request the BS assigns available time-slots to the requesting MS and sends this allocation in a timeslot map in the next frame (not shown in the figure).

For real-time packet flows, an extended resource allocation scheme can be used which performs reservation along multiple superframe periods for as long as the packet flow exists. Table 3.1 is the list of parameters taken from various sources and used in our analysis. For a particular VHO instance we also define T_{wlan}^{fail} to be the average elapsed time from VHO initiation until WLAN coverage fails.

Now consider the conventional non-VHSN handoff case. When an MS is moving out of WLAN coverage, it eventually detects a decrease in RSSI, and determines if a WLAN-to-WLAN handoff is possible. There are various mechanisms for reducing WLAN scanning latency which would streamline this stage, such as proactive scanning and using AP neighbour lists [FT05]. Once a VHO is triggered, the MS uses its cellular

Table 3.1: Vertical Handoff Delay Parameters

T_{frame}	Cellular frame duration, (4.845 msec default [JG97])
T_{mac}^u	Cellular radio channel access latency, (two or four frame times)
T_{mac}^d	Cellular base station radio channel access latency, (two frame times)
T^u, T^d	Up/Downlink packet transfer delay
T_{vho}	Latency for establishing the cellular leg of the call (VHO latency).
T_{core}	Mean cellular core network delay, (20msec)
T_{net}	Mean Internet transit delay, (50msec [Goo99])
$T_{connect}$	VHSN to MS redirection-link establishment time, (500msec)
T_{wlan}^{fail}	Elapsed time from VHO initiation to WLAN coverage failure (2.0sec)

interface to establish a connection through the cellular core network to the anchor node. This connection will serve as the third leg of the call as labelled in Figure 3.4 with the number 3.

We define the VHO delay, T_{vho} , to be the time it takes for the third leg of the call (i.e., Leg 3 in Figure 3.4) to be established. It is clear from the literature that call setup times in existing cellular networks are often long and highly network specific. In the ensuing results we will use the collected data previously described in Figure 3.3 for T_{vho} , but it should be noted that in some cases smaller or larger latencies can be experienced. The numbers quoted represent current network latency examples which could eventually change, but this is not expected to happen very quickly since cellular core networks tend to evolve very slowly. It should be noted however, that if call connection latencies were to become almost instant, there would be no role for the VHSN.

The outage period during which the MS is completely disconnected is given by $T_{vho} - T_{wlan}^{fail}$. During this time, packets arriving for the MS through the WLAN interface will be lost. In a separate study T_{wlan}^{fail} for different WLAN deployments [SAT06] has been characterized. The results suggest that T_{wlan}^{fail} can easily vary over a large range of values with typical averages of 2 to 4 seconds. For our purpose, T_{wlan}^{fail} is chosen to be a parameter.

3.3.1 Packet Loss Performance

The initial delay given in Equation 3.1 can be written as,

$$T_{initial} = 2\eta^{data}T_{frame} + 2T_c + T_{connect}, \quad (3.2)$$

where η^{data} is the number of interactions required between the MS and BS before it is granted access to data slots and relates the cellular radio access delay component for acquiring timeslots to the frame duration. Also, T_c , is the cellular link transmission delay. To estimate the number of lost packets during VHO we assume a quasi-deterministic packet arrival at the WLAN which is reasonable for real-time CBR voice. At the time when the first data packet sent by the VHSN is received by the MS, there should be nearly an average of $\lambda T_{initial}$ packets in the VHSN destined to the MS, where λ is the packet arrival rate for the flow. Therefore, after the initial VHSN link setup phase the remaining time for the packet at the head of the queue to reach the MS within its deadline is $D - T_{net} - T_{initial}$, where D is the maximum allowable delay for a given packet and T_{net} is the CN to VHSN delay. Note that the time for this packet to get from the VHSN to the BS and from there to the MS is $2T_c$ and has been included in the computation of $T_{initial}$. The condition under which the first packet will be considered lost is given by $T_{initial} > D - T_{net}$.

In our case, $T_c = 1/\lambda$, since the packets are given periodic timeslots to be sent by the VHSN during each service interval. Therefore, the remaining time to deadline for the i^{th} packet at the moment when it reaches the head of VHSN's packet queue can be written as,

$$D - T_{net} - (T_{initial} - T_{arrival}^i) + T_c; \quad (i \geq 1), \quad (3.3)$$

where $T_{arrival}^i = (i - 1)/\lambda$ is the expected arrival time of the i^{th} packet, starting from when VHSN begins to buffer packets for the MS. Noting that it takes $2T_c = 2/\lambda$ to transmit a packet from the VHSN to the MS, for the $(i + 1)^{th}$ packet to be lost we should have,

$$D - T_{net} - T_{initial} + (i - 1)/\lambda < 0. \quad (3.4)$$

Hence the number of packets lost during VHSN vertical handoff can be estimated by,

$$L = \max\{0, \lceil \lambda(T_{net} + T_{initial} - D) \rceil\}, \quad (3.5)$$

where $T_{initial}$ is given by Equation 3.2. Equation 3.5 expresses the maximum number of packets lost if WLAN coverage is lost immediately after initiating a VHO. In most cases however, this is not the case and it takes T_{wlan}^{fail} for the mobile station WLAN link to be lost after VHO triggering. As a result, λT_{wlan}^{fail} packets can be saved by being received through the WLAN interface of the MS. Therefore, the actual number of packets lost can be estimated by,

$$L^* = \max\{0, \lceil \lambda(T_{net} + T_{initial} - D - T_{wlan}^{fail}) \rceil \}. \quad (3.6)$$

Using Equations 3.2 and 3.6 one can find the maximum time that a VHSN-MS redirection link establishment can take before experiencing a packet loss (i.e., $T_{connect}^{max}$). Our simulation results have shown that queueing effects in the nodes are negligible compared with various latencies and other sources of delay that have already been discussed. Our simulation results presented in the next section verify this assumption.

3.3.2 VHSN Overhead and Scalability

Using the VHSN will incur additional resource usage compared with the non-VHSN case. Assuming that each connection requires B_f^{wlan} bps while on the WLAN, and the resource usage for VHO signalling is S bits per VHO, then the total resource usage during one VHO is $2B_f^{wlan}T_{vho} + S$ bits. Usually WLAN voice connections use higher bit-rates and better codec quality than when operating through the cellular network. If WLAN voice encoding rates are then used during VHSN connection recovery the incremental bandwidth required during a VHO may be higher than that for native cellular vocoding rates, i.e., $B_f^{cellular} = (1/K_{BW})B_f^{wlan}$ bps. As a worst case we assume that WLAN vocoding rate is used. Accordingly the total bandwidth requirement for VHSN-assisted vertical handoff for a given VHO rate, λ_{vho} , is given by,

$$BW_{total}^{vhsn} = (2B_f^{wlan}T_{vho} + S)\lambda_{vho}(bps). \quad (3.7)$$

As an example, we consider a WLAN deployment with an average number of 15 active users, each performing a vertical handoff with probability 0.1%. As a result the average VHO arrival rate from such a WLAN is $\lambda_{vho} = 0.015/sec$. We take the

Table 3.2: Percentage of Additional VHSN Bandwidth Overhead Normalized to Cellular Call Bandwidth

	GSM 6.10	G.726-32	G.711
Bit rate(Kbps)	13.2	32	64
K_{BW}	1	2.42	4.85
BW_{total}^{VHSN} (normalized)	18%	43.56%	87.3%

VHO latency to be $T_{vho} = 6$ sec, and the cellular network is assumed to support 80 GSM6.10 calls. Table 3.2 lists the percentage of average additional bandwidth consumed when using the VHSN scheme normalized with respect to a cellular call bandwidth, for three different WLAN voice coding examples. Note that the values listed are total overheads normalized with respect to a single voice connection, e.g., a value of 100% consists of the bandwidth required for one additional voice connection per BS. This table also illustrates the increase in bandwidth overhead as the gap between the WLAN and cellular flow bit rate increases. For example, if both WLAN and cellular flows use the same bandwidth, then providing VHSN handoff for such a WLAN hotspot would only cost on average 18% of the bandwidth that is consumed by an ordinary cellular call. Even if G.711 rates are used, the average overhead is less than that (i.e., 87.3%) of a single additional voice connection. This result is typical of other examples we have done and in general we find that the bandwidth overhead for VHSN usage is very low since cellular call setup latencies, i.e., T_{vho} , are typically 2-3 orders of magnitude smaller than typical call durations.

We have also considered the scalability of the VHSN approach. To express the effect of policy-based resource limitations, we use a simple analytic model. We assume that VHOs arrive according to a Poisson process with mean rate, λ_{vho} , and that the cellular network at most allows a fixed number of C channels to be used by the VHSN. Assuming a VHO latency of T_{vho} , we can approximate the probability of VHO rejection at the VHSN by the probability of more than C VHO arrivals during an arbitrary period of length, T_{vho} . When the probability of rejection is low, it can

Table 3.3: VHSN Scalability, Max λ_{VHO} (1/sec) at 0.1% Rejection Probability

Max Channels, C	Max λ_{VHO}
1	0.02
2	0.05
4	0.2
6	0.4
8	0.62
10	0.88

be shown that this approximation is very good and is given by,

$$P_{reject} = 1 - \sum_{k=0}^C \frac{e^{-\lambda_{vho}T_{vho}} (\lambda_{vho}T_{vho})^k}{k!}. \quad (3.8)$$

The model can also easily be extended to the case where a VHO requires more than one cellular channel from the VHSN, for example, if different coding schemes are employed at the WLAN and cellular network. The probability of rejection can be further reduced by taking advantage of the time it takes for a MS to lose the WLAN link after initiating a VHO, T_{wlan}^{fail} . During this time, a MS can continue communication on the WLAN interface and wait for the VHSN to grant its request for a redirection link. Table 3.3 shows the values of maximum mean VHO arrival rate that can be supported assuming a 0.1% rejection probability. Using this approach we can see that up to 0.2 VHO's per second can be handled provided up to 4 channels are made available to the VHSN. This is a very high vertical handoff rate for a typical WLAN deployment. Note that VHOs are rare events and their arrival rate depends on the user residence time inside the entire WLAN coverage region. For example, if there are 1000 users in a WLAN out of which 20% have an active call, and the mean residence time is one hour, then the VHO rate will be roughly 0.05 per second.

3.3.3 Call Blocking and Dropping Performance

Due to the additional bandwidth requested by vertical handoff calls from the cellular network, the BS responsible for accepting these calls is likely to incur a higher

dropping probability for VHO calls compared to when implementing conventional soft handoff without using a VHSN. On the other hand the VHSN-assisted scheme provides for less outage and packet loss, which in turn decreases VHO dropping. To study the blocking and dropping probability for the VHSN system, we model a cellular base station using a Markov process with vertical handoff and new call request arrivals obeying a Poisson arrival process. We also assume call duration and handoff latency to be exponentially distributed random variables. In addition, there is no call/VHO queueing and each new call request requires one channel from the cell whereas a VHO call initially requests more channels depending on the value for K_{BW} . This model and these assumptions are validated in the next section using computer simulations.

Figure 3.7 shows the two dimensional Markov process modeling the channel occupancy of a BS with a maximum of C channels¹. For the sake of simplicity we show the case where C is an even number and $K_{BW} = 1$, that is, both WLAN and cellular calls use the same voice vocoding. We assume that handoff and new calls are not queued and are rejected immediately if no resources are available. Each state is represented by (h, c) , which corresponds to the number of channels currently occupied by handoffs, and the number of channels occupied by conventional cellular calls. The Markov process can be numerically solved in order to find blocking and dropping probabilities. These probabilities are given by,

$$P_{block} = \sum_{i=0}^{C/2} Pr\{(h, c) = (2i, C - 2i)\}, \quad (3.9)$$

and

$$P_{drop} = P_{block} + \sum_{i=0}^{C/2} Pr\{(h, c) = (2i, C - 2i - 1)\}, \quad (3.10)$$

respectively. Note that the value of both summations in Equations 3.9 and 3.10 depend on the value of K_{BW} , which in this case is set to 1. The blocking and dropping probabilities for a system using conventional non-VHSN soft handoff are equal.

¹We ignore the possibility of calls ending during the handoff process.

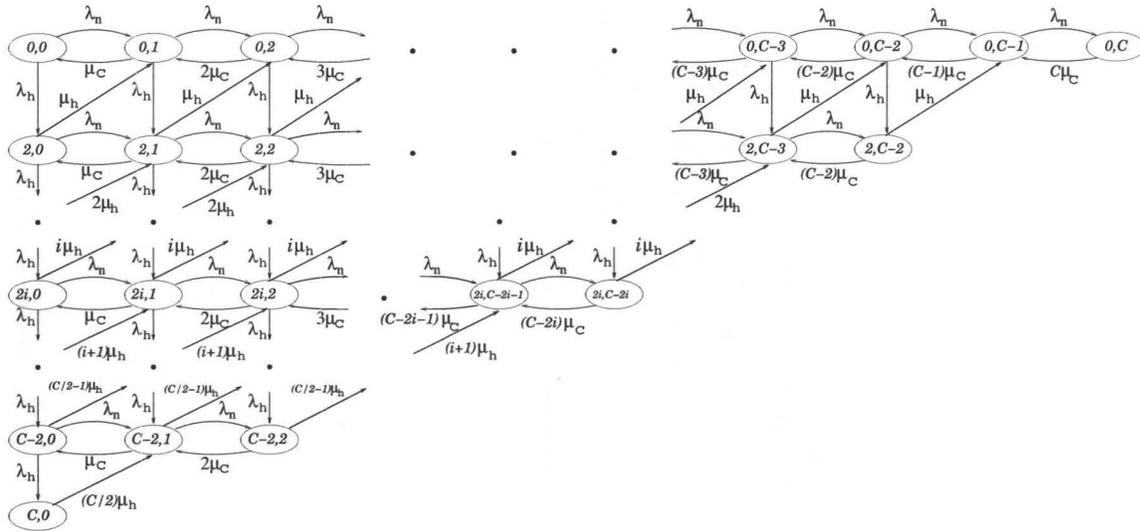


Figure 3.7: Markov Process for Cellular Base Station Channel Occupancy (VHSN Case).

3.4 Simulation Results

In addition to the analytic results simulations are performed for conventional and VHSN soft vertical handoff using a software developed in C++. The simulator consists of a standard event scheduling and execution engine which works with the different components of the system being modeled. Figure 3.8 shows a high level block diagram of the system model used for simulation. This diagram is closely related to Figure 3.4 and models all the major components in the system such as the mobile-station, cellular base station, anchor-node, called-node (CN), WLAN access point, and the VHSN as separate objects with well defined interfaces. The cellular medium is also modeled as a separate object containing the information for the entire cellular super-frame and individual timeslots, e.g., their position in the super-frame and whether they are occupied or free. The LAN to which the WLAN is connected is also modeled as a separate object, WLAN-Backbone, which is nothing but a passive entity serving as a connection point between WLAN-AP and VHSN.

Also shown, are the important queues in each of these components which store packets received for their corresponding entity through one of its interfaces or packets which are generated or forwarded by that entity to another. For instance, all the

queues for MS, BS, and VHSN, which have cellular interfaces are connected to the Cellular Medium object which arbitrates their access to the air interface.

The Mobile Station object is able to send/receive packets to/from two objects, WLAN-Backbone and Cellular Medium. The decision to choose one or both of these objects for communication is made by the VHO Protocol Control object which takes care of progressing through and tracking various VHO stages. The same object is also responsible for controlling the queues of the VHSN in order for the VHO protocol to be properly realized. For example, when the MS triggers a VHO and issues a HO_SIG packet to VHSN stating that it is about to perform vertical handoff, the VHSN switches its queues to WLAN-Backbone. This simulates the effect of VHSN collecting packets that arrive on the WLAN for the access point (WLAN-AP). In addition, when the temporary cellular redirection path between the VHSN and MS is established, the VHO Protocol Control entity available in the VHSN will switch its queues to both WLAN-Backbone and Cellular Medium. The VHO Protocol Control entity present in Anchor-Node is different from other instances of this class. This object deals with the VHO stage from the Anchor-Node's perspective. More specifically, it only keeps track of the condition of the three different legs of the connection as introduced previously in Figure 3.4. This object does not need to know the exact stage of the VHO protocol.

As our simulation scenario, we consider voice connections using the GSM 6.10 13.2 Kbps codec which is typically used in GSM cellular networks. In addition, we assume a cellular network having a TDMA radio interface as described before, containing 4×21 timeslots per superframe, 20 of which are for data. Each data timeslot contains 256 payload bits resulting in a total cell throughput of roughly 1Mbps. Table 3.4 shows the values for the set of parameters used in the equations provided in Section 3.3 and this is the complete set of parameters used in our simulations. The real-time packet arrival rate is taken to be one of the typical values used for the packetization period in the G.711 voice codec. On the WLAN side we assume the IEEE 802.11e QoS is used and that VHSN signalling is sent at the highest priority level. We also assume that 32 Kbps G.726-32 vocoding is used for calls in the WLAN. In addition, we have calculated 95% confidence intervals for our results by running the simulator more than 40 times for each set of input parameters.

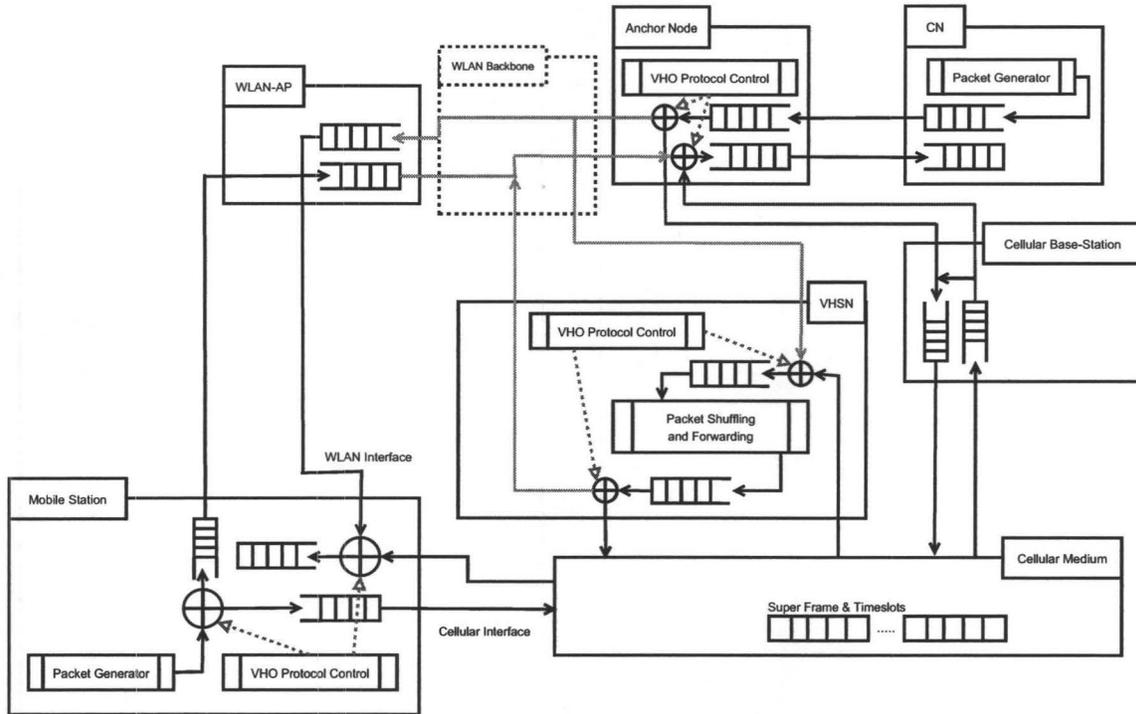


Figure 3.8: *Diagram Showing the System Model Used for Simulation.*

We evaluated the performance of the cellular BS in the presence of WLAN-to-cellular vertical handoffs. These results are shown in Figure 3.9 for a fixed cellular new call arrival rate of $0.250/sec$. Figure 3.9 compares two sets of curves corresponding to VHSN and non-VHSN vertical handoff performance. Each set consists of three curves showing cellular new call blocking probability (with a 95% confidence interval of 4.0% and 7.7% error around the mean for the VHSN and non-VHSN cases, respectively), vertical handoff dropping probability (2.4%, 7.0%) and base station utilization (0.2%, 0.2%) for a range of λ_{vho} between 0.01 and 0.20 arrivals per second. These curves show a wide range of operating conditions from a BS utilization of nearly 60% to a cell with about 90% utilization. The figure characterizes how a cellular base station would respond to the VHSN scheme in terms of new call blocking probability and how this affects the VHO dropping probability. For both VHSN and non-VHSN schemes the BS utilization is nearly the same as can be seen from the figure. Hence, the actions of the VHSN does not cause any significant utilization change for the BS.

Table 3.4: Parameter Values

Parameter	Value
η^{data}	4.5
N (Number of data time-slots in one superframe)	80
Superframe length	4 frames
$T_{timeslot}$	230.7 μ sec
Bits per timeslot	256
T_{frame}	4.845msec [JG97]
C (Maximum channels in one cell)	80
T_{net}	50msec [Goo99]
T_{vho}	6.0sec [STZK06]
D (Max allowable delay for real-time packets)	200msec
λ_{new} (Background cellular new call arrival rate)	0.250/sec
λ_{vho} (WLAN-to-Cellular vertical handoff arrival rate)	0.005 - 0.200 /sec
λ (Real-time packet arrival rate)	0.050/sec
T_c^{data} (Real-time packet service interval)	19.38msec
B_f^{wlan}	32Kbps (G.726-32)
$B_f^{cellular}$	13.2Kbps (GSM 6.10)
K_{BW}	2.42
Mean cellular call duration	180sec
$T_{connect}$	500msec
T_{wlan}^{fail}	Mean=2.0sec

In a loosely coupled architecture, VHO requests are not distinguished from cellular calls, resulting in equal dropping and new call blocking probabilities. The minor difference as seen in Figure 3.9 is due to the slightly higher bandwidth requirement for VHO requests. For the VHSN case both VHO and new call requests are dealt with in the same way by the cellular network, however, VHO requests require an additional link. According to Figure 3.9, at a low VHO arrival rate of 0.06/sec, the dropping rate for the non-VHSN and VHSN cases are 0.1% and 3% respectively. This difference becomes less as λ_{vho} increases resulting in higher load on the BS. For example, at a moderate VHO arrival rate of 0.10/sec one would experience a 0.7% dropping rate for the non-VHSN case compared to 10% for the VHSN approach. Although using VHSN increases the VHO dropping rate due to a lack of channels at the BS, if the VHSN is not used, VHO calls are far more likely to lose their link during the outage

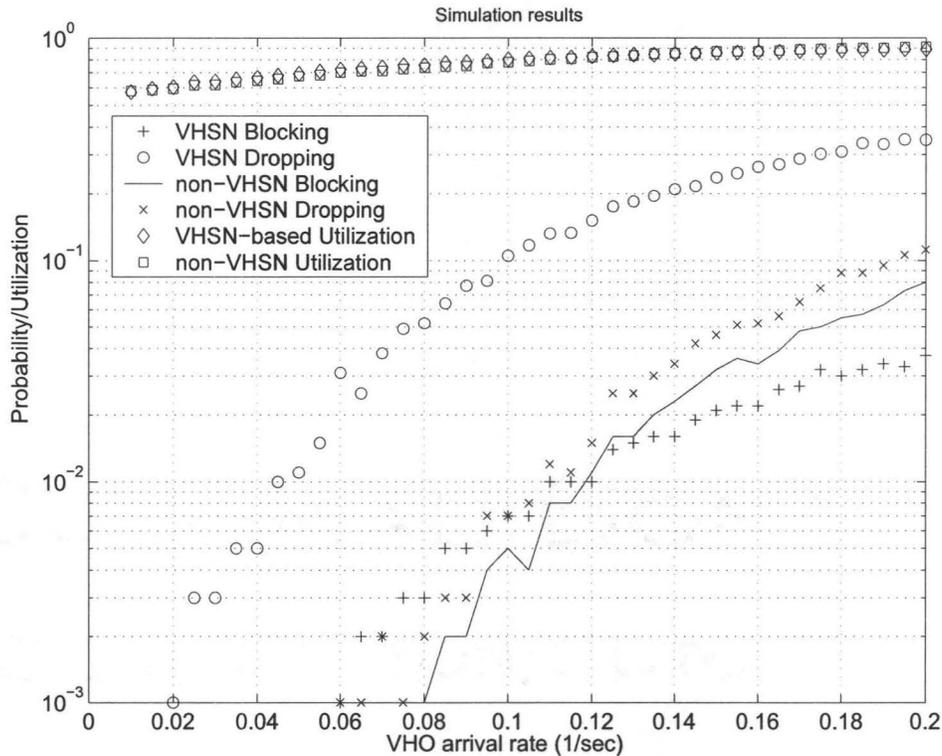


Figure 3.9: Simulation Results of New Call Blocking and VHO Dropping Probabilities. Comparison of VHSN-Assisted and Conventional Soft Handoff.

period. This can be realized by considering the large number of packets which are lost during a VHO as shown in Figure 3.13 and will be shortly discussed. Note that in this chapter we only refer to VHO droppings which happen as a result of having no free channels.

Figure 3.10 shows the same curves derived by numerically solving the Markov process given in Figure 3.7 for $K_{BW} = 2.42$. The analysis is performed for a mean VHO delay of 6.0 sec which is held fixed for different values of VHO arrival rate and assuming there are 80 channels per BS. Blocking and dropping probabilities are then calculated using Equations 3.9 and 3.10. This figure, however, shows only five curves instead of six, since new call blocking and vertical handoff dropping probabilities for the non-VHSN case are both equal. This is because a VHO arrival is just as any new call arrival to the cellular BS and requires the same amount of resources as well.

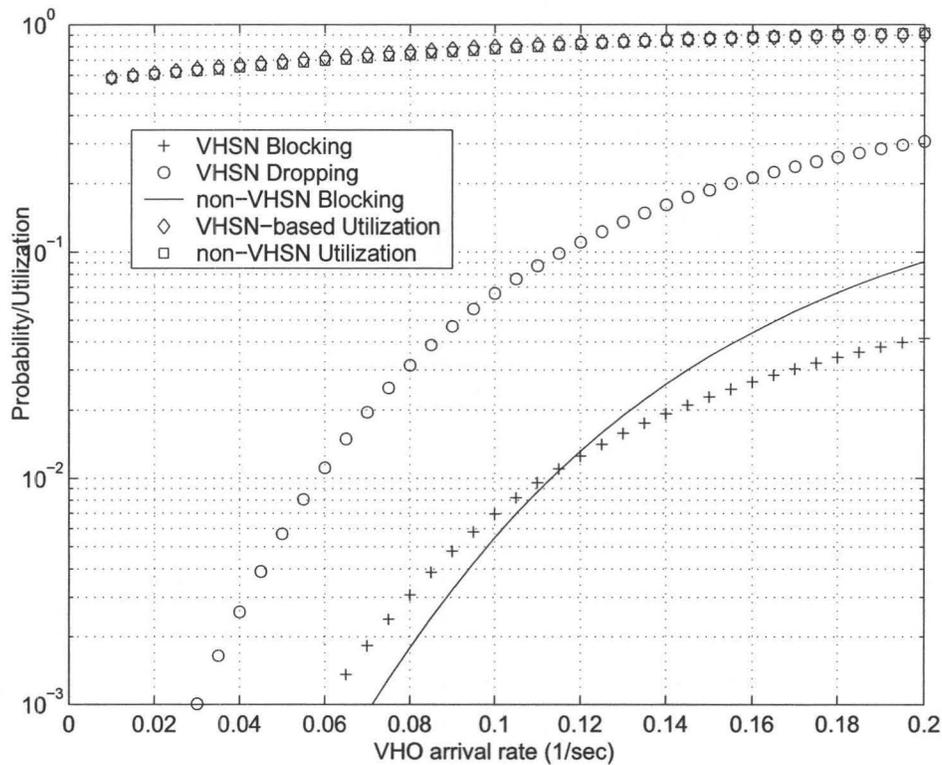


Figure 3.10: Analytic Results for New Call Blocking and VHO Dropping Probabilities. Comparison of VHSN and Conventional Soft Handoff.

Comparing Figures 3.9 and 3.10 it can be seen that the analytic results provide a good approximation to the simulations. This helps to validate the integrity of the simulator.

Figure 3.11 more clearly compares VHO dropping and new call blocking probabilities using the simulation and analysis for the VHSN case. Both results for new call blocking are seen to agree very closely. In terms of VHO dropping rate the analytic model underestimates the true values but is still a reasonable approximation. Figures 3.9, 3.10 and 3.11 illustrate the capacity tradeoff of using VHSN which results in an increase in the VHO dropping rate. Without specific changes on the cellular network side, VHO calls will not be treated any differently than new calls, which is true for any loosely coupled architecture.

Using the VHSN, when the load is relatively high (e.g., for $\lambda_{vho} > 0.100$ in our

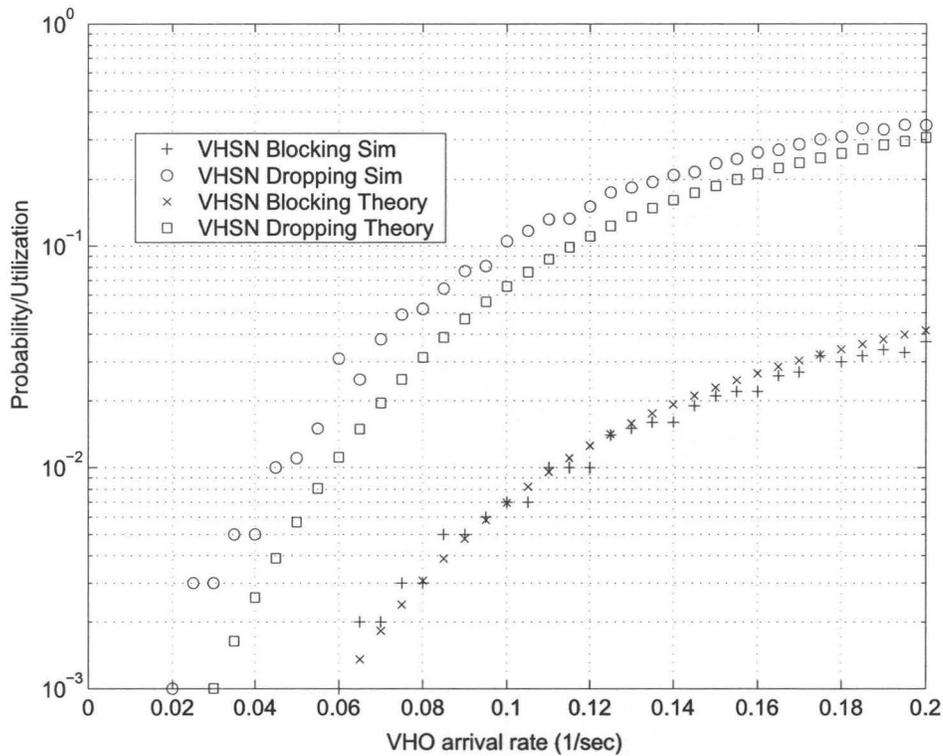


Figure 3.11: *Comparison of Analysis and Simulation.*

results) there is no negative effect on cellular new call blocking probability. This is because new call requests are implicitly prioritized over VHO requests since the bandwidth request per call is lower. For this reason VHO calls are more likely to be dropped. For lower load cases however, new call blocking can be slightly increased due to the additional bandwidth used by VHO calls. This increase is unlikely to cause an undesirable new call blocking probability since at low load, blocking is already very low. This behaviour can be verified in Figures 3.9, 3.10, and 3.11.

The mean, standard deviation, and histogram for packet loss during a VHO is shown in Figures 3.12 and 3.13 from our simulations and analysis. Figure 3.12 shows that less than three packets are lost on average for the VHSN case, which is a very small amount compared to roughly 211 packets lost in the non-VHSN case. The 95% confidence interval for the VHSN and non-VHSN packet loss is 0.7% and 0.1% error around the mean, which suggests that these results are accurate. It can be

seen that mean packet loss for both the VHSN and non-VHSN cases are relatively constant across the range of VHO arrival rates considered. This is in accordance with the analytical formulation of these quantities given in Equation 3.6 for the VHSN case. For the non-VHSN case, the analytical packet loss is calculated by counting the average number of packets received during the outage interval, yielding the following simple equation: $(T_{vho} - T_{wlan}^{fail})\lambda$.

Figures 3.12(a) and 3.12(b) also show the total mean and standard deviation of packet loss across all values of VHO arrival rate as a flat line. The figure also compares mean packet loss for all λ_{vho} for the simulation and analytic cases, which were found to agree very closely, with less than 5% error. For the non-VHSN case packet loss tends to have a higher deviation in the absolute value sense, however, when normalized to the mean, the VHSN case shows more relative variability with a standard deviation to mean ratio of 2.22 compared to 0.40 for the non-VHSN case. This is due to the long outage period associated with the non-VHSN case which dominates the total outage period of a vertical handoff.

We have plotted the collected pdf of packet loss using our simulation results for the VHSN and non-VHSN cases in Figure 3.13. These results suggest that almost 80% of the vertical handoffs do not experience a single packet loss when using the VHSN while only about 5% of the non-VHSN vertical handoffs experience zero packet loss. The probability of having at least one packet loss during a VHO is also analytically derived for both of the VHSN and non-VHSN cases using Equation 3.6 for the value of packet loss under the VHSN scheme and the previously derived simple equation $(T_{vho} - T_{wlan}^{fail})\lambda$ for the non-VHSN packet loss value. These analytic results are also included in Figure 3.13 and validate our simulations with good accuracy. According to our simulations, 21.76% of the VHO calls experience at least one packet loss in the VHSN case, this value is found to be 20.15% from our analytical model.

From these results we can conclude that by using the VHSN, calls that succeed in obtaining a VHSN link experience greatly improved outage and packet loss performance. However, VHO calls may experience a higher dropping rate than in the non-VHSN case. Nevertheless, as it can be seen from the packet loss figures, if the VHSN is not used, a far higher number of calls may be considered as dropped due to

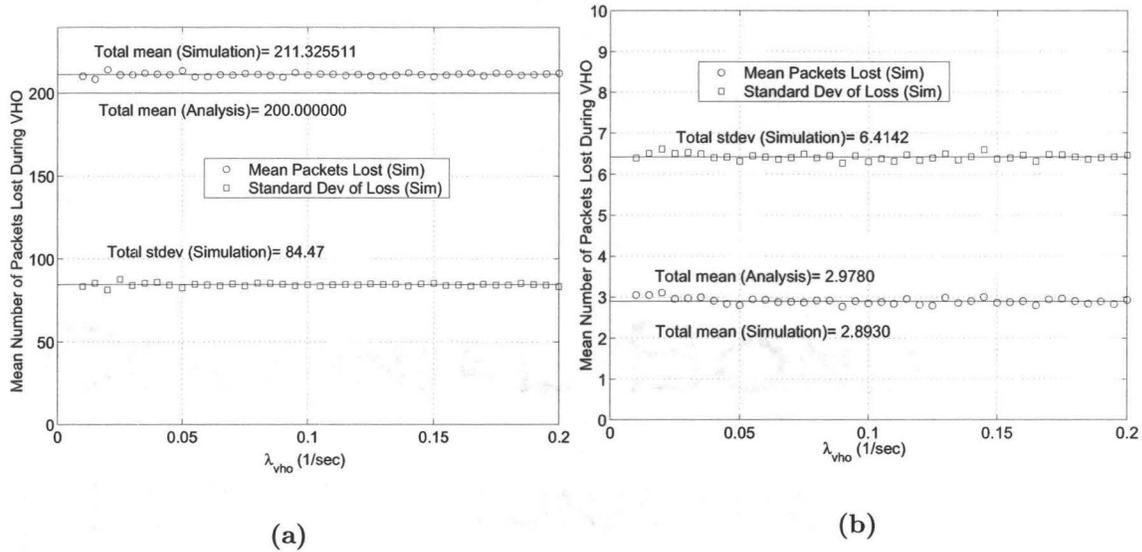


Figure 3.12: Mean and Standard Deviation For Packet Loss During a VHO for (a) the Non-VHSN and (b) the VHSN-Assisted Vertical Handoffs vs. the VHO Arrival Rate.

many packets being lost.

Figure 3.14 shows the mean, standard deviation and pdf for initial VHO delay for the VHSN approach, as the time it takes from when VHO is triggered by the MS to when the first redirected packet is received. The figure shows that the initial delay has an average of about 538 msec with a standard deviation of 16.22 msec across all values of VHO arrival rate. The 95% confidence interval on this parameter is 0.2% error around the mean. It is apparent from Figure 3.14(a) that initial delay is relatively insensitive to VHO arrival rate. This can also be verified from Equation 3.2 which shows the relationship between this parameter and cellular link latency, namely, T_c . Cellular link latency lumps the queuing effects related to packet transmission and contributes to the small variations visible in the pdf for T_{initial} as shown in Figure 3.14(b). This is desirable since it implies that packet loss as a result of unexpectedly long air interface and queuing latencies is unlikely to happen. Nevertheless, packet loss can still happen if the WLAN coverage is lost very quickly. Figure 3.14(a)

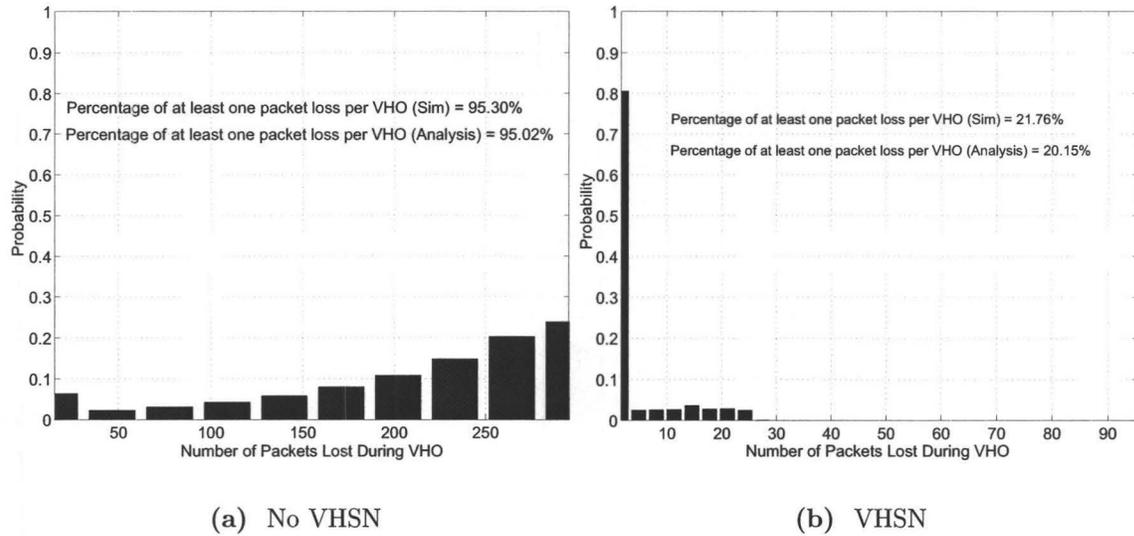


Figure 3.13: Pdf of VHO Packet Loss

also compares the analytic value for $T_{initial}$ derived according to Equation 3.2 to the one derived from simulations and shows that they are close (554msec and 538msec respectively).

To better illustrate the sequence of events during a vertical handoff, we have included packet traces that occur at the MS. These have been generated by our simulator using a cellular new call arrival rate of 0.250/sec, a VHO arrival rate of 0.100/sec, a VHO latency of 3 seconds, $T_{wlan}^{fail} = 2.0sec$ and $T_{connect} = 0.5sec$. Figures 3.15(a) and 3.15(b) show packet trace samples for the non-VHSN and VHSN cases, respectively. One can easily see that the VHSN is able to quickly redirect the flow of packets through the cellular BS so that the MS does not experience noticeable outage and packet loss compared to the non-VHSN case, yielding zero packet loss in this case. The duplicate packets received by the MS through the old WLAN link and the cellular redirect link to the VHSN are also shown in Figure 3.15(b). These duplicate packets continue to arrive until the WLAN coverage is lost.

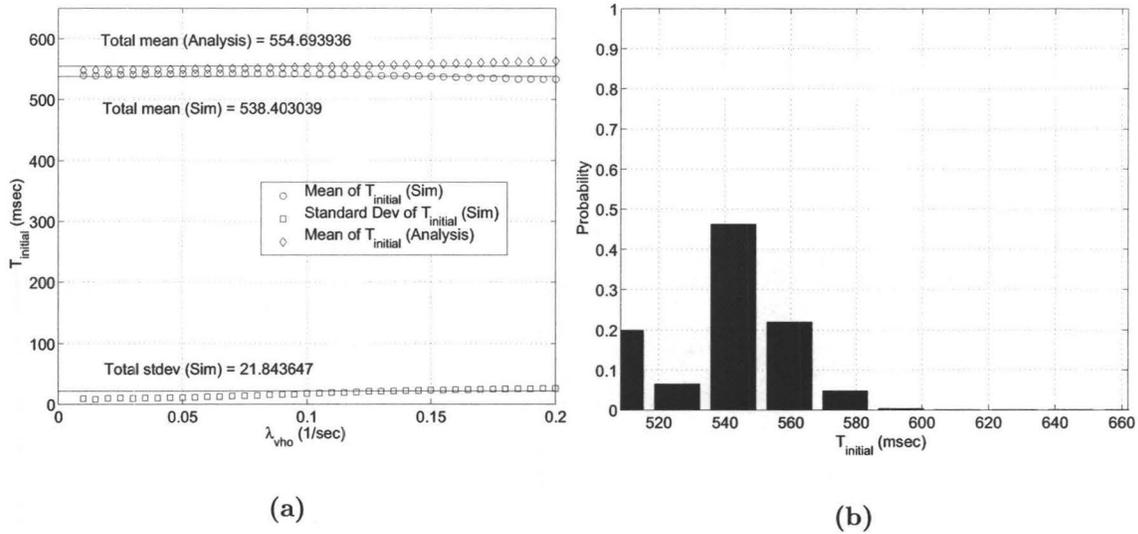


Figure 3.14: Mean (a) and pdf (Simulation), (b) of Initial Delay vs. VHO Arrival Rate.

3.5 Summary

This chapter proposes the use of dynamic client-based cellular media redirection during vertical handoff. This is done by installing a special cellular end station (i.e., the VHSN) on the WLAN infrastructure. A media stream is redirected through this node during the time when soft handoff is taking place. The VHSN connects to the cellular base station like any other cellular station through a standard cellular radio interface.

The proposed mechanism requires no upgrade to WLAN access points or the cellular base station, and is independent of the mobility management scheme used as long as soft handoff is provisioned. The main advantages of this scheme include the following. 1. There are no significant requirements for change to typical WLAN and cellular infrastructure/protocol functionality beyond allowing for cellular recovery-link forwarding, i.e., fast packet redirection is done using a conventional cellular client interface; 2. The proposed scheme is scalable in that the number of simultaneous link recoveries is only limited by bandwidth restrictions imposed by the cellular network.

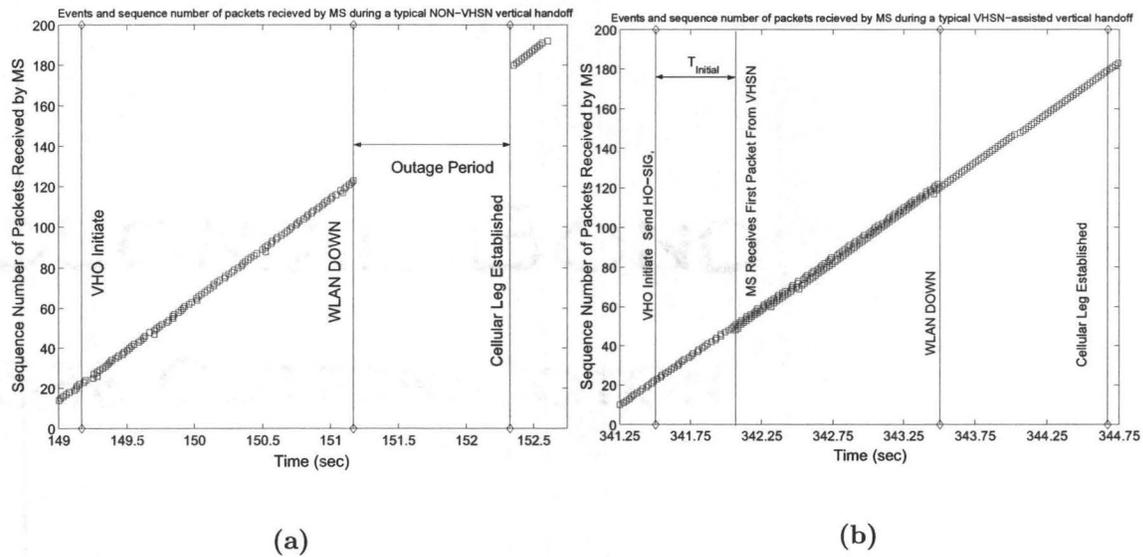


Figure 3.15: Series of Events and Packet Sequence Numbers Received by the MS During (a) Conventional Soft and (b) VHSN-Assisted Vertical Handoff.

For this reason a single VHSN can support a large population of dual mode MSs; 3. The mechanism is link-based and therefore independent of the higher layer mobility management mechanism used.

The results show that packet loss can be substantially reduced or in many cases eliminated. However, due to the additional cellular capacity used by the VHSN there is a tradeoff between VHO dropping probability when using the VHSN and packet loss and outage during VHO when using conventional soft handoff. However, the VHSN approach is shown not to significantly degrade the cellular new call blocking probability. In the next chapter, another issue will be addressed which also contributes to VHO dropping.

Chapter 4

The WLAN Capacity Deficit Problem

In the previous chapter an approach was proposed to reduce the VHO *link-dropping* rate, which happens when a vertical handoff loses its WLAN link due to experiencing an RSSI which is below the minimum required for decoding the signal. In this chapter, another type of VHO dropping, namely, *capacity-dropping* will be discussed. Capacity-dropping occurs when the bandwidth requirement of a VHO call cannot be satisfied by the WLAN AP as a result of its reduced bitrate. The total VHO dropping rate is the sum of its capacity and link-dropping components.

In this chapter, the VHO capacity-dropping for an IEEE 802.11 WLAN deployment is calculated using analysis and simulations. In addition, a bandwidth reservation scheme is employed to reduce capacity-dropping and its performance is evaluated using analysis and simulations. It is assumed that a vertical handoff from the WLAN to the cellular network is handled according to a soft VHO scheme similar to that employed in [STZK06] and illustrated in Figure 3.1.

Table 4.1: Relationship Between a Single G.711 VoIP Call Bitrate, RSSI, and its Bandwidth Consumption on an IEEE 802.11g AP.

Bitrate	Minimum RSSI	Normalized Bandwidth
54Mbps	-72dBm	0.0182
48Mbps	-72dBm	0.0185
36Mbps	-73dBm	0.0200
24Mbps	-77dBm	0.0227
18Mbps	-80dBm	0.0256
12Mbps	-82dBm	0.0312
9Mbps	-84dBm	0.0370
6Mbps	-90dBm	0.0476
2Mbps	-91dBm	0.1667
1Mbps	-95dBm	0.3333

4.1 System Model

There is an upper limit to the number of voice calls¹ that a single WLAN access point can accommodate. One of the most widely used voice encoding standards is ITU G.711, which uses a data rate of 64Kbps. The capacity of an AP for voice calls is calculated based on maintaining a minimum Mean Opinion Score (MOS) per call, which is a function of the bit error rate (BER). To maintain the target MOS, a call experiencing a high BER must shift down to a lower bitrate. As the bitrate of a call is reduced, it consumes more capacity from its AP. Here the terms capacity and bandwidth are used in a broad sense and refer to the amount of resources used to realize a connection. In the case of IEEE 802.11 WLANs, bandwidth refers to the amount of time that a packet transaction takes normalized to the frame length. Therefore, when a call uses a low bitrate it will require more time (hence bandwidth) to send the same number of bits. Various papers have provided voice call capacity calculations for an IEEE 802.11 access point for different voice encodings [PA04]. Table 4.1 presents the normalized bandwidth consumed by a single G.711 VoIP call associated with an IEEE 802.11g AP. This table is derived by extending the capacity

¹The discussion will focus on voice calls using VoIP over WLAN as an example of a real-time network application.

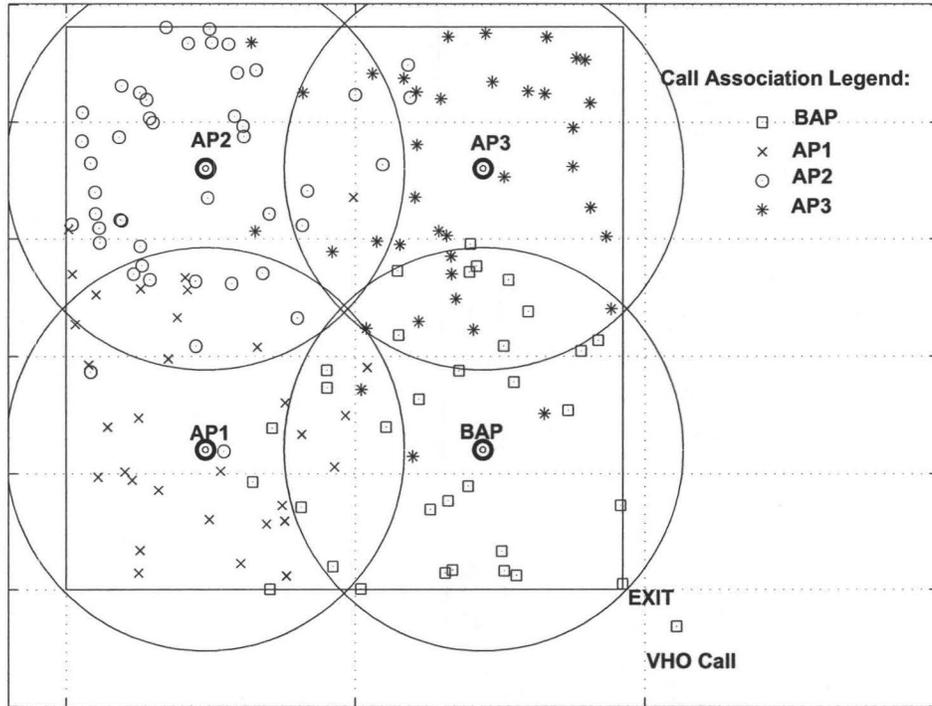


Figure 4.1: Floor Plan of the Coverage Area Considered. There Are Four APs One of Which Covers the Exit and is Named BAP. The Coverage Area of Each AP Corresponding to S_{min} is Shown. Also Active MSs Holding Calls, Which are Distributed Throughout the Coverage Area are Shown. In Addition a Single MS is Performing a VHO While Exiting the Building.

calculation method of [PA04] to an IEEE 802.11g AP. It also shows the receiver sensitivity in terms of minimum RSSI for each bitrate, based on typical Cisco AP specifications [Cisb].

Figure 4.1 shows an example of a WLAN deployment with four APs. A *boundary AP* (BAP) is defined as one which provides coverage for an exit point of the deployment. The APs provide coverage such that a minimum RSSI of S_{min} is received from at least one of the access points at any indoor location with a certain probability (typically above 95%). This is done to ensure that all indoor mobile stations (MS) experience a minimum bit rate of at least R_{min} , with a high probability. System designers can control the capacity of the WLAN by limiting R_{min} . For example, if

a high capacity deployment is desired, then MSs will be able to communicate with higher bitrates. Therefore, S_{min} must be increased accordingly, which means that the intended coverage area of each AP is reduced. The coverage area of each AP corresponding to S_{min} is shown in Figure 4.1. This figure also shows calls² which are distributed throughout the coverage area.

Despite using more APs to increase S_{min} , some mobile stations (MS) may remain connected to an AP with an RSSI less than S_{min} . This may also happen when MSs located outside the desired coverage area connect to the APs. Such stations may consume a large portion of the WLAN capacity. To prevent this, it is common to put a lower limit on the call bitrate. This technique is called, *bitrate clipping*, and is enforced through various information element advertisements made by the APs at the time of association. Under this policy, calls which go below the clipping bitrate, R_{min} , will be disconnected by the AP.

Figure 4.1 shows a single call that is performing a VHO while exiting the indoor coverage region. A vertical handoff is triggered whenever a MS holding a call experiences an RSSI below S_{tp} . The triggering threshold should be set low enough to prevent false triggering while a user is indoors, i.e., $S_{tp} < S_{min}$. False triggering is undesirable because it wastes MS resources by causing unnecessary signalling overhead as well as preventing users from enjoying the higher bandwidth provided by the WLAN. It also causes the user to unnecessarily connect to the cellular network.

Conversely, the VHO threshold should be set high enough such that the VHO procedure can finish successfully before losing the WLAN link. The WLAN link is lost when its bitrate becomes less than the clipping rate, R_{min} . The clipping rate is normally set well above the minimum bitrate achievable for the specific WLAN technology. Therefore, even if a soft VHO technique similar to the one illustrated in Figure 3.1 is employed, it is still possible to experience outage and packet loss if the WLAN link is lost before completing the VHO.

²In this context calls and active mobile stations refer to the same concept.

4.2 Limited Data-rate Usage

A recent measurement based study [SAT06] has shown that in many loosely coupled scenarios, seamless WLAN-to-cellular handoff is difficult to achieve due to large VHO latencies. It was found that unless the WLAN link is permitted to rate adjust itself to low bitrates (e.g., 1-2 Mbps for IEEE 802.11b/g), a call is very likely to lose its link before completing the VHO process. This, however, may not be possible in many Wi-Fi deployments which use bitrate-clipping. If a MS requires lower bitrates during a VHO, then these may not be allowed causing the WLAN link to be dropped before the VHO is complete. Conversely, if low data rates are permitted without any restriction, then it is difficult to ensure high WLAN capacity.

The above discussion presents a dilemma. If a high capacity deployment is desired then mobile stations must not be allowed to hold links at low data rates. But if this is enforced, then vertical handoff performance will be compromised. This conclusion is supported by various studies showing that maintaining the WLAN link down to low bitrates severely reduces the capacity of a WLAN [Bic05].

To address the link rate adaptation issue the *Limited Data-rate Use (LDU)* approach proposed in [ASKT06] is adopted. The LDU scheme allows stations to hold links down to low bitrates only for a very limited time period, long enough to finish a vertical handoff. It has been shown that this technique allows for successful VHO, and at the same time protects the capacity of the AP.

4.3 WLAN Capacity Deficit

Using LDU significantly reduces the VHO link dropping rate. However, it creates the following problem. If the mobile station is allowed to hold the link while it rate adjusts to low data rates, then the capacity effects on the WLAN AP can be very negative. According to [PA04], a G.711 VOIP call operating at 1 Mbps consumes about one third of the available (IEEE 802.11a/b/g) bandwidth while it is holding the link. However, since the VHO may only last for a short period of time, the effects on *average* AP bandwidth are very small. The problem is that in order for the link to

drop to 1-2 Mbps, this amount of bandwidth must be available with a high probability during the short VHO execution phase. We refer to this as the *transient bandwidth* requirement, and will show that this can severely affect the performance of the VHO.

To evaluate the performance of the LDU scheme, a computer simulation was conducted for a WLAN deployment consisting of four IEEE 802.11g APs covering a square indoor region, which has an exit point at one of its corners (Figure 4.1). A fixed number of MSs move according to a random way-point mobility model with wall reflections [BHPC04]. This model is shown to result in a uniform user distribution in steady state. Each user initiates new calls with exponentially distributed inter-arrival times, resulting in a total mean new call arrival rate, λ_{new} , for the entire indoor region. Call length is exponentially distributed with a mean of 180 seconds. User residence time is also exponentially distributed with an average duration of $1/\lambda_{exit}$. This is the average time a user (with or without a call) spends inside the coverage area. Exiting the indoor region results in a VHO if a user bears an active call. The mean vertical handoff rate is denoted by, λ_{vho} , and the latency of a VHO, T_{vho} , is chosen to be 4 seconds such that VHO link dropping is negligible. Note that only real-time voice calls are considered in the simulations. Data traffic can consume any free bandwidth and has less priority compared to voice. Different choices for handling data traffic will be discussed in Section 4.4.

Horizontal handoffs (i.e., AP-to-AP) occur if the signal level of an ongoing call drops below a certain threshold which is above S_{min} . Hysteresis is used for horizontal handoffs (HHO) and leverages the coverage overlap between adjacent APs to effectively queue these handoff requests resulting in very low HHO dropping rates [Sto02]. The access point transmission power is fixed at 30mW (15dBm) which is typical for IEEE 802.11g. An exponential path loss plus lognormal variation propagation model is used with an exponent of n and a standard deviation of σ [Rap96]. In the presented results the 95% confidence intervals are also included with the data. Table 4.2 lists the important parameters and their values used in the simulations. System parameters are chosen such that a typical new call blocking rate of about 1% is experienced at the AP. Note that due to their inherently different propagation environments, different propagation parameters are used for the indoor and outdoor segments.

Table 4.2: Simulation Parameters

Parameter	Value	Description
S_{min}	-72dBm	Minimum RSSI across the WLAN deployment
S_{tp}	-75dBm	VHO triggering threshold
R_{min}	24Mbps	WLAN clipping rate
n_{in}	2.7	Indoor path loss exponent [Rap96]
σ_{in}	3.5dB	Indoor lognormal shadowing std. dev. [Rap96]
n_{out}	3.0	Outdoor path loss exponent [Rap96]
σ_{out}	7.0dB	Outdoor lognormal shadowing std. dev. [Rap96]
λ_{new}	0.2335 1/sec	Total new call arrival rate
$1/\lambda_{exit}$	9000 sec	Average user residence time in the indoor region
λ_{vho}	0.0172 1/sec	Average VHO rate
$1/\mu_{call}$	180 sec	Average call duration
v_{user}	1m/sec	User mobility speed (pedestrian)
A_{WLAN}	9248 m^2	Area of the designated indoor WLAN coverage
N_{ap}	4	Number of access points in the deployment
T_{vho}	4sec	VHO latency

Table 4.3 presents the simulation results for the performance evaluation of the limited data rate use scheme. The results are for the boundary AP, as it is significantly affected. Horizontal handoff (HHO) performance is acceptable, having a dropping rate of about 0.02%. Vertical handoff dropping due to link loss (VHO link-dropping), is nearly 0.04% which is also very small because of the LDU scheme. However, vertical handoff dropping as a result of capacity deficit at the boundary AP (VHO capacity-dropping), is now 1.61%. The total VHO dropping rate, which is the sum of link and capacity dropping is thus, 1.65% which is an unacceptable value and is even larger than the new call blocking rate. In general, it is desirable to have call dropping rates which are one to two orders of magnitude less than call blocking rates. The rationale behind this is that it is more acceptable for a system to block calls which have not already started than to drop those which are currently ongoing. In the next section, a bandwidth reservation scheme is proposed to reduce the VHO capacity-dropping rate.

Table 4.3: Performance Evaluation of the LDU scheme.

Parameter	Value (Percentage)	95% Confidence Interval
New call blocking	1.03%	[0.99% 1.07%]
HHO dropping	0.0225%	[0.0199% 0.0252%]
VHO link-dropping	0.0377%	[0.0304% 0.0450%]
VHO capacity-dropping	1.61%	[1.50% 1.72%]
VHO total dropping	1.65%	[1.54% 1.76%]

4.4 The Cellular Guard Channel Scheme

The WLAN capacity deficit problem can be solved using bandwidth reservation. In this section the classical cellular guard channel approach is adopted to reserve bandwidth for vertical handoffs. The well-known guard channel (GC) approach is designed to improve handoff performance in cellular networks by reserving a certain number of channels, G , out of the total N channels at each cell to accommodate handoffs from other cells [RRT96] [Gue88]. This way the handoff dropping rate can be substantially reduced. When a new call initiates in a cell, it will only be accepted if there are no more than $N - G$ occupied channels, otherwise, the new call will be blocked. On the other hand, handoffs coming from adjacent cells will be accepted if there is at least one channel available.

In conventional handoff, resources are normally reserved in the *forward* direction, i.e., on the base station that is the target of the handoff. However, in the case of WLAN to cellular VHO, bandwidth must be reserved in the *reverse* direction on the WLAN AP which is hosting the rate-adjusted link, i.e., the boundary AP. This adapted version of the GC approach, will be called the *static reservation* scheme (SR). The static reservation scheme pre-allocates a portion, G_{vho}^{static} , of the boundary AP's bandwidth, which will be only assigned to VHO calls if it is needed.

Although the focus of this thesis is on the capacity ramifications for real-time connections, a discussion of best effort data traffic is also necessary. One common strategy for data traffic bandwidth provisioning is to assign a predetermined portion

of the AP bandwidth only to data. This provides data traffic with a guaranteed bandwidth at the expense of reducing the voice call capacity. In such an allocation strategy, two approaches can be taken for handling VHO voice calls which need extra bandwidth. First, the data portion of the bandwidth can be regarded as a hard reservation which will not be shared with VHO calls. Therefore, such a VHO call will be dropped if there is not enough bandwidth available in the voice portion of the allocation. Alternatively, VHO calls can take priority over data traffic and be allowed to use the data bandwidth for a short while. This way, no vertical handoff call will be dropped due to a capacity deficit and there will be no need for bandwidth reservation.

Preallocating a fixed portion of the bandwidth to data limits the voice capacity of an AP. A voice-based WLAN deployment is designed to meet a certain voice traffic demand which is expressed in terms of the new call blocking probability for voice calls. Therefore, it is proposed that voice traffic always takes priority over data in accessing the medium. Data traffic will use this bandwidth whenever it is available. In this case, there is no explicit bandwidth provisioning for the data traffic. Since data traffic tends to be bursty and has a low average bandwidth requirement, this way of bandwidth arbitration will suffice user requirements in most cases. Nevertheless, it is possible to accommodate for a higher data traffic volume in a given environment by tailoring the WLAN deployment for a lower voice call blocking rate. This in effect reduces the AP voice traffic load in order to support more data. This data allocation strategy is adopted throughout the rest of the thesis.

4.5 Performance Evaluation

In this section simulations are performed to evaluate the performance improvements gained using the SR approach. A WLAN deployment such as the one shown in Figure 4.1 is considered and the simulation scenario is as previously described in Section 4.3. The parameters used are also listed in Tables 4.1 and 4.2.

Figure 4.2 shows the new call blocking, horizontal handoff dropping, and vertical handoff dropping probabilities for different fractions of guard bandwidth under the

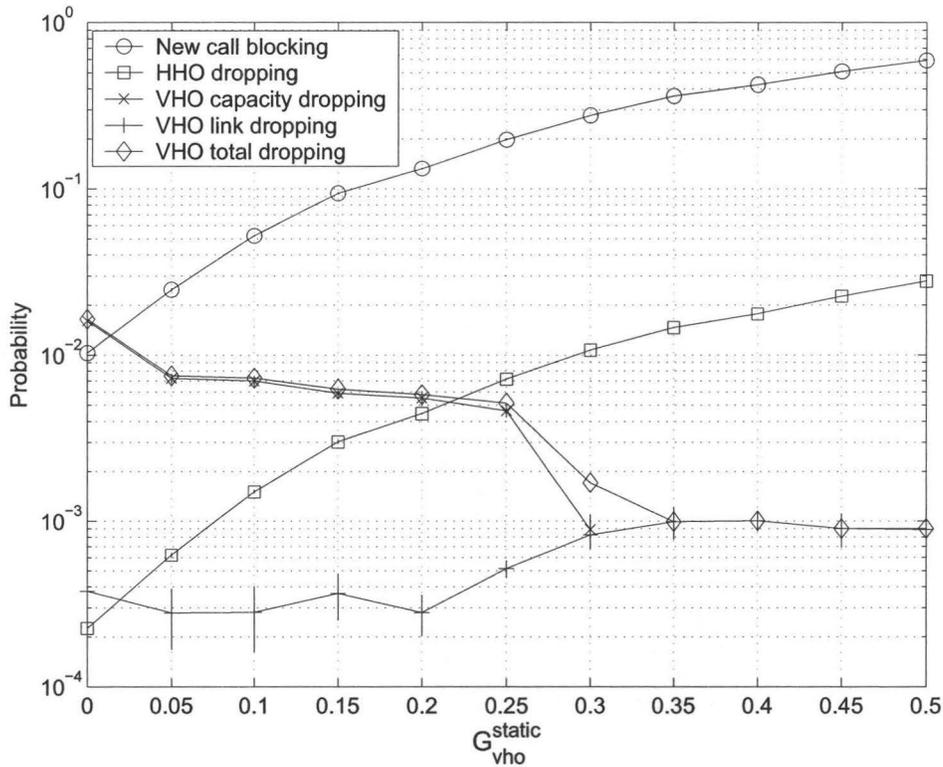


Figure 4.2: BAP Blocking and Dropping Probabilities vs. Fraction of Static Guard Bandwidth Reservation.

static reservation scheme. The figure also shows two different components of the VHO dropping rate, namely, link-dropping and capacity-dropping. For the chosen set of parameters, link-dropping is negligible compared to capacity-dropping. The new call blocking rate corresponds to the boundary AP and the HHO dropping rate belongs to those calls that perform horizontal handoffs to the boundary AP. Other APs will experience only minor performance changes and are not considered.

According to Figure 4.2, it is possible to substantially reduce VHO capacity-dropping with a large bandwidth reservation. For example, capacity-dropping can be reduced from 1.6% to almost zero if 35% of the BAP bandwidth is reserved for vertical handoffs. Note that the total VHO dropping rate is lower bounded by link-dropping and saturates at a maximum of 0.1% for the given set of parameter values.

The significant reduction in VHO dropping rate, however, comes at the price of an increase in new call blocking and HHO dropping rates for the BAP. In the example shown, this causes new call blocking to become 36 times larger going from 1.0% to almost 36.3% and HHO dropping to roughly increase by two orders of magnitude from about 0.02% to 1.46%. Since it is considered better to block new calls than to drop horizontal handoff (HHO) calls, the increase in HHO dropping can alone render the SR scheme unsuitable for the chosen parameters. Any scheme which improves VHO capacity-dropping should maintain a reasonable new call blocking rate and HHO dropping rate.

The source of the elevated new call blocking rate in the SR scheme is due to the following two differences between VHO and other calls. First, VHO calls only need the reserved bandwidth for a very short period of time ($\leq T_{vho}$), compared to a call holding time, e.g., (180sec). In addition, VHO events rarely occur. These two factors result in a poor utilization of the reserved bandwidth. Second, a call performing vertical handoff may need a far larger bandwidth reservation than an ordinary call to guarantee its success since it may drop to a very low bit rate. For example, if a VHO call holds the link down to 1 Mbps, it will require 33% of the AP bandwidth otherwise it will be dropped or experience unacceptable outage. For these reasons the reserved bandwidth needed to achieve a reasonable VHO dropping rate will be far larger than the typical values required by cellular networks.

4.6 Modeling and Analysis

In this section, a mathematical model is used to evaluate the new call blocking and VHO capacity-dropping rates for static guard bandwidth reservation. The number of voice calls at the BAP is modeled using a standard birth-death Markov process with arrivals due to new calls which originate in the BAP coverage area and horizontal handoffs that arrive from adjacent APs. For the sake of simplicity it is assumed that all non-VHO calls have the same bitrate and therefore need the same amount of bandwidth from the BAP. This translates into a maximum nominal call capacity, C , for the BAP which will be used to develop the Markov chain.

The capacity of an AP is expressed as the number of equal bandwidth calls that it can support. Let $P_{rx}(x)$ be the received power in dBm from the AP at a distance x . According to the exponential path loss with lognormal shadowing propagation model one can write,

$$P_{rx}(x) = P_{tx} - PL_0 - 10n_{in} \log(x) + X, \quad (4.1)$$

where, P_{tx} is the transmission power of the AP, PL_0 is the reference path-loss at a one meter distance from the AP and X is a zero mean Gaussian random variable with standard deviation of σ_{in} representing the indoor lognormal shadowing component. It follows that the probability of having a call at location x with a bitrate of $r(i)$ is $q_i(x)$ and can be derived as,

$$\begin{aligned} Pr \{S_{r(i)} \leq P_{rx}(x) < S_{r(i-1)}\} = \\ Pr \{S_{r(i)} - P_{tx} + PL_0 + 10n_{in} \log(x) \leq X < S_{r(i-1)} - P_{tx} + PL_0 + 10n_{in} \log(x)\} = \\ q_i(x) = erf \left(\frac{S_{r(i-1)} - P_{tx} + PL_0 + 10n_{in} \log(x)}{\sigma_{in}} \right) - erf \left(\frac{S_{r(i)} - P_{tx} + PL_0 + 10n_{in} \log(x)}{\sigma_{in}} \right), \quad (4.2) \end{aligned}$$

where, $erf(\cdot)$ is the standard Gaussian error function, $S_{r(i)}$ is the minimum RSSI corresponding to bitrate $r(i)$, $i = 1..M$ and $S_{r(0)} = +\infty$, where $r(i+1) < r(i)$. Consequently, one can express the average bandwidth, $b(x)$, consumed by a call at distance x from the AP as,

$$b(x) = \sum_{i=1}^M q_i(x) B_{r(i)}, \quad (4.3)$$

where $B_{r(i)}$ is the bandwidth required to operate at a bitrate of $r(i)$. The coverage area periphery of an AP is determined by the distance, d_{AP} , at which the signal strength of the AP is above S_{min} with some probability (e.g., $> 90\%$). Integrating Equation 4.3 over the disk representing the coverage area of the AP and dividing it by the actual area of the coverage disk,

$$B(d_{AP}) = \frac{2}{d_{AP}^2} \int_{x=0}^{d_{AP}} b(x) x dx, \quad (4.4)$$

where $B(\cdot)$ approximately represents the average normalized bandwidth used by a call within the coverage region of the AP as a function of the access point range, d_{AP} . Hence the nominal capacity of such an AP is,

$$C(d_{AP}) = 1/B(d_{AP}). \quad (4.5)$$

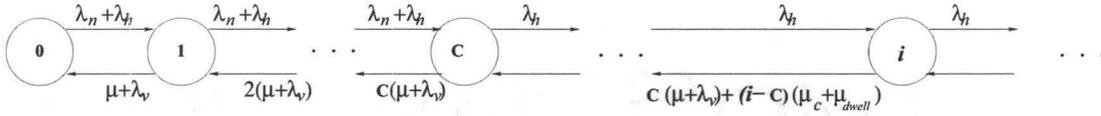


Figure 4.3: Markov Chain Modeling the State of an AP With Queuing of HHO Requests.

When there is not enough bandwidth available at the BAP, the HHO requests are queued while in the coverage overlap between adjacent APs and the BAP. In order to take the effect of these HHO requests into account, the state space of the Markov chain is extended to infinity, where states that exceed C account for those HHO calls that are waiting for bandwidth to be freed so they can handoff to the BAP. This method of modeling the BAP is common when HHO request queueing is used and can be found along with derivations of new call blocking and HHO dropping rates in [Sto02]. The Markov Chain used to model this is shown in Figure 4.3 and its state probability mass function (i.e., pmf) and new call blocking probability, P_{nb} , can be derived as [Sto02],

$$\pi_k = \begin{cases} \frac{(\lambda_n + \lambda_h)^k}{k!(\mu + \lambda_v)^k} \pi_0 & 0 \leq k \leq C \\ \frac{(\lambda_n + \lambda_h)^C \lambda_h^{k-C}}{C!(\mu + \lambda_v)^C \prod_{j=1}^{i-C} [C(\mu + \lambda_v) + k(\mu_c + \mu_{dwell})]} \pi_0 & k \geq C \end{cases}, \quad (4.6)$$

$$P_{nb} = 1 - \sum_{i=0}^{C-1} \pi_i, \quad (4.7)$$

where λ_n and λ_h correspond to the new call and horizontal handoff arrival rates, respectively. Also, $1/\mu$ refers to the average call residence time in one AP, while $1/\mu_c$ is the average call duration. The average amount of time that a HHO request remains in the coverage overlap between two adjacent APs before losing its connection is also modeled by the parameter, $1/\mu_{dwell}$. Note that the vertical handoff rate per call is denoted as λ_v to differentiate between this and the total vertical handoff rate, λ_{vho} . Moreover, the probability of being in state zero, π_0 , for this Markov chain is expressed as [Sto02],

$$\pi_0 = \left\{ \sum_{i=0}^C \frac{(\lambda_n + \lambda_h)^i}{i!(\mu + \lambda_v)^i} + \sum_{i=C+1}^{\infty} \frac{(\lambda_n + \lambda_h)^C}{C!(\mu + \lambda_v)^C} \frac{\lambda_h^{i-C}}{\prod_{j=1}^{i-C} [C(\mu + \lambda_v) + j(\mu_c + \mu_{dwell})]} \right\}^{-1}. \quad (4.8)$$

In order to use the aforementioned Markov model, it will be assumed that VHO events are rare compared to new call and HHO arrivals and that once a call triggers a VHO it will immediately exit the system. With this assumption, the effect of ongoing vertical handoffs in terms of bandwidth consumption on overall BAP performance and load can be ignored. This is a very reasonable assumption for a large range of system parameter values. Throughout the analysis we are interested in the state probability distribution of the BAP whenever a VHO event has occurred. That is, we would like to calculate $Pr\{N|VHO\}$, where N is the number of calls in the BAP. Using Bayes' Theorem one can write,

$$Pr\{N|VHO\} = \frac{Pr\{VHO|N\}}{Pr\{VHO\}} \pi_N, \quad (4.9)$$

where, $Pr\{VHO\}$ is the probability of having a VHO. According to the Markov chain in Figure 4.3 one can write for $0 < N < C$,

$$Pr\{VHO|N\} = \frac{N\lambda_v}{\lambda_n + \lambda_h + N(\mu + \lambda_v)}, \quad (4.10)$$

and assuming that $Pr\{N > C\}$ is negligible yields the following equation for $Pr\{VHO\}$

$$\sum_{k=0}^C \frac{k\lambda_v}{\lambda_n + \lambda_h + k(\mu + \lambda_v)} \pi_k. \quad (4.11)$$

Noting that $\mu \gg \lambda_v$, one can ignore λ_v in the denominator of Equations 4.10 and 4.11 arriving at,

$$Pr\{VHO|N\} = \frac{N\lambda_v}{\lambda_n + \lambda_h + N\mu}, \quad (4.12)$$

$$Pr\{VHO\} = \sum_{k=0}^C \frac{k\lambda_v}{\lambda_n + \lambda_h + k\mu} \pi_k. \quad (4.13)$$

Therefore, substituting π_k from Equation 4.6, the ratio $\frac{Pr\{VHO|N\}}{Pr\{VHO\}}$ can be written as,

$$\frac{N\mu}{(\lambda_n + \lambda_h + N\mu)\pi_0 \sum_{k=0}^C \frac{k(\lambda_n + \lambda_h)^k}{k!\mu^k(\lambda_n + \lambda_h + k\mu)}}. \quad (4.14)$$

Note that the states which have a non-negligible probability of occurring are those close to the mean state number (i.e., \bar{N}). It can be shown that, $\frac{Pr\{VHO|N\}}{Pr\{VHO\}}$, approaches

one when N becomes close to \bar{N} . It then follows from Equation 4.9 that states which are close to the mean and have a higher chance of occurrence will closely obey the unconditional steady state pmf. For lower values of N , however, $\frac{Pr\{VHO|N\}}{Pr\{VHO\}}$ is small but π_N is also small, meaning that such states do not contribute much to the steady state pmf of the Markov chain. More importantly, $\frac{Pr\{VHO|N\}}{Pr\{VHO\}}$, becomes independent of λ_v when $\mu \gg \lambda_v$. Figure 4.4 compares the steady state pmf for the Markov chain presented in Figure 4.3 to its conditional state pmf given a VHO event has occurred. It can be observed that conditioning on the occurrence of a VHO event does not substantially change the state pmf, therefore it is assumed that a VHO event will sample the steady state of the Markov process used to model BAP. This is a key property that will be taken advantage of during the rest of the analytical development.

Using the random way-point mobility model with wall reflections, the user population will be uniformly distributed throughout the coverage area at steady state [BHPC04]. Therefore the new call arrival events will be uniformly distributed throughout the coverage area and equally divided between all the APs, hence,

$$\lambda_{new}^{bap} = \lambda_{new}/N_{ap}. \quad (4.15)$$

The value of λ_{hho}^{bap} is estimated by running separate simulations for the mobility model. An analytical derivation of the HHO arrival rate can also be found in [BHPC04]. Knowing these two arrival rates one can find the new call blocking and HHO dropping rates for the BAP according to [Sto02] and also as mentioned in Equation 4.7 for the blocking rate.

The next step is the derivation of VHO capacity-dropping rate. A vertical handoff will be dropped if the maximum amount of bandwidth required by a call during its vertical handoff is not free at the BAP. Therefore, one needs to know the probability with which a given bitrate (hence bandwidth requirement) is the minimum experienced by a call during its VHO. With this information, the VHO capacity-dropping probability can be calculated as follows,

$$P_{cd} = \sum_{i=1}^M P_{r(i)} Pr\{B_{bap}^f < B_{r(i)}\}, \quad (4.16)$$

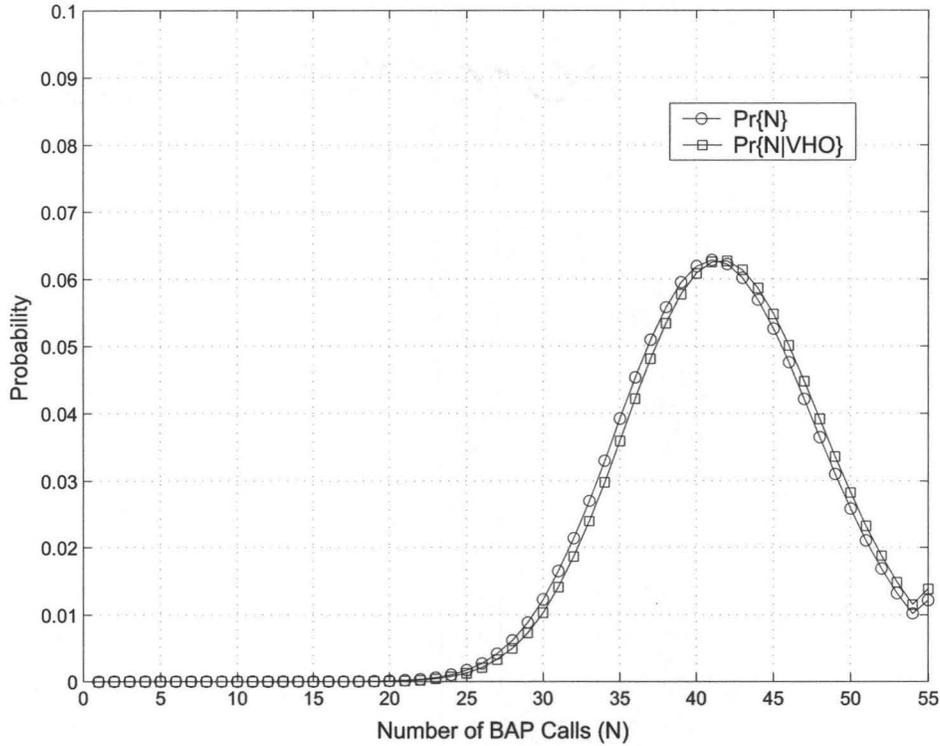


Figure 4.4: Steady State pmf for the Markov Chain in Figure 4.3 and Conditional State pmf Given a VHO Occurrence.

where $P_{r(i)}$ is the probability that a VHO call goes to a minimum bitrate of $r(i)$ during its execution, B_{bap}^f is the free bandwidth available at the BAP, $B_{r(i)}$ is the amount of bandwidth required by the VHO call to operate at bitrate, $r(i)$, and M is the maximum number of bitrates. Note that $r(1)$ and $r(M)$ correspond to the highest and lowest bitrates that are supported by the WLAN technology respectively. B_{bap}^f and $B_{r(i)}$ are translated into the equivalent number of calls based on the average bandwidth per call, $B(\cdot)$, which enables discretization of the state space and the use of a Markov chain to model the BAP.

During a VHO, the RSSI changes based on the exponential path loss and lognormal shadowing propagation model. The time variations in the RSSI experienced by a VHO call while moving away from the BAP can be modeled as a *coloured Gaussian noise*

process with the following exponential correlation function [ZH96],

$$E[X(t)X(s)] = \sigma^2 e^{-\lambda|t-s|}. \quad (4.17)$$

To achieve this, the differential equation governing the coloured noise is given by,

$$\frac{dX(t)}{dt} = -\lambda X(t) + G(t), \quad (4.18)$$

where $G(t) \sim N(0, \sigma)$ is the white Gaussian noise process used to generate $X(t)$, σ is the standard deviation of the coloured noise process which will be set to σ_{out} and $1/\lambda$ is the correlation coefficient in units of time and is set to 20 seconds in our case [ZH96].

Deriving the probabilities, $P_{r(i)}$, from this model is sophisticated and is not the objective of this work, so these probabilities are estimated from running independent simulations for a single VHO while exiting the WLAN for different values of VHO latency, T_{vho} .

The effect of the static bandwidth reservation, G_{vho}^{static} , on the state pmf of the BAP can be included by reducing the maximum call capacity to $\hat{C} = \lfloor C(1 - G_{vho}^{static}) \rfloor$. The state numbers beyond \hat{C} will represent queued HHO requests. Consequently, having the state pmf for the BAP, one can easily calculate VHO capacity-dropping using the modified form of Equation 4.16, replacing bandwidth with its equivalent number of calls as,

$$P_{cd} = \sum_{i=1}^M P_{r(i)} Pr\{N_{bap}^f < N_{r(i)}\}. \quad (4.19)$$

Figure 4.5 shows the performance results for the static reservation scheme derived using Equations 4.7 and 4.19 and also compares them to the simulation results presented previously in Figure 4.2. The set of parameters used are the same as those given in Table 4.2. The analytic new call blocking curve closely matches the simulation results, while for VHO capacity-dropping, the analysis is roughly in agreement with the simulations. This is due to the simplifying assumption that VHOs are rare events and that the probability of more than one happening at the same time is zero, whereas in the simulations, there is a chance that two VHOs overlap and thus consume more resources from the BAP such that one of them would be more likely to

drop leading to a higher P_{cd} for the simulations. In the case shown here, there is a 7% chance that VHO events overlap.

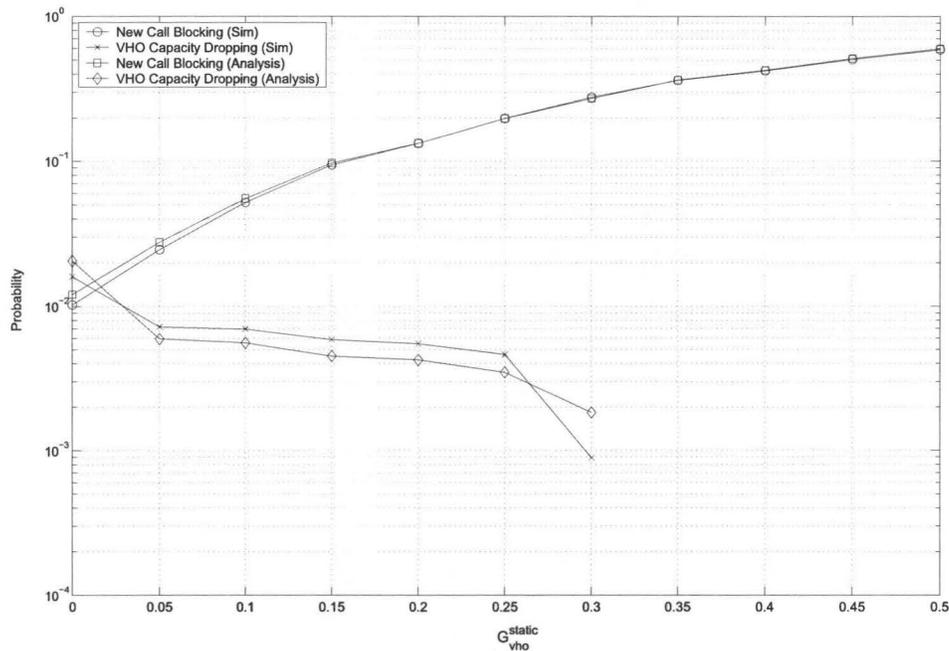


Figure 4.5: Comparison Between the Analytical and Simulation Results for the Static Reservation Scheme.

When no reservation is performed, i.e., the left most point on the curves, analytical VHO capacity-dropping is higher than the simulation results. This is due to the error in the analytical model which estimates a marginally higher load for the BAP. However, when bandwidth reservation is made, the simulated VHO capacity-dropping will become higher than the analytical estimate due to overlapping vertical handoffs.

4.7 Summary

In this chapter, the WLAN capacity deficit problem was presented and the resulting VHO capacity-dropping rate was evaluated using analysis and simulations. It was concluded that capacity-dropping can be significant and by using the static bandwidth

reservation method at the boundary AP, it can be reduced to almost zero. However, the penalty of employing the static reservation scheme is an unacceptable increase in the new call blocking and HHO dropping rates. This is due to the fact that in the worst case, vertical handoff calls need a very large amount of bandwidth. This is however, required for a short period of time. In the next chapter this characteristic of a VHO call is used to propose a scheme for reducing vertical handoff capacity-dropping while minimizing the increase in new call blocking and HHO dropping probabilities.

Chapter 5

Transient Bandwidth Reservation

A vertical handoff call can require a large amount of bandwidth for a short period of time. In addition, a vertical handoff event is rare compared to new call and HHO arrival events, therefore, this large amount of bandwidth is only required occasionally when a VHO is taking place. As a result, the static bandwidth reservation scheme, which books a fixed fraction of the boundary AP's bandwidth will under-utilize this bandwidth. Based on this observation, a *transient bandwidth reservation* scheme is proposed to mitigate the capacity deficit problem faced by VHO calls.

Once a VHO is triggered, a certain fraction of the total BAP bandwidth is reserved for future use by the VHO call. This bandwidth is then released after the VHO has finished execution. It is possible that the desired reservation bandwidth is not available at the boundary AP when it is needed. To overcome this, a bandwidth securing mechanism is used which is based on forced horizontal handoffs with an attempt to *momentarily* evict as many non-VHO calls as possible from the BAP to neighbouring WLAN APs. This mechanism is referred to as *momentary forced HHO* (MFH) and calls which are momentarily forced to do a handoff are referred to as MFH-calls. It is worth mentioning that AP-to-AP handoffs can be executed in a very short amount of time thus not affecting the quality of service for those calls that need to be forced into HHO. There have been many schemes proposed in [MSA04] [SMA04] [RS05] [VK04] which try to reduce AP-AP handoff latency in WLANs using different techniques such as context transfer and neighbour graphs.

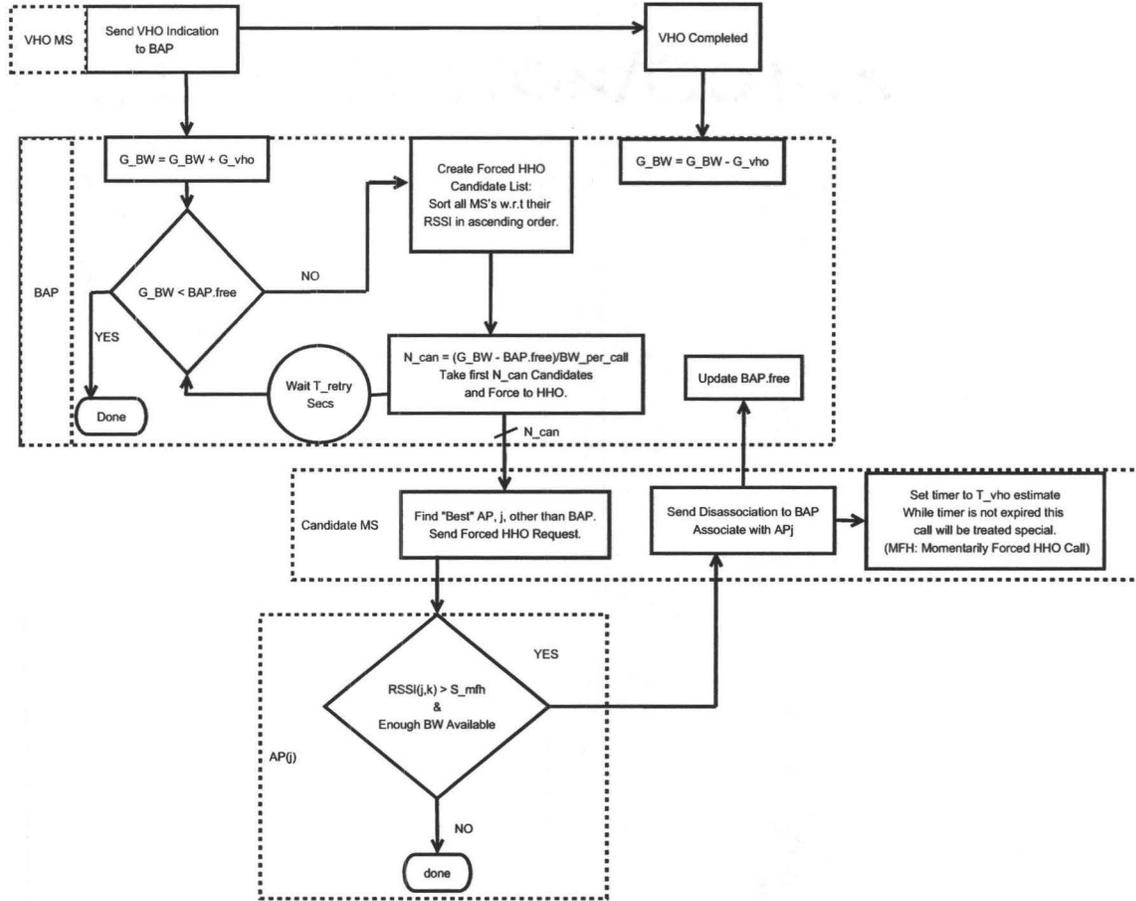


Figure 5.1: Block Diagram for the TR-MFH Scheme.

Figure 5.1 presents a high level flow diagram showing the operation of the transient reservation scheme with momentarily forced horizontal handoff, abbreviated as TR-MFH. When a VHO is triggered, that call will send an indication to the boundary AP, upon which the boundary AP will try to reserve the fraction, $G_{vho}^{transient}$, of its total bandwidth. The actual signalling involved in this process can be implemented in many ways based on IAPP¹ for IEEE 802.11 WLANs. If the BAP is already serving a VHO call then it will increase the amount of bandwidth reservation, G_BW , by $G_{vho}^{transient}$. If this is not available, the BAP will initiate the momentarily forced HHO

¹Inter Access Point Protocol

(MFH) phase by selecting candidate non-VHO calls and requesting them to switch to a suitable neighbouring AP. The BAP picks candidates starting from those calls which have a link with a weaker RSSI level hoping that they are closer to neighbour APs, thus increasing the chance of successful forced HHO. The BAP will pick N_{can} candidate calls for forced HHO, enough to free up bandwidth such that the new reservation amount, G_{BW} , can be accommodated. The BAP will then wait for T_{retry} before re-evaluating the amount of bandwidth deficit. If there is any lack of free bandwidth, it will recalculate N_{can} and attempt another round of forced HHOs until the reservation can be secured or the amount of reservation decreases (e.g., due to a VHO completion). The MFH-retry period T_{retry} should be large enough to allow for forced HHOs. In addition, MFH-retry cannot possibly happen after VHO completion, it follows that, $T_{hho} < T_{retry} < T_{vho}$, where T_{hho} is the latency for performing a HHO; in the simulations $T_{retry} = 1sec$.

In order to be accepted by an adjacent AP, the signal strength of an MFH-candidate call at that AP should be larger than a predefined value, S_{mfh} . This value may be less than the minimum RSSI level, S_{min} , set for other calls and could cause the MFH-call to operate with bitrates less than R_{min} . This is to allow for more acceptance probability for MFH-calls when they are asked to switch. This issue will be discussed when presenting the simulation results in Section 5.1.

Figure 5.2 is a re-drawing of Figure 4.1 and illustrates the floor plan of the WLAN coverage area considered for evaluation and shows that by selecting $S_{mfh} < S_{min}$ it is possible to increase the likelihood of finding MFH-candidate calls and the probability of successful VHO completion. The dotted circles indicate the range for accepting MFH-calls which in this case is shown to be larger than the standard AP coverage area. Those calls that belong to the BAP and are inside the S_{mfh} region are appropriate candidates for the MFH procedure. Note that by starting from calls that have lower WLAN link RSSI, the BAP can more effectively choose MFH-candidates. This method of candidate selection, however, is not guaranteed to select all calls that have RSSI to a neighbour AP higher than S_{mfh} , and as a result some of the MFH-requests may fail.

Upon successful forced HHO, a call will be labelled as an MFH-call for a certain

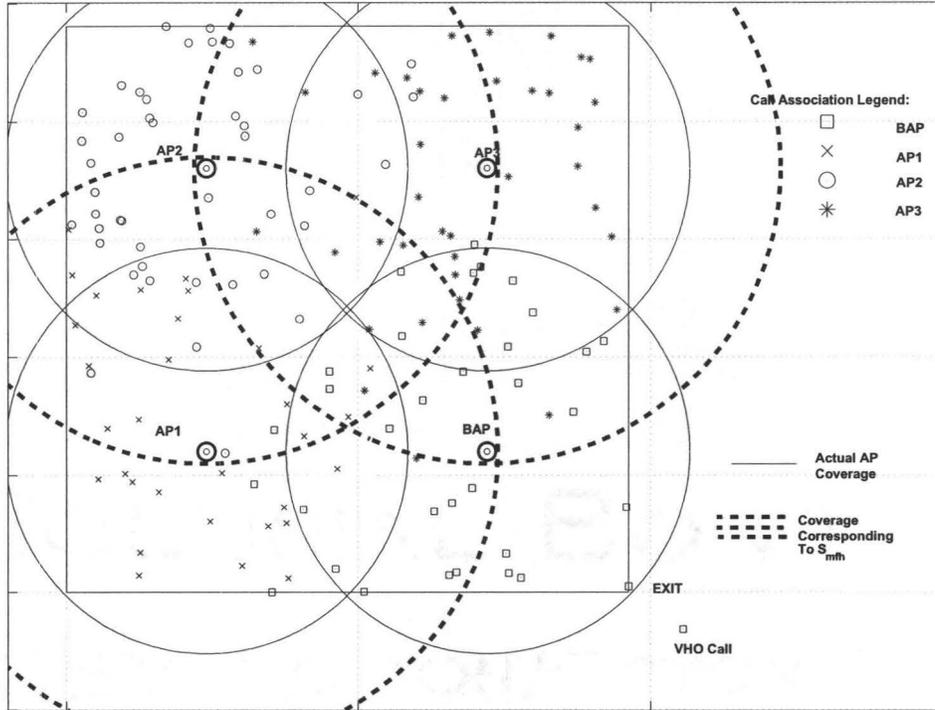


Figure 5.2: *The Extended MFH-Region in the TR-MFH Scheme.*

duration of time, T_{mfh} . During this time, an MFH-call is allowed to operate with an RSSI less than S_{min} as long as it is larger than S_{mfh} . More importantly, if an MFH-call experiences a low signal strength with its current (adjacent) AP and performs a horizontal handoff back to the BAP it will take priority over all other calls and can use the bandwidth reservation made for VHO calls. If the amount of free bandwidth including the reservation at the BAP is not enough to accommodate a *returning MFH-call*, any ongoing VHO call will be terminated and its bandwidth claimed by the MFH-call. This is to prevent the unfairness that is created towards calls with a low RSSI which are candidates for forced HHO but also happen to return to the BAP after a short time. If this strict prioritization is not enforced then these calls could suffer higher HHO dropping rates. However, after the T_{mfh} timer is expired, these calls will be regarded as ordinary calls. The value of T_{mfh} is chosen to be slightly larger than the estimated value of T_{vho} .

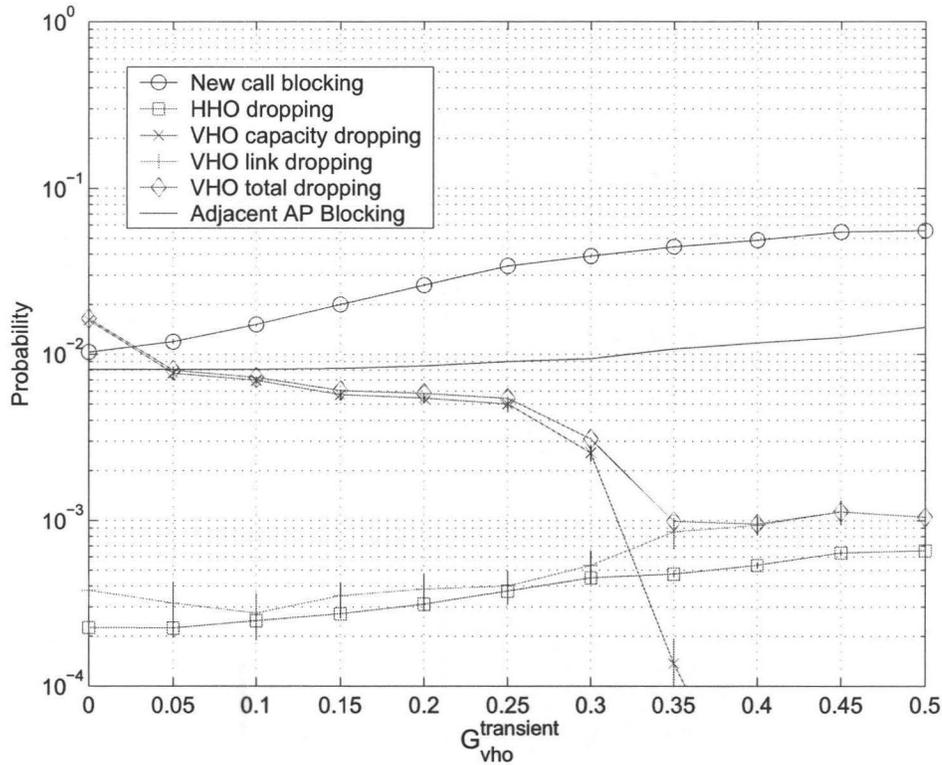


Figure 5.3: Boundary AP New Call Blocking, HHO Dropping and VHO Dropping versus Transient Guard Bandwidth Reservation Using Momentary Forced HHO (TR-MFH).

Note that the MFH process may not be able to free enough bandwidth, resulting in a reservation which is less than the pre-determined fraction. Nevertheless, it is still possible for a vertical handoff to finish if it does not drop to bit rates which need more bandwidth than is currently reserved. The following section presents simulation results for the TR-MFH scheme employed by a WLAN deployment as shown in Figure 5.2, when $S_{mfh} = -74dBm$ and other parameters are set according to Table 4.2.

5.1 Performance Evaluation

Figure 5.3 illustrates the call level performance at the boundary AP when the transient bandwidth reservation approach is employed. It can be observed that to achieve

close to zero VHO capacity-dropping probability, 40% of the bandwidth needs to be reserved for the duration of one vertical handoff once a VHO has been triggered. This is slightly larger than the 35% reservation required for the SR scheme to achieve almost the same VHO capacity-dropping (Figure 4.2). For the TR-MFH approach, new call blocking increases from 1.0% (no scheme) to almost 4.9% when $G_{vho}^{transient} = 0.40$. This is very small compared to the 36% new call blocking rate achieved under the SR scheme for almost the same VHO capacity-dropping. More importantly, the HHO dropping rate for the TR-MFH scheme increases from 0.02% when no reservation is made to 0.05% when $G_{vho}^{transient} = 0.40$, which is significantly lower than the 1.46% HHO dropping as a result of using the SR scheme (Figure 4.2). This shows that the proposed approach clearly outperforms SR by providing the same VHO performance while affecting other calls much less.

The TR-MFH scheme, however, requires more bandwidth reservation to achieve the same VHO capacity-dropping as SR. This is due to the possibility of not being able to force enough calls to switch in order to secure the bandwidth reservation, $G_{vho}^{transient}$, whereas for SR the bandwidth reservation is guaranteed. However, according to the simulation results, the MFH mechanism can efficiently secure the required bandwidth since the difference between $G_{vho}^{transient}$ and G_{vho}^{static} for the same amount of capacity-dropping rate is relatively small.

By momentarily forcing calls to switch to adjacent APs, TR-MFH can potentially increase the new call blocking rate at those APs. Figure 5.3 also includes the new call blocking for AP 1 which is adjacent to the BAP. According to this figure, the new call blocking in AP 1 increases from 0.8% to only 1.2% when 40% of the BAP bandwidth is reserved upon VHO triggering. This negligible increase is due to the temporary nature of the forced HHO procedure and also the fact that there are more than one AP adjacent to the boundary AP in many deployment scenarios. Hence, it is expected that the effect of the MFH procedure on the neighbour APs will be negligible. The 0.2% difference in blocking rate for the BAP and AP 1 when no reservation is made is attributed to the slight increase in the BAP's load as a result of VHO calls employing the LDU scheme.

It was previously stated that having $S_{mfh} < S_{min}$ may result in MFH-calls with

bitrates less than R_{min} at neighbour APs. This can potentially increase the new call blocking rate at these APs. However, as suggested in Figure 5.3 the new call blocking increase due to employing the TR-MFH scheme is small and the contribution of these calls to the blocking rate is even smaller since they constitute a small fraction of the total MFH-calls.

When all other parameters are the same, one would not expect the proposed scheme to outperform the static guard bandwidth approach in terms of VHO capacity-dropping. Among capacity reservation schemes the amount of VHO capacity-dropping under static guard bandwidth reservation is a lower bound for a given bandwidth reservation ratio, G_{vho} . This can be verified by comparing capacity-dropping curves in Figures 4.2 and 5.3. For example, to achieve a VHO capacity-dropping of 0.5%, 25% of the bandwidth needs to be reserved upon a VHO being triggered under the transient reservation scheme, however, by interpolating the curves presented in Figure 4.2, only 22.5% static reservation is enough to achieve the same capacity-dropping rate. Furthermore, an upper bound on VHO capacity-dropping is achieved when no reservation is performed. In the simulated scenario, this is nearly 1.6%.

From a new call blocking and HHO dropping perspective, a lower bound is achieved when no capacity is reserved for vertical handoffs. This is the conventional case when no scheme is employed and corresponds to the leftmost point on all of the curves where the reservation is zero. Moreover, for a given amount of bandwidth reservation, the upper bound on these parameters is achieved when using the static guard bandwidth scheme. Any scheme based on capacity reservation would have its VHO capacity-dropping, new call blocking, and HHO dropping rates somewhere between those for the no-reservation and static guard bandwidth schemes. A further discussion of these issues will be presented in Section 5.2.

These results and many others that are not included indicate that the transient reservation approach when combined with the momentary forced HHO mechanism, performs exceptionally well. While its VHO capacity-dropping rate is practically the same as that of the lower bound achieved under the static guard bandwidth scheme, it also provides new call blocking and HHO dropping rates which are close to the lower bound achieved when no reservation is made. It should be noted, however, that

the proposed mechanism achieves the same VHO capacity-dropping at a higher guard bandwidth reservation ratio, but this does not significantly increase new call blocking and HHO dropping and therefore is not a drawback.

5.2 Modeling and Analysis

To derive the new call blocking rate for TR-MFH, B_{TR} , the Markov model introduced earlier in Figure 4.3 and analyzed in Equation 4.6 will be used with a maximum call capacity of C . In addition, define the events,

$$Y = \text{At least one ongoing VHO upon a new call arrival,} \quad (5.1)$$

$$X = \text{Not enough bandwidth for a new call.} \quad (5.2)$$

Consequently, the new call blocking rate can be calculated as the sum of two conditional blocking rates,

$$B_{TR} = Pr\{X|\bar{Y}\}(1 - P_{vho}) + Pr\{X|Y\}P_{vho}. \quad (5.3)$$

The Markov model in Figure 4.3 describes the birth death process which governs the BAP. For reasonably low values of new call blocking rate, the system can be approximated as one with infinite capacity for new call arrivals, i.e., $C = \infty$. This approximation reduces the Markov chain to an M/M/ ∞ queue since call arrivals are assumed to be a Poisson process and a call occupies a channel for an exponential period of time before it either completes, moves to another WLAN AP, or exits the system by performing a VHO. Note that user residence time inside the WLAN deployment and call duration are assumed to be exponentially distributed. Therefore, due to Burke's theorem the number of calls within the BAP at any time is Poisson distributed with the mean $\rho = \frac{\lambda_n + \lambda_h}{\mu + \lambda_v}$. Note that the waiting time of HHO requests in the coverage overlap between adjacent APs is neglected. The call departure process from the BAP is also Poisson according to the same theorem. A call departs either due to a VHO with mean rate of λ_v , or because of moving to another AP or finishing, with the total rate of μ . This happens independently for each call departing the BAP. Therefore, at the time of departure, a call will initiate a VHO with probability

$\alpha = \frac{\lambda_v}{\lambda_v + \mu}$. Since the departure process is Poisson, the VHO process which is sampling it with fixed probability α is also Poisson with a mean of $\lambda_{vho} = \alpha\rho$. Therefore the probability of having at least one VHO when a call arrives, P_{vho} , is approximated by the probability of a VHO occurring anytime during T_{vho} seconds, it then follows that,

$$P_{vho} = 1 - e^{-\lambda_{vho}T_{vho}}. \quad (5.4)$$

Based on the assumption that VHO events are rare, the probability of having more than one VHO when a new call arrives can be ignored. Hence from Equations 5.3 and 5.4 the new call blocking rate can be written as,

$$\begin{aligned} B_{TR} &= Pr\{N_{bap} \geq C\}(1 - P_{vho}) + Pr\{N_{bap} \geq \lfloor C(1 - G_{vho}^{transient}) \rfloor\}P_{vho} \\ &= 1 - F_{bap}(\lfloor C(1 - G_{vho}^{transient}) \rfloor - 1) - \\ &\quad e^{-\lambda_{vho}T_{vho}}(F_{bap}(C - 1) - F_{bap}(\lfloor C(1 - G_{vho}^{transient}) \rfloor - 1)), \end{aligned} \quad (5.5)$$

where F_{bap} is the cumulative pmf of the Markov chain analyzed in Equation 4.6 to model the BAP state.

According to Equation 4.7 one can write the blocking probability for the no-reservation case as,

$$B_{NR} = 1 - F_{bap}(C - 1), \quad (5.6)$$

and the blocking probability for the static reservation scheme is upper bounded by,

$$B_{SR} \leq 1 - F_{bap}(\lfloor C(1 - G_{vho}) \rfloor - 1), \quad (5.7)$$

where the bound becomes tighter as B_{SR} decreases. Then from Equations 5.5, 5.6 and 5.7 one can write,

$$\begin{aligned} B_{TR} &= B_{NR}(1 - P_{vho}) + Pr\{N_{bap} \geq \lfloor C(1 - G_{vho}) \rfloor\}P_{vho} \\ &\geq B_{NR}(1 - P_{vho}) + B_{SR}P_{vho}. \end{aligned} \quad (5.8)$$

This is a lower bound on the blocking probability for the TR-MFH scheme. It also follows that,

$$B_{NR} \leq B_{TR} \leq B_{SR}. \quad (5.9)$$

According to Inequality 5.9 when the VHO arrival rate is small enough, the blocking rate of the TR-MFH scheme is lower bounded by that of the no-reservation scheme and upper bounded by the blocking rate of the SR approach. For reasonably small values of B_{SR} one can rewrite Inequality 5.8 as the following approximate equation,

$$B_{TR} \approx B_{NR}(1 - P_{vho}) + B_{SR}P_{vho}. \quad (5.10)$$

This equation suggests that the value of B_{TR} approaches B_{SR} as the VHO arrival rate increases and approaches B_{NR} when it decreases. Therefore for a reasonable system having a small VHO arrival rate (e.g., an order of magnitude less than the new call arrival rate) the new call blocking rate of the TR-MFH scheme is expected to be very close to when no reservation scheme is employed. This is a very important property of the TR-MFH scheme.

The next performance metric of interest is the VHO capacity-dropping rate, $P_{cd}^{transient}$, which is derived using Equation 4.19 with the exception that the effect of momentary forced HHO on reducing the number of BAP calls should be taken into consideration. Let us assume that at the time of VHO initiation there are N_{bap} calls being serviced by the BAP. The BAP will try to secure G_{vho} bandwidth and reserve it for the VHO call. There are two possible outcomes to this operation, either there is enough free bandwidth at the BAP or there is not, causing the initiation of the MFH procedure. The condition under which there will be no need for forced HHOs is expressed as,

$$N_{bap}^f = C - N_{bap} \geq [CG_{vho}].$$

It follows that,

$$N_{bap} \leq C - N_G, \quad (5.11)$$

where $N_G = [CG_{vho}]$ is the amount of bandwidth reservation in terms of the equivalent number of calls. On the other hand, if the reserve bandwidth cannot be secured then it is required that N_{req} calls be forced to adjacent APs. The required number of calls for forced HHO can be expressed as follows,

$$N_{req} = N_G + N_{bap} - C. \quad (5.12)$$

Consequently, the MFH procedure will go over all calls associated to the BAP and will find N_{can} valid candidates for forced HHO, out of which a number, N_{mfh} , will be accommodated by the adjacent APs. Note that N_{can} is the actual number of MFH-candidates which also fulfill the RSSI requirements for being accepted by neighbour APs. Some MFH candidates can only switch to one adjacent AP while others might have two or more options. To make the derivations easier, it is assumed that all the MFH candidates will be able to find an adjacent AP with enough free bandwidth. Therefore, the possibility of a lack of capacity at the adjacent APs is ignored. The ramifications of this assumption will be discussed, arguing that it is justified in most reasonable cases. Hence, the eventual number of MFH-calls can be expressed as,

$$N_{mfh} = \max\{0, \min\{N_{req}, N_{can}\}\}. \quad (5.13)$$

In addition, if there is no need for forcing calls out of the BAP, then N_{req} will be less than zero resulting in $N_{mfh} = 0$. Consequently, the number of calls remaining in the BAP at the end of MFH phase is,

$$\widehat{N}_{bap} = N_{bap} - N_{mfh}. \quad (5.14)$$

The required number of calls for forced HHO, N_{req} , is a deterministic function of the number of calls in the BAP at the time of a VHO initiation, N_{bap} . Moreover, observe that if N_{bap} is known, then N_{can} obeys the Binomial distribution. Therefore the conditional probability of having N_{can} MFH candidates given that there are N_{bap} calls associated to the BAP is,

$$f_{(can|bap)}(N_{can}|N_{bap}) = \binom{N_{bap}}{N_{can}} P_{can}^{N_{can}} (1 - P_{can})^{(N_{bap} - N_{can})}, \quad (5.15)$$

where P_{can} is the probability that a given call associated to the BAP is also a valid MFH-candidate and needs to be calculated. Figure 5.4 shows the WLAN deployment map along with the intended coverage area for the BAP corresponding to an RSSI of larger than S_{min} and the MFH-range for all three adjacent APs, in which the RSSI is greater than S_{mfh} . For the sake of simplicity, these areas are shown as perfect circles. Nevertheless, since the propagation shadowing components have a zero mean and are

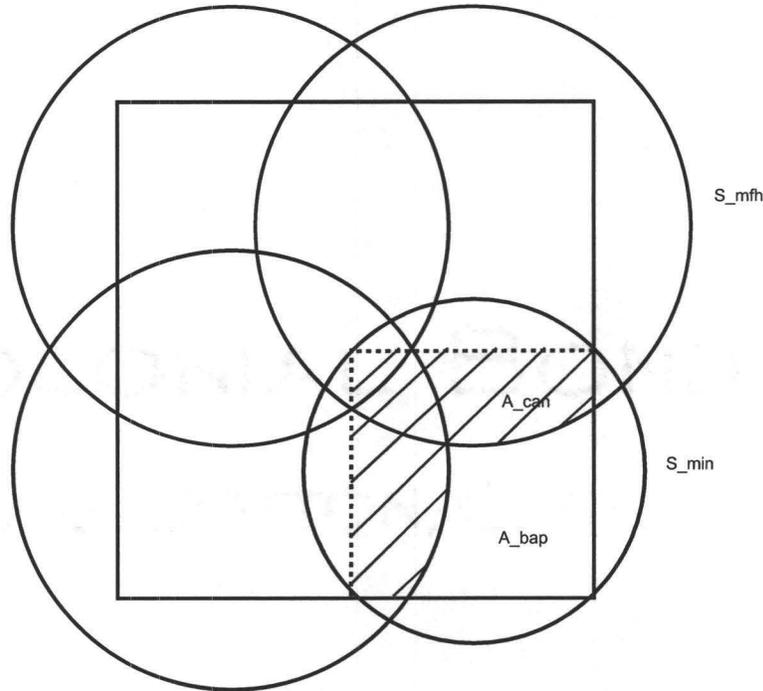


Figure 5.4: *The WLAN Deployment Showing the BAP Coverage Region and the MFH Region for Other APs. The Shaded Area Indicates the Valid MFH Region.*

equally distributed around that mean, the coverage area can be approximated as a circle.

The goal is to find the conditional probability of a call being a valid MFH-candidate, given that it belongs to the coverage area of the BAP. Considering the uniform user distribution throughout the WLAN coverage area, one fourth ($1/N_{ap}$) of the calls will be associated to the BAP on average, hence the effective coverage area for the BAP is the dotted small square, with an area of $A_{bap} = A_{WLAN}/N_{ap}$. This takes into account the fact that calls inside the coverage overlap regions are equally likely to belong to the two overlapping APs. The same reasoning can be used to find the valid MFH-region which is the intersection of the coverage area for BAP and the MFH-regions for all other neighbour APs. The effective area for this region is shaded

in Figure 5.4 and its area is denoted by A_{can} , which is calculated as,

$$A_{can} = r_{mfh}^2 [3\pi/4 - 1 - \arccos(h/r_{mfh}) - \arccos(\sqrt{2}h/r_{mfh})] + h[\sqrt{r_{mfh}^2 - h^2} + \sqrt{2r_{mfh}^2 - 4h^2}] - 3h^2, \quad (5.16)$$

where h is the distance of any AP from the closest deployment boundary and r_{mfh} is the radius of the MFH-circle corresponding to an RSSI of S_{mfh} . Note that although this derivation is specific to the class of WLAN deployments shown in Figure 5.4, the same concept and methodology applies to any given deployment. Given A_{can} , one can easily compute the probability, P_{can} , of a call being a valid MFH-call as the ratio A_{can}/A_{bap} . However, this probability is actually lower due to returning MFH-calls before a VHO is completed. This happens if an MFH-call within the shaded region in Figure 5.4 is able to move into the unshaded region belonging to the BAP coverage area in a time less than T_{vho} . Figure 5.5 is used to model the effect of these MFH-calls which shows a given MFH-call, M , that can potentially return to the BAP. Observe that this MS is travelling at a constant speed, v , along a direction, θ , which is uniformly distributed in the interval $[0, 2\pi)$. The mobility extent of such a call during an ongoing VHO can be represented by a circle with the radius $L = vT_{vho}$. Note that it is implicitly assumed that such a call is unlikely to change its course during T_{vho} which is a short period of time compared to the mean travelling time of the user (60 seconds for the simulations). Hence for the MFH-call, M , to return to the BAP during T_{vho} , its mobility circle should intersect with the MFH-region boundary as shown in Figure 5.5. Moreover, let the angle of the arc created as a result of this intersection be α . Then the probability that an MFH-call, M , returns to the BAP before the end of the current VHO is $\alpha/2\pi$. Let x be the shortest distance from user, M , to the MFH-region boundary, then it follows that the probability of user, M , returning to the BAP equals $\alpha(x)/2\pi$ and can be expressed as,

$$P_{ret}(x) = \frac{2(r_{mfh} - x)}{\pi} \arcsin(h/r_{mfh}) \arccos\left(\frac{2r_{mfh}x - v^2T_{vho}^2 - x^2}{2vT_{vho}(r_{mfh} - x)}\right). \quad (5.17)$$

Define, A_{ret} , as the effective area corresponding to all MFH-calls which may return to the BAP during T_{vho} by integrating $P_{ret}(\cdot)$ over the shaded region shown in Figure

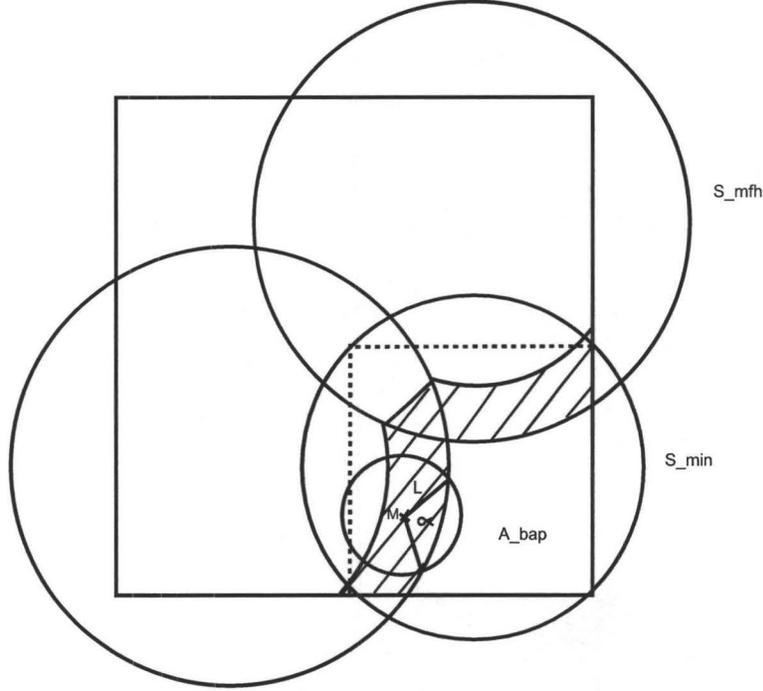


Figure 5.5: The Shaded Area is Where Valid MFH-Users May Return to the BAP After Being Forced to Perform a HHO to a Neighbouring AP.

5.5. This region contains all MFH-calls that may return to the BAP. The resulting effective MFH-return area is,

$$A_{ret} = 2 \int_{x=0}^{vT_{vho}} \int_{\phi=\phi_1}^{\phi_2} P_{ret}(x) dx d\phi, \quad (5.18)$$

where ϕ_1 and ϕ_2 are the angles corresponding to the extreme directions which result in an MFH-call to return to the BAP. Hence, P_{can} is expressed as,

$$P_{can} = \frac{A_{can} - A_{ret}}{A_{bap}}. \quad (5.19)$$

Next, the pdf of N_{mfh} will be derived. There are three possible cases in terms of what N_{mfh} will be. The first is the *no-deficit* case where there is already enough free bandwidth available at the BAP to secure the G_{vho} reservation. From Equation 5.11 it follows that the probability of this case is,

$$P_{NDF} = F_{bap}(C - N_G), \quad (5.20)$$

where, F_{bap} , is the BAP state cdf. In this case, N_{mfh} , will be zero and the number of calls that remain in the BAP after VHO initiation will clearly still be N_{bap} . The second case is when the number of MFH-calls are limited by the requirement, N_{req} . In this case there are plenty of MFH-candidates to satisfy the reservation, thus at the end of the MFH procedure, there will be exactly N_G free call capacity available at the BAP, and $N_{mfh} = N_{req}$. Note that $N_{req} = 0$ also results in N_G being zero, however, this is included as part of the no-deficit case. Hence, the probability of the exact case can be derived as,

$$P_{EXC} = Pr\{N_{req} \leq N_{can} \cap N_{req} \geq 1\}. \quad (5.21)$$

The *candidate-constrained* case is when the number of successful MFH-calls, N_{mfh} , is limited by the number of MFH-candidates rather than the actual requirement. This can happen as a result of a poor choice of S_{mfh} leading to small overlap between the coverage region of the BAP and the MFH-region of the adjacent APs. The probability of such a case is found to be,

$$P_{CDC} = Pr\{N_{req} > 0 \cap N_{can} \leq N_{req} - 1\}. \quad (5.22)$$

For this case N_{mfh} will be equal to N_{can} . This probability can be expressed in more detail as follows,

$$P_{CDC} = \sum_{k=0}^{N_g-1} f_{can|bap}(k|C - N_g + k + 1) f_{bap}(C - N_g + k + 1). \quad (5.23)$$

Alternatively, for the exact case one can write,

$$P_{EXC} = 1 - P_{NDF} - P_{CDC}. \quad (5.24)$$

In order to calculate VHO capacity-dropping from Equation 4.19, one needs the pdf for the free call capacity at the BAP after the MFH procedure, \widehat{N}_{bap}^f , which is in general expressed as,

$$\widehat{N}_{bap}^f = C - N_{bap} + N_{mfh}. \quad (5.25)$$

Hence, according to the previously mentioned cases the free capacity can be either less than, greater than, or exactly equal to the reservation amount and it follows that,

$$\widehat{N}_{bap}^f = \begin{cases} N_G - m, & (1 \leq m \leq N_G) & \text{Candidate Constrained Case} \\ N_G + m, & (1 \leq m \leq C - N_G) & \text{No Deficit Case } (N_{req} < 0) \\ N_G, & & \text{Exact Case and } (N_{req} = 0). \end{cases} \quad (5.26)$$

Substituting these conditions with their probabilities, the pdf of free call capacity at the BAP can be obtained as,

$$\widehat{N}_{bap}^f = \begin{cases} N_G - m & \sum_{k=0}^{N_G-m} f_{can|bap}(k|C - N_G + k + m)f_{bap}(C - N_G + k + m) \\ N_G + m & f_{bap}(C - N_G - m) \\ N_G & 1 - F_{bap}(C - N_G - 1) - \\ & \sum_{k=0}^{N_G-1} f_{can|bap}(k|C - N_G + k + 1)f_{bap}(C - N_G + k + 1). \end{cases} \quad (5.27)$$

Equation 5.27 can be used with Equation 4.19 to derive the VHO capacity-dropping rate.

Figure 5.6 compares the analytical and simulation results for VHO capacity-dropping, $P_{cd}^{transient}$, and new call blocking, B_{TR} , for TR-MFH derived from Equations 4.19 and 5.5, respectively. The set of parameters used for this scenario are those listed in Table 4.2. According to Figure 5.6, the analytical results closely follow the simulation results. Unlike the SR scheme, the analytical new call blocking rate overestimates the simulations while analytical VHO capacity-dropping is still an underestimation at some points.

As mentioned in the previous chapter, the difference between the analytical and simulation results for the first point on the curves is due to the error of the analytical model in capturing the load on the boundary AP. Note that the first point on the curves corresponds to no bandwidth reservation. Vertical handoff capacity-dropping is also estimated to be larger than that obtained from simulation, since the analytical model estimates a load slightly higher than its actual value for the BAP. This results in less available bandwidth in the BAP according to the analytical model.

As the total amount of requested reservation becomes larger, the MFH procedure is more likely to fail in securing *all* of the bandwidth, thus, making the blocking rate from simulations decrease below the analytical estimate. Note that for calculating

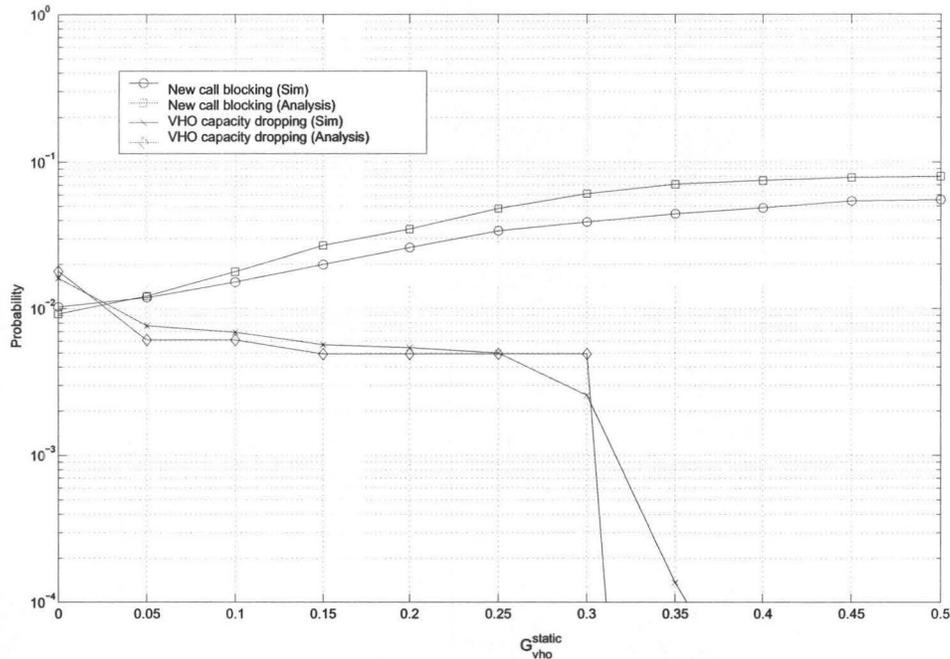


Figure 5.6: Comparison Between the Analytical and Simulation Results for the TR-MFH Scheme.

the analytical new call blocking rate it is assumed that all the requested bandwidth will be successfully reserved. However, this assumption becomes less appropriate with the increase in VHO overlap probability as well as with higher values of G_{vho} . Therefore, as suggested by Figure 5.6, the analytical blocking rate becomes larger than the simulation blocking rate as G_{vho} increases.

With regard to VHO capacity-dropping, the analytical results are below the simulations for small fractions of bandwidth reservation. Similar to the SR case illustrated in Figure 4.5, this occurs due to overlapping VHO events in the simulations, whereas in the analysis, it is assumed that vertical handoffs are rare events. This causes more bandwidth deficit at the BAP and hence capacity-dropping will be slightly higher in the simulations. However, unlike SR, overlapping VHOs will result in more bandwidth reservation. This increase becomes significant for larger values of $G_{vho}^{transient}$ and causes the simulation-based capacity-dropping to become closer to the analytical

estimate. This behaviour is clearly seen in Figure 5.6.

One of the assumptions in developing this model is that *capacity-constrained* cases in which the number of MFH-calls is limited to the free capacity of the neighbour APs, are assumed to never occur. According to Figure 5.6, the comparison of VHO capacity-dropping obtained from this model and simulations suggests that such an assumption is justified at least for the set of parameters used. However, it should be noted that for very large values of G_{vho} or when the effective G_{vho} is large due to frequent overlapping VHOs, the required amount of MFH calls will be high and may not be accommodated by the adjacent APs. However, such cases are not practical because a relatively moderate bandwidth reservation of less than 35% is generally sufficient to eliminate VHO capacity-dropping. In addition, VHO events occur much less frequently than new call arrivals in any reasonable WLAN deployment, hence the effective G_{vho} would also remain close to the per-VHO G_{vho} .

An important parameter which provides insight into the operation of the TR-MFH scheme is the number of successful MFH-calls, N_{mfh} , expressed previously in Equation 5.13. Next, the pdf, mean, and the standard deviation of N_{mfh} will be calculated. N_{mfh} is non-negative and always less than N_g . It will take the value of zero when there is no capacity deficit or that there are no MFH-candidates found. The probability of the latter event is small compared to the former, and can be safely ignored. Hence, with probability $P_{NoDeficit}$, N_{mfh} will be zero. The case where $N_{mfh} = N_g$ is also interesting since it corresponds to the BAP having no free call capacity, thus requiring N_g calls to be successfully forced out. This is a worst case situation for the system, however. For all the other values of N_{mfh} , there are two possible causes: either it is the number of required MFH-calls, N_{req} , and there are sufficient number of candidates ($N_{can} \geq N_{req}$), or it is the number of candidates, N_{can} , which is less than the requirement enforced by the MFH procedure ($N_{can} < N_{req}$). It

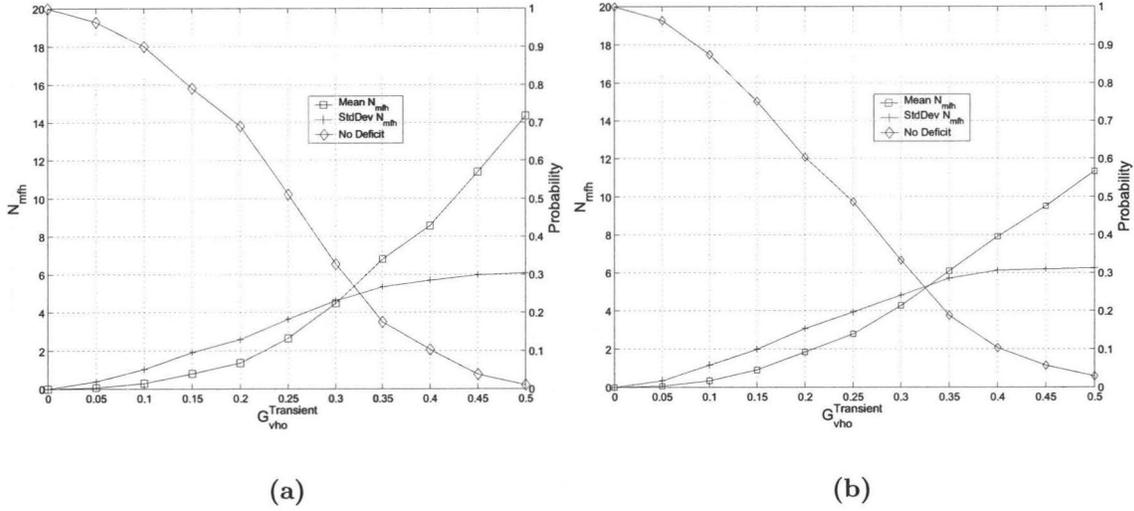


Figure 5.7: Analytical (a), and Simulation (b) Results for Mean and Standard Deviation of the Eventual Number of MFH-Calls, N_{mfh} , vs. the Fraction of Transient Bandwidth Reservation at the BAP. Also Included on the Right Side Axis is the Probability of the No-Deficit Case.

follows that the pdf for N_{mfh} can be calculated as,

$$f_{mfh}(k) = \begin{cases} P_{NoDeficit} & k = 0 \\ (1 - F_{can|bap}(N_g - 1, C)) f_{bap}(C) & k = N_g \\ (1 - F_{can|bap}(k - 1, C + k - N_g)) \times \\ f_{bap}(C + k - N_g) + & 0 < k < N_g \\ \sum_{m=C+k-N_g+1}^C f_{can|bap}(k, m) f_{bap}(m) & . \end{cases} \quad (5.28)$$

Figure 5.7(a) plots the mean and standard deviation for N_{mfh} which is derived from Equation 5.28 for the setup evaluated in Figure 5.6. This is also compared with the same results obtained using simulations as shown in Figure 5.7(b). The maximum error in estimating the simulation results is 15%. However, as the reservation fraction increases beyond 0.4, the simulation based value of $E[N_{mfh}]$ becomes lower than the one derived using Equation 5.28. This is due to capacity-constrained cases happening more frequently as more calls have to be switched to adjacent APs. It was previously mentioned that the effect of capacity-constrained situations will be neglected in favour

of obtaining a simple model. It is worth mentioning that although the actual value of N_{mfh} may deviate from the analytical value, this happens at large bandwidth reservation fractions for which the capacity-dropping is almost zero. Therefore, as illustrated in Figure 5.6, the model provides a good estimation of the performance metrics which are of interest.

Figure 5.7 also shows the likelihood of having no capacity deficit at the BAP, $P_{NoDeficit}$, for different values of G_{vho} . It can be observed that as the transient bandwidth reservation fraction increases from very low to moderate values, $E[N_{mfh}]$ increases more steeply. This is explained by noting that the percentage of No-Deficit cases declines sharply as G_{vho} becomes larger. Thus when G_{vho} is about 0.25, the majority of VHO initiations will result in the MFH procedure being invoked. After this point, $P_{NoDeficit}$ has no significant effect and $E[N_{mfh}]$ will increase linearly with G_{vho} due to the linear increase in the amount of capacity deficit.

5.3 Effect of Vertical Handoff Latency and Link-Dropping Rate

One of the factors influencing the use of bandwidth reservation to overcome the capacity deficit problem is the vertical handoff latency, T_{vho} . Vertical handoff latency determines the lowest bitrate that will be used by a call and the duration of using it. In this section the effect of increasing T_{vho} on call level performance (i.e., new call blocking and VHO dropping rates) at the boundary AP will be considered. According to Figure 5.8, increasing T_{vho} to seven seconds will result in more VHO calls being dropped due to capacity shortage when no VHO improvement scheme is implemented. This case corresponds to the leftmost point on the curve having zero guard bandwidth. For example, according to Figure 5.8, VHO capacity-dropping with no reservation is about 3.7% compared to 1.6% for $T_{vho} = 4sec$ (Figure 5.3). The increase in VHO capacity-dropping can be explained by noting that the longer it takes to perform a VHO, the more likely a MS is to experience lower link RSSI and hence to shift down to lower bit rates. In addition, the time during which such a MS is operating at low

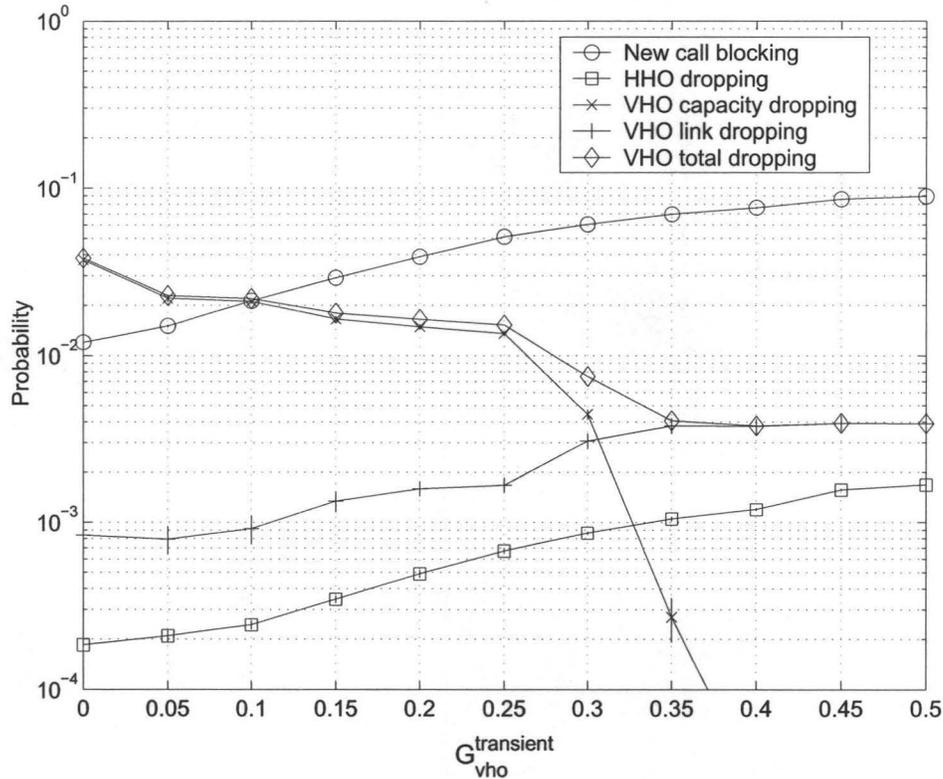


Figure 5.8: Boundary AP New Call Blocking, HHO Dropping and VHO Dropping vs. Transient Guard Bandwidth Reservation Using Momentary Forced HHO (TR-MFH) When VHO Latency is 7 Seconds.

bitrates is extended. Consequently, a VHO call requires more bandwidth from the BAP and for a longer period as T_{vho} increases resulting in a higher chance of VHO capacity-dropping.

Figure 5.9 plots the probability of the minimum bitrate a VHO call experiences for different values of T_{vho} . This figure is generated by running a simulation for the RSSI of a single user as it moves away from the BAP and is translated into bit rates using Table 4.1. Clearly, lower bit rates are experienced less frequently by VHO calls, however, they require a larger amount of bandwidth. For example, a link operating at 1 or 2Mbps requires almost 33% or 17% of the AP bandwidth. Increasing the VHO latency from four to seven seconds causes P_{1Mbps} and P_{2Mbps} to go from 0.6% and

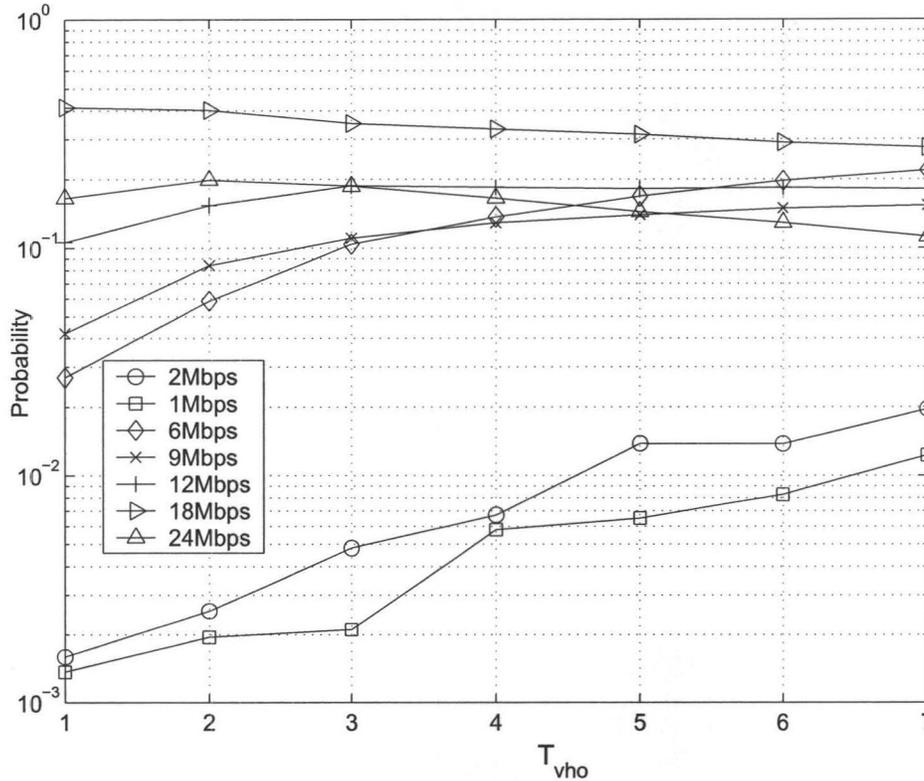


Figure 5.9: Probability of Minimum Bitrate During a VHO for Various Values of VHO Latency.

0.7% to 1.2% and 2%, respectively, which in turn results in the capacity-dropping to increase from 1.6% to 3.7% when no reservation scheme is employed. To get almost zero capacity-dropping, one needs to reserve 40% of the bandwidth upon a VHO initiation when $T_{vho} = 7sec$. Note that according to Figure 5.3, the same amount of capacity-dropping was reached with 35% bandwidth reservation.

As T_{vho} becomes larger, VHO link-dropping also starts to increase, putting a lower bound on the VHO total dropping rate. Clearly, link-dropping increases since the MS is more likely to lose the link as a vertical handoff procedure takes longer to complete. Total VHO dropping will eventually be dominated by link-dropping. For this set of parameter values, and a T_{vho} of four seconds, VHO link-dropping is almost

negligible and less than 0.1% (Figure 5.3). However, as shown in Figure 5.8, when T_{vho} increases to seven seconds, VHO link-dropping increases to 0.4%. Consequently, reducing VHO capacity-dropping to much less than 0.4% in the latter case will not significantly affect the total dropping rate. Nevertheless, a total dropping rate of 0.4% is still a significant improvement considering its original value of 3.7%.

Link-dropping is directly affected by the system configuration such as the VHO algorithm used and its latency, the WLAN deployment, user mobility pattern, receiver sensitivity and wireless propagation. It is expected that a deployment is designed such that link-dropping is within reasonable target bounds provisioned by the network administrator. Obviously, if a given deployment has an unacceptable link-dropping rate then there is no point in trying to decrease VHO capacity-dropping.

In addition, VHO link-dropping and capacity-dropping are correlated. If link-dropping is very low then capacity-dropping will also be low. Therefore, there will be less motivation for a scheme that improves VHO capacity-dropping. An extreme example of such a case is when a VHO is performed almost instantaneously. Also if link-dropping is high, then capacity-dropping can potentially be high. From this argument it can be concluded that for scenarios where link-dropping is high but still reasonable (e.g., < 1-2%), the potentially large amount VHO capacity-dropping can be effectively reduced using the TR-MFH approach. One such case will be addressed in Section 5.7.

5.4 Effect of Boundary AP Load

Another factor influencing the amount of VHO capacity-dropping is the load on the boundary AP. The AP load determines the amount of capacity deficit experienced by VHOs. As the BAP becomes more populated with calls, the likelihood of a VHO call experiencing bandwidth deficit increases. The load of an AP is affected by the total call arrival rate, which is both a function of user population density and per user call arrival rate. Figure 5.10 shows the call level performance of the BAP when no reservation scheme is used for different values of per AP new call arrival rate.

The new call blocking rate can be taken as an indicator for AP load. Even at

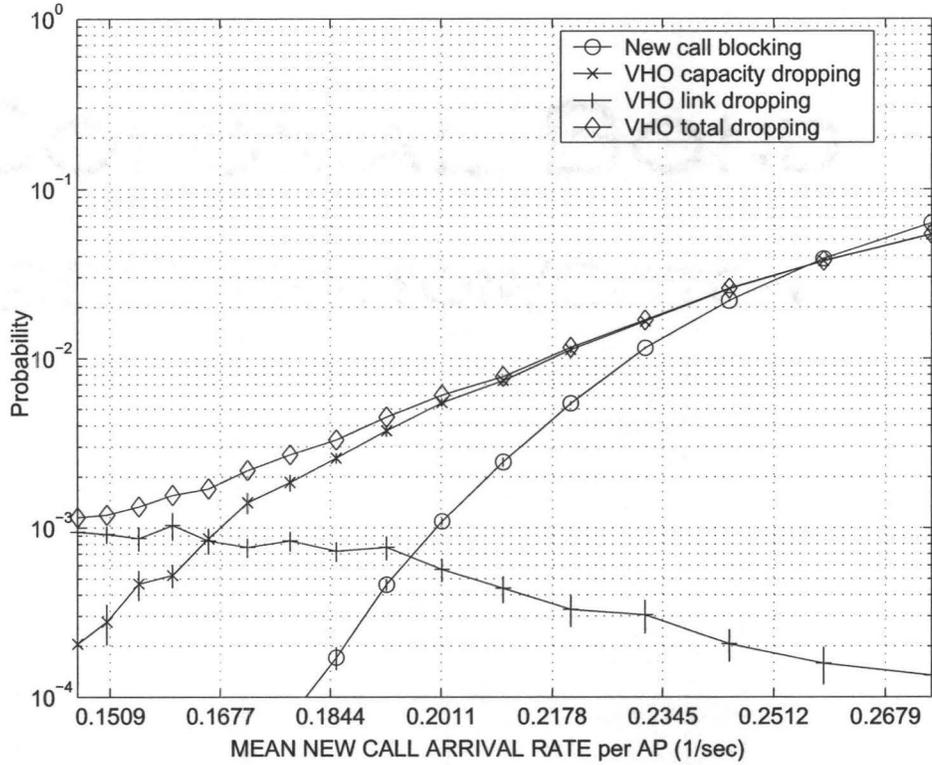


Figure 5.10: VHO Capacity-Dropping and New Call Blocking Rates for the BAP vs. per AP New Call Arrival Rate With the Other Parameters as in Table 4.2, When No Reservation Scheme is Employed.

small arrival rates where the new call blocking probability is extremely low, VHO capacity-dropping is still significantly higher than the new call blocking rate. For example, given a new call arrival rate of $\lambda_{new} = 0.1844/sec$ for each AP, the new call blocking rate is almost 0.02% while VHO capacity-dropping is at 0.25%. This means that VHO calls will be dropped almost ten times as often as new calls are blocked. According to Figure 5.10, capacity-dropping decreases at a much lower rate than new call blocking. This is attributed to the fact that VHO calls require a larger amount of bandwidth compared to a new call.

The results in Figure 5.10 suggest possible VHO dropping reduction by using a bandwidth reservation scheme even at a low AP load. However, one should be

aware of the lower bound on VHO total dropping which is due to link-dropping as mentioned in the previous section. Therefore, when the BAP load is below a certain threshold, the improvements in total VHO dropping become marginal. For example, Figure 5.10 suggests that for $\lambda_{new} < 0.15/sec$, VHO capacity-dropping will be much less than the 0.09% link-dropping, hence there is no advantage in employing any bandwidth reservation scheme.

In contrast, as the BAP load increases there will be an increase in VHO capacity-dropping, which calls for employing TR-MFH. This scheme in turn invokes the MFH procedure. The effectiveness of TR-MFH, as mentioned earlier, is dependent on the probability of the following three cases: candidate-constrained, no-deficit and exact (Section 5.2). Clearly, as the load on the access points including the BAP becomes higher, the likelihood of a no-deficit case declines. However, due to the larger call population at the BAP, the probability of candidate-constrained cases also decreases thus balancing the increase in capacity-dropping for a given amount of bandwidth reservation. On the other hand, as the load also increases in adjacent APs, the possibility of having a *Capacity-Constrained* case increases. In other words, it becomes more likely that MFH-candidates are not accepted by neighbour APs due to a lack of bandwidth. Therefore, the assumption previously made about having a negligible number of capacity constrained cases may no longer hold. This will reduce the effective bandwidth reservation at the BAP for the TR-MFH scheme and will eventually result in less reduction in VHO capacity-dropping.

5.5 Effect of Vertical Handoff Rate

VHO events happen occasionally compared to new call arrivals. In Section 5.2 an analytical model for TR-MFH was derived by assuming that VHO arrivals are rare events. This assumption was later justified by showing that simulation and analytical results closely agree. TR-MFH takes advantage of the rarity of vertical handoffs and provides sufficiently low VHO capacity-dropping while maintaining a new call blocking rate which is significantly lower than that of SR. In this section, an example will be investigated for which such an assumption is not valid and VHO events happen

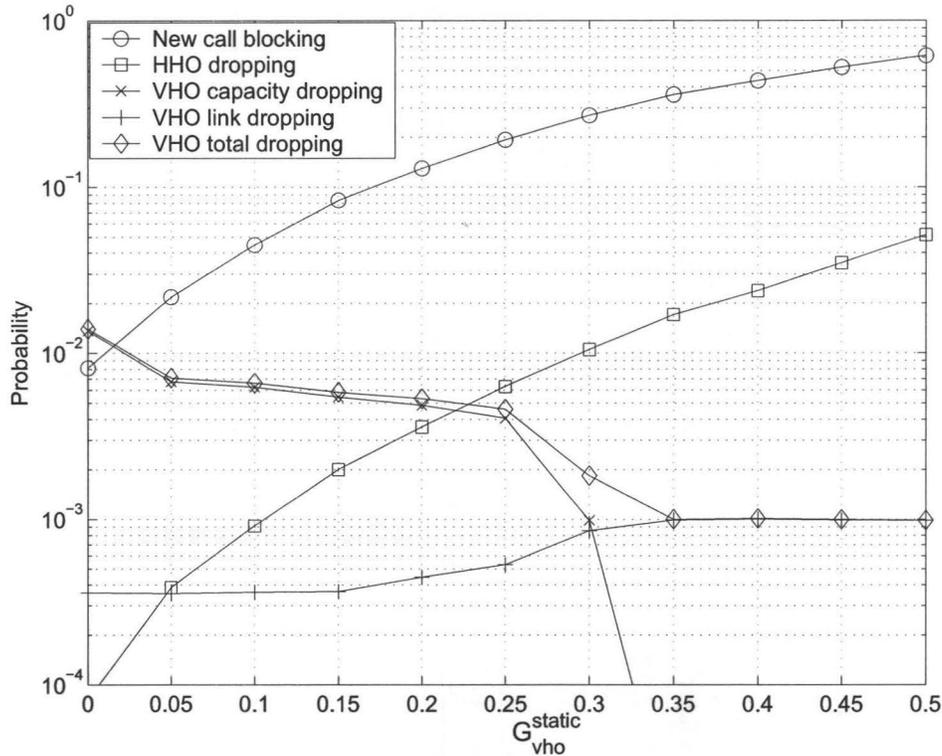


Figure 5.11: Performance of the SR Scheme Under a high VHO Arrival Rate, $\lambda_{vho} = 0.168/sec$.

more frequently. TR-MFH reserves a certain fraction, G_{vho} , of BAP's total bandwidth per VHO initiation. Therefore, as VHO events become more likely to overlap, it is reasonable to expect that more bandwidth would be reserved. In addition, this bandwidth would be reserved more frequently. Therefore, it is desirable to investigate the performance of TR-MFH under such circumstances and compare it to that of SR.

First consider the effect of increasing the VHO arrival rate on SR performance. This is presented in Figure 5.11 showing the performance of the BAP in terms of new call blocking rate and VHO dropping rate when the VHO arrival rate is approximately 0.16 calls/sec, which is almost ten times larger than the original value of 0.016/sec considered in Figure 4.2. Interestingly, Figures 5.11 and 4.2 show that the performance of SR is almost unaffected by the VHO arrival rate. The slight decrease

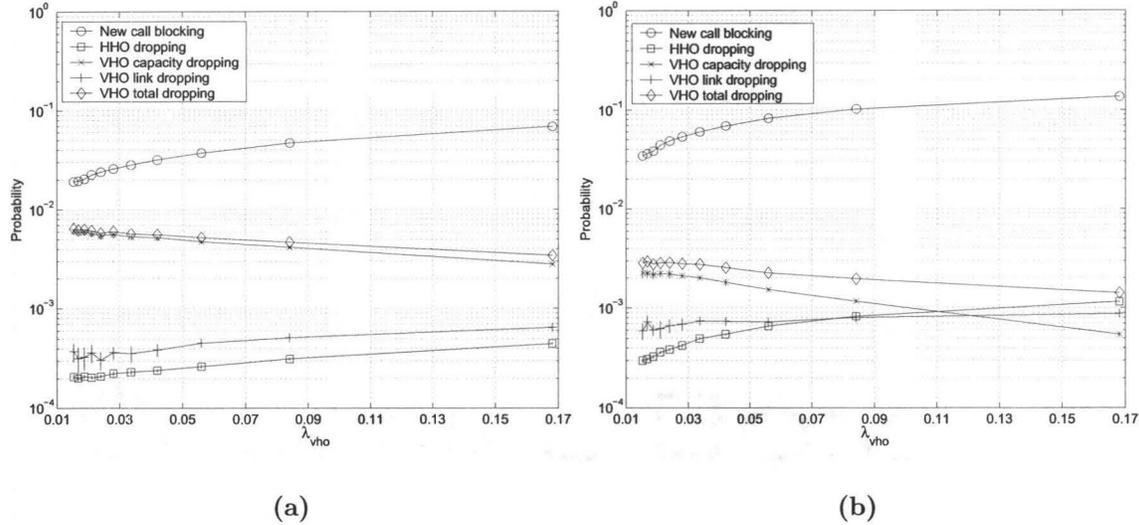


Figure 5.12: Performance Evaluation of the Transient Bandwidth Reservation Approach for Increasing Values of λ_{vho} , When the Reservation Fraction is (a) 15% and (b) 30%.

in new call blocking from 1% (Figure 4.2) to 0.8% and in VHO capacity-dropping from 1.6% to 1.4% is due to the reduction in the BAP's load as a result of more calls exiting the WLAN. Similar to the results in Figure 4.2, nearly 35% bandwidth reservation is sufficient to reduce VHO capacity-dropping to almost zero. This is at the cost of increasing new call blocking to 36% and HHO dropping to 1.7%. These values are roughly equal to those for the system with a lower VHO arrival rate studied earlier in Section 4.4.

Figure 5.12 presents the simulation results for TR-MFH over a wide range of VHO arrival rates. The range of λ_{vho} corresponds to having users spend on average between 150 to 17 minutes within the designated WLAN deployment area. These values are only examples of the amount of time a user may spend in a given WLAN deployment and actual values will be specific to the deployment size, user mobility pattern and time of day. Figures 5.12(a) and 5.12(b) show the performance of the TR-MFH scheme when 15% and 30%, respectively, of the BAP's bandwidth is reserved upon a VHO being triggered. Unlike SR, TR-MFH is significantly affected by the rate of

VHO arrivals. In fact, for $G_{vho} = 0.15$, the new call blocking rate increases from 2% to 7% as λ_{vho} rises from 0.016/sec to 0.168/sec. By increasing the VHO arrival rate, bandwidth reservations will be performed more frequently and they can overlap. This is also verified by Equation 5.3 which clearly shows that new call blocking is directly related to the probability of having an ongoing VHO. More importantly, overlapping VHO events cause overlapping bandwidth reservations which result in a larger amount of bandwidth being allocated to vertical handoffs. This further decreases VHO capacity-dropping as clearly shown in Figure 5.12(a). For instance, VHO capacity-dropping is cut in half when the VHO arrival rate spans the range on the X-axis of Figure 5.12(a). Interestingly, this suggests that lower amounts of bandwidth reservation will suffice to achieve the same capacity-dropping as the VHO arrival rate becomes higher. This is an advantage of TR-MFH over SR. Increasing the bandwidth reservation would only make things better in terms of capacity-dropping. As Figure 5.12(b) shows with $G_{vho} = 0.3$, capacity-dropping will be reduced by more than a factor of four from 0.22% to 0.05% when λ_{vho} changes from 0.016/sec to 0.168/sec. However, this comes at the price of an increase in new call blocking, which changes from 3.4% to 13.7%.

The increase in new call blocking rate may in theory cause SR to be a better option. Nevertheless, the additional reduction in capacity-dropping using TR-MFH should not be neglected. When the VHO arrival rate is sufficiently high, one can achieve the same capacity-dropping by reserving less bandwidth under TR-MFH. Moreover, as Figure 5.12 suggests, the new call blocking curve for TR-MFH becomes flatter as the VHO arrival rate increases. This is because once there are many overlapping VHOs, the instantaneous amount of reservation will be limited to the maximum capacity of the BAP and will not increase with λ_{vho} . One can also see from inspecting Figures 5.12 and 5.11 that it is less likely for the new call blocking rate of TR-MFH to be larger than that of SR as G_{vho} increases. For example, with $G_{vho} = 0.15$, SR exhibits a new call blocking rate of 8% which is close to the maximum new call blocking rate of 7% achieved by the same amount of G_{vho} under TR-MFH. However, for $G_{vho} = 0.30$, SR results in a 27% blocking rate which is far more than the 13.7% maximum blocking rate achieved by TR-MFH.

It is worth mentioning that the analytical model developed for TR-MFH does not provide a close estimate when λ_{vho} is large due to two reasons. First, it assumes that all the G_{vho} bandwidth is secured, which is not necessarily true especially when the total reservation demand becomes large. This results in a higher analytical new call blocking compared to the simulation results. In addition, not taking into account simultaneous bandwidth reservations due to overlapping VHOs, causes the analytical capacity-dropping to be fixed for different values of λ_{vho} and exceeds the simulation results as λ_{vho} and G_{vho} become larger. By taking into account the effect of candidate-constrained cases and the pdf for N_{mfh} provided in Equation 5.28, as well as including the possibility of having two overlapping VHOs, it may be possible to refine the analytical model. It should be noted, however, that the lower bound on B_{TR} presented by Inequality 5.8 is valid for such cases.

In summary, when the rate of vertical handoffs out of the WLAN becomes larger, TR-MFH reduces VHO capacity-dropping, however, at the expense of a larger new call blocking rate. In addition, the increased demand of TR-MFH for bandwidth reservation in such circumstances, means that the potential effectiveness of this approach would be limited by the candidate-constrained cases, which is the subject of discussion in Section 5.6.

5.6 Effect of MFH Signal Level Threshold

The performance of the transient reservation approach relies on the effectiveness of the momentary forced HHO (MFH) mechanism. The MFH procedure is in turn influenced by the value of S_{mfh} . As more coverage overlap exists between WLAN access points, higher values of S_{mfh} will be sufficient for guaranteeing that enough MFH-calls are accepted by neighbour APs. Interestingly, dense deployments, which are the main focus of this work, usually have a higher coverage overlap.

As more bandwidth reservation needs to be made, the MFH procedure is more likely to become candidate-constrained. Figure 5.13(a) plots for different values of S_{mfh} , the probability of the MFH procedure running into a candidate-constrained case, P_{CDC} , as a function of the total bandwidth reservation demand. Note that the

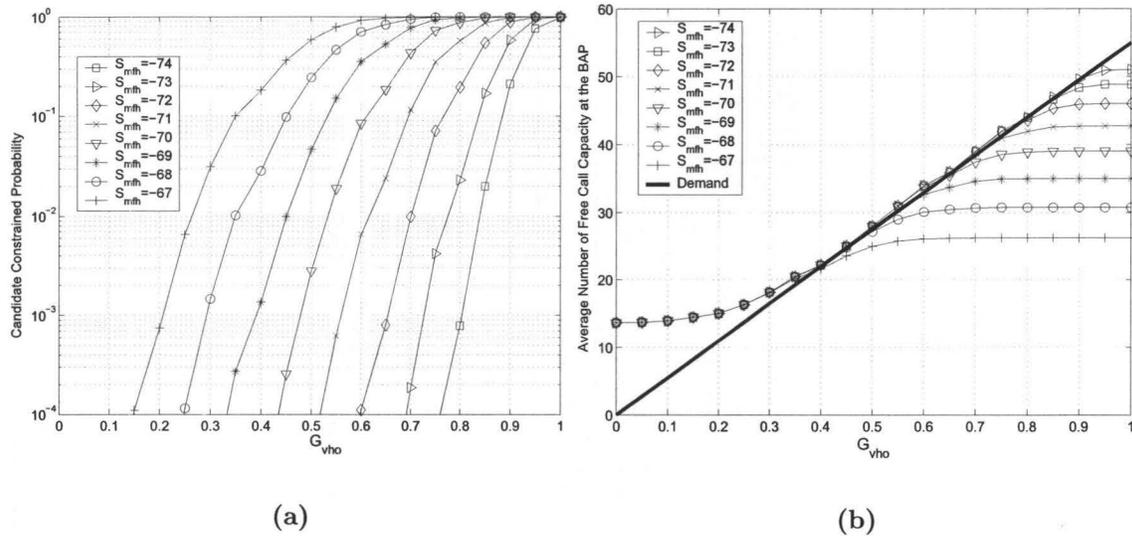


Figure 5.13: The Effect of S_{mfh} (in dBm) on the Likelihood of Having a Candidate-Constrained Case (a) and the Average Free Call Capacity at the BAP After Running the MFH Procedure (b).

total reservation demand can be higher than the per VHO reservation if VHO events overlap.

According to Figure 5.13(a), as S_{mfh} becomes larger and the MFH region is reduced, the probability of becoming candidate-constrained for a certain value of G_{vho} quickly increases. For instance, when half of the bandwidth is to be reserved per VHO, going from $S_{mfh} = -67\text{dBm}$ to -70dBm will cause P_{CDC} to decline from 60% to 0.3%. The optimal value for S_{mfh} depends on the desired target VHO capacity-dropping. For example, if 2% of the VHO calls use the lowest bitrate of 1Mbps and they need roughly 35% of the BAP bandwidth to successfully finish at that bitrate, then to achieve a target capacity-dropping of, 0.1%, S_{mfh} should be chosen such that $2\% \times P_{CDC} < 0.1\%$. According to Figure 5.13(a), this will require S_{mfh} to be about -68dBm.

Intuitively, one should make sure that the amount of reservation needed to take

care of the worst-case capacity-dropping is not significantly affected by the candidate-constrained cases and the target capacity-dropping is achievable. This ensures that the MFH region will be large enough so that sufficient calls are forced out to adjacent APs and there will be room at the BAP for securing the reservation. Figure 5.13(b) shows the analytical average free call-capacity at the BAP after applying the MFH procedure for different values of S_{mfh} , versus the reservation demand. A thick straight line denoted by *Demand* shows the bandwidth reservation demand in number of calls. If the MFH procedure is able to provide the BAP with at least *Demand* call capacity, then the VHO call will finish without being dropped.

According to Figure 5.13(b) and for the range of S_{mfh} shown, as long as the total bandwidth demand from all concurrent VHO calls is less than 40%, there will be enough free capacity at the BAP on average and VHO calls will not be dropped. However, for $G_{vho} > 40\%$ the average free bandwidth at the BAP will start to saturate at lower values of bandwidth reservation demand. For instance, when S_{mfh} changes between -67dBm and -74dBm, the average free capacity at the BAP will saturate somewhere between 26 and 51 calls, respectively.

Once the average free bandwidth saturates, there is no point in further increasing G_{vho} . An effective rule of thumb for determining S_{mfh} is that the free bandwidth at the BAP should not be below the demand. The bandwidth demand is in turn determined by the probability of overlapping VHOs, the minimum bitrate that they experience, and the amount of bandwidth requirement at that bitrate.

Controlling the value of S_{mfh} in a deployment is important since setting it too low will cause calls with lower bitrates to be associated to the adjacent APs when they have a much higher bitrate at the BAP. This means that more bandwidth is been taken away from the adjacent AP than is been freed from the BAP. Also it is better not to allow calls that are deep inside the BAP coverage area to associate to the adjacent APs for interference reasons. Therefore it is important to choose the proper value for S_{mfh} .

Many deployments do not have the same AP layout and propagation exponent that is used in this model, therefore, S_{mfh} cannot be used as a reliable control parameter for every situation. For example, if the BAP is only adjacent to one other

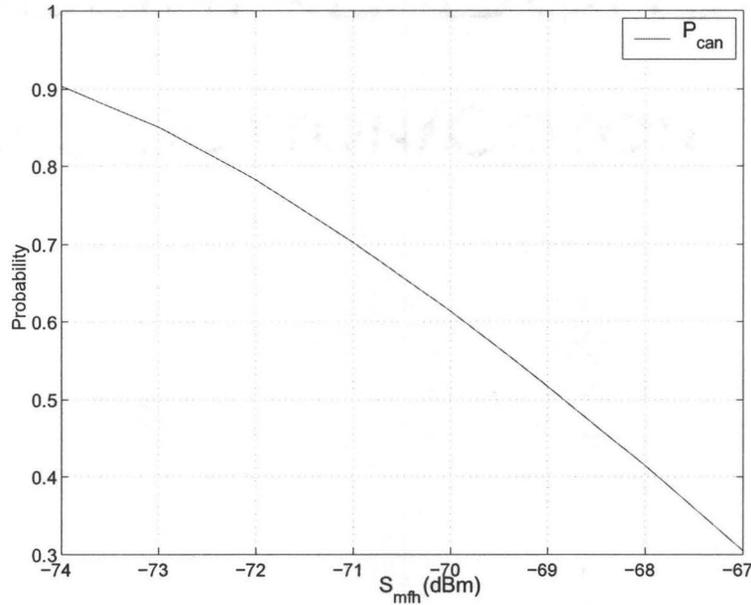


Figure 5.14: Probability That a Given Call Associated With the BAP is a Valid MFH Candidate for Different Values of S_{mfh} .

AP then the MFH region will be smaller. The major parameter which affects P_{CDC} is the probability that a given call associated to the BAP is a valid MFH-call. This probability is denoted by, P_{can} , and is plotted for the aforementioned deployment scenario versus S_{mfh} in Figure 5.14. Note that this is a strong function of S_{mfh} and is also related to the structure of the deployment.

5.7 Effect of Wireless Propagation Parameters

The propagation parameters affect vertical handoff performance in two different ways. Firstly, outdoor propagation parameters determine the likelihood of losing the WLAN link during a VHO. Larger path loss exponents and more shadowing result in a higher chance of a VHO call being dropped due to losing its link. In the same way, outdoor propagation parameters also influence the probability of a VHO call experiencing different bitrates. This in turn affects the bandwidth requirement of a VHO call,

Table 5.1: 6Mbps WLAN Deployment Parameters

Parameter	Value	Description
S_{min}	-80dBm	Minimum RSSI across the WLAN deployment
S_{tp}	-85dBm	VHO triggering threshold
R_{min}	6Mbps	WLAN clipping rate
n_{in}	3.0	Indoor path loss exponent
σ_{in}	5.0dB	Indoor lognormal shadowing std. dev.
n_{out}	3.5	Outdoor path loss exponent
σ_{out}	5.0dB	Outdoor lognormal shadowing std. dev.
$1/\lambda_{new}$	2080sec	Average time between call requests for a user

which translates directly into VHO capacity-dropping. Consequently, the higher the link-dropping rate, the higher will be the chance of VHO capacity deficit. This was previously concluded in Section 5.3.

Propagation parameters can also affect VHO capacity-dropping by influencing the range of an AP and hence its load. As mentioned earlier in Section 5.4, a higher load on the WLAN access points leads to more bandwidth deficit for vertical handoffs. For the capacity-dropping to be high, the load of an AP should be such that a significant fraction of the VHO calls experience a bandwidth deficit.

In this section, a different deployment with larger propagation parameters has been considered which results in less AP coverage area for a given minimum bit rate and higher link-dropping probability. Table 5.1 lists those parameters that have been modified from Table 4.2. The deployment parameters are tuned such that a bitrate of at least 6Mbps is experienced by all users. The new call arrival rate is also adjusted such that a typical 1% new call blocking is achieved. This results in the same deployment area as before, however, the minimum RSSI at the deployment boundaries is now much lower and equal to -80dBm. A vertical handoff is triggered at -85dBm which is 5dB below the edge of the coverage area making the possibility of false VHO triggering slim (0.1%).

Figure 5.15(a) plots new call blocking and VHO dropping for this deployment under the static reservation scheme. As expected, capacity-dropping for this deployment

is much higher than the previous one evaluated in Figure 4.2. For example, when no reservation is employed, capacity-dropping for the current deployment is 10% compared to 1.6% for the previous one. An almost negligible VHO capacity-dropping of 0.02% is achieved when half of the BAP bandwidth is statically reserved. Note that for the previous deployment it was possible to reach almost zero capacity-dropping at $G_{vho} = 0.35$. This is due to the larger propagation path loss exponents which cause a VHO call to spend more time using very low bitrates.

Larger path loss also means higher values of VHO link-dropping rate, which as shown clearly in Figure 5.15(a) amounts to 1.4%. This is much higher than the 0.1% link-dropping for the previous deployment. Consequently, in such a deployment one would experience a higher total dropping rate, which is dominated by link-dropping as VHO capacity-dropping improves. In fact, a minimum total VHO dropping of 1.4% is achieved at $G_{vho} = 0.35$. Although it is not possible to reduce VHO dropping any further, this is still a significant reduction from 10%.

An important difference between the current deployment and the one discussed in Chapter 5 is that for small fractions of bandwidth reservation in the previous deployment, it is possible to largely reduce VHO capacity-dropping. This is due to many VHO calls not shifting down to very low bitrates. For example, a 5% static bandwidth reservation resulted in a reduction of VHO capacity-dropping from 1.6% to 0.7%, according to Figure 4.2. In contrast, for the current deployment evaluated in Figure 5.15, the same 5% static reservation would result in almost no change in VHO capacity-dropping. In fact, a reasonable decline does not occur until 25% of the BAP's bandwidth is reserved. This is due to many VHO calls using very low bitrates in the latter deployment.

Another issue arising with low bit rate deployments is the possibility of ordinary calls not finding enough bandwidth to shift down to lower bitrates while they are still within the designated coverage area of the WLAN. For example, according to Table 4.1, going from a bitrate of 54Mbps to 6Mbps requires an additional 3% free bandwidth at the AP supporting such a call. However, such low bitrates, although allowed to exist in a deployment are relatively rare. More importantly, there will also be calls which will increase their bitrate as a result of improvements in their RSSI.

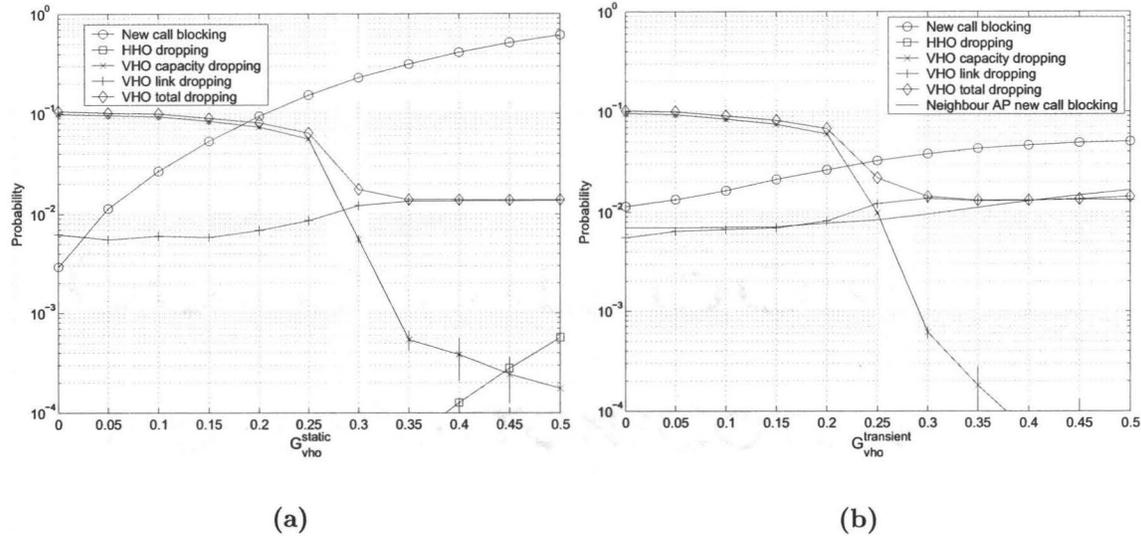


Figure 5.15: Influence of Static (a) and Transient (b) Bandwidth Reservation Schemes When Different Deployment and Wireless Propagation Parameters are Used.

This deficiency has been considered for the current deployment and it was found that 0.13% of the ordinary calls are forced to terminate due to a lack of bandwidth to accommodate for their bitrate changes. This is a relatively large amount compared to the 0.3% new call blocking rate (i.e., no scheme) shown in Figure 5.15(a). Therefore, it is best to statically reserve a fraction of the bandwidth in every AP (not just the BAP), to accommodate these calls. Extensive simulations for various scenarios and deployments show that for reasonable deployments a small amount of reservation is enough. Hence 5% of the bandwidth will be reserved for this purpose. Furthermore, this and any additional bandwidth reserved upon VHO triggering is shared by both new calls which are shifting to lower bitrates and VHO calls. The constant 5% reservation reduces new call dropping due to bitrate down-shifting to 0.005%; this is considerably less than the new call blocking rate. In addition, the new call blocking within the WLAN is now almost 1% according to Figure 5.15(a), which is a typical value.

The performance of the TR-MFH scheme under the current deployment scenario

is also evaluated in Figure 5.15(b) where a fixed 5% reservation is implicitly made in addition to the transient reservation, and the total reservation amount is again shared between bitrate down-shifting calls and VHO calls. Therefore, $G_{vho} = 0.10$ actually corresponds to 15% bandwidth reservation. By employing the TR-MFH approach it is possible to reduce VHO total dropping from 10% to 1.4%. New call blocking, however, rises to just below 4% which is significantly lower than the 30% blocking rate of SR for the same total capacity-dropping. In addition, the effect of the MFH procedure on adjacent AP new call blocking is plotted in Figure 5.15(b) as the line with no markers, which shows that a 30% transient reservation required to reduce total VHO dropping to 1.4% changed adjacent AP new call blocking from 0.70% to almost 0.95%. Hence, it is concluded that the TR-MFH approach outperforms SR and is a suitable option for tackling the VHO bandwidth deficit problem in WLANs.

5.8 Effect of Call Bitrate

Until now only VoIP calls using the common G.711 codec have been considered. In this section, the effect of changing the codec will be studied. To make fair comparisons the call arrival rate to the system will be tuned such that all scenarios result in the same new call blocking rate for the APs. The results for using various codecs will be given based on the analytical models derived in Chapters 4 and 5 for the static and transient reservation schemes.

The bandwidth deficit problem for IEEE 802.11 WLANs is aggravated due to the following reasons:

1. For each voice packet, the overhead for MAC, IP, UDP and RTP headers is large compared to the payload size. This results in a relatively larger bandwidth per call than what it needs to be at any given bit rate.
2. The physical layer (PHY) overhead associated with IEEE 802.11 is large and is always transmitted at the lowest PHY rate. This introduces a fixed and large bandwidth penalty to every packet transmission. Moreover, the ratio of the bandwidth penalty to payload increases as a call goes to higher bit rates,

spending less time transmitting the headers and payload while the PHY overhead still takes the same amount of time to transmit.

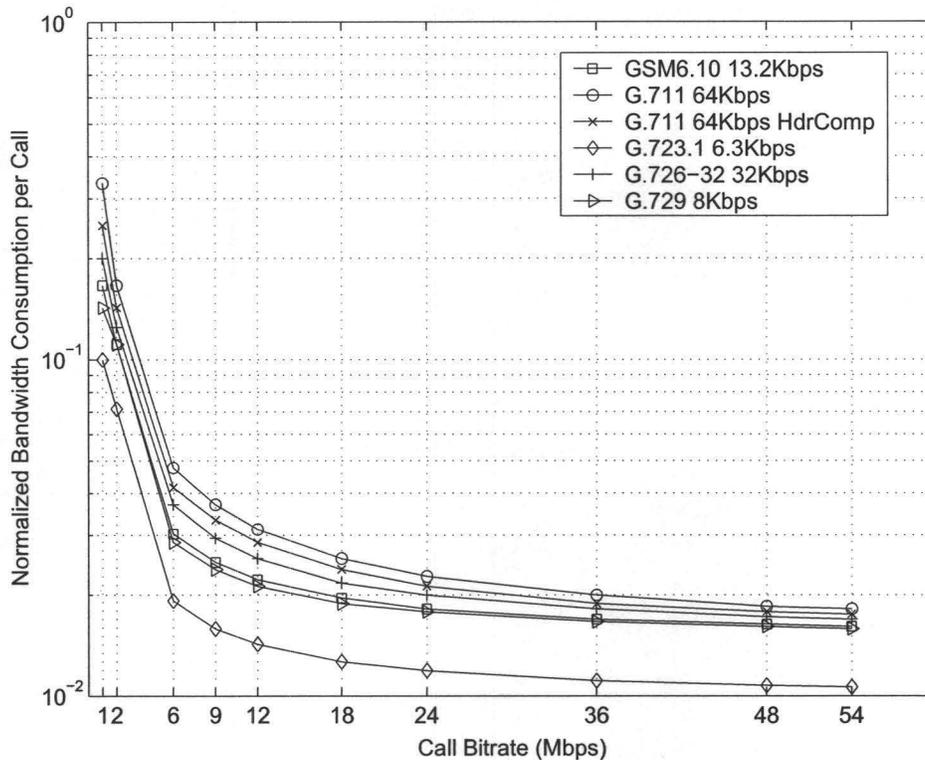


Figure 5.16: Fraction of an IEEE 802.11g AP Bandwidth Consumed by One Call as a Function of Its Bitrate for Some Well-Known Voice Codecs.

The combination of these two effects causes the huge bandwidth requirement of low bitrate calls as well as capacity saturation as bitrate is increased. Figure 5.16 shows the amount of normalized bandwidth used by a single voice call when using some example codecs versus the call bitrate for an IEEE 802.11g access point. The bandwidth calculations are performed by adopting the analytical method for calculating IEEE 802.11b voice capacity presented in [PA04]. There are two key features of the curves shown in Figure 5.16 which affect the capacity deficit problem. The first and most obvious is the amount of bandwidth required for operating at

the lowest bitrate². This dictates how much bandwidth reservation is required to reduce capacity-dropping. The other feature is the ratio of bandwidth at the lowest bit rate to nominal bandwidth used by a call. Without loss of generality it can be assumed that this is the ratio of bandwidth at the lowest bit rate to the bandwidth consumed at the maximum bitrate. Intuitively, this ratio represents the maximum weight of a VHO call compared to that of an ordinary call. This affects the value of capacity-dropping compared to new call blocking when no reservation is made.

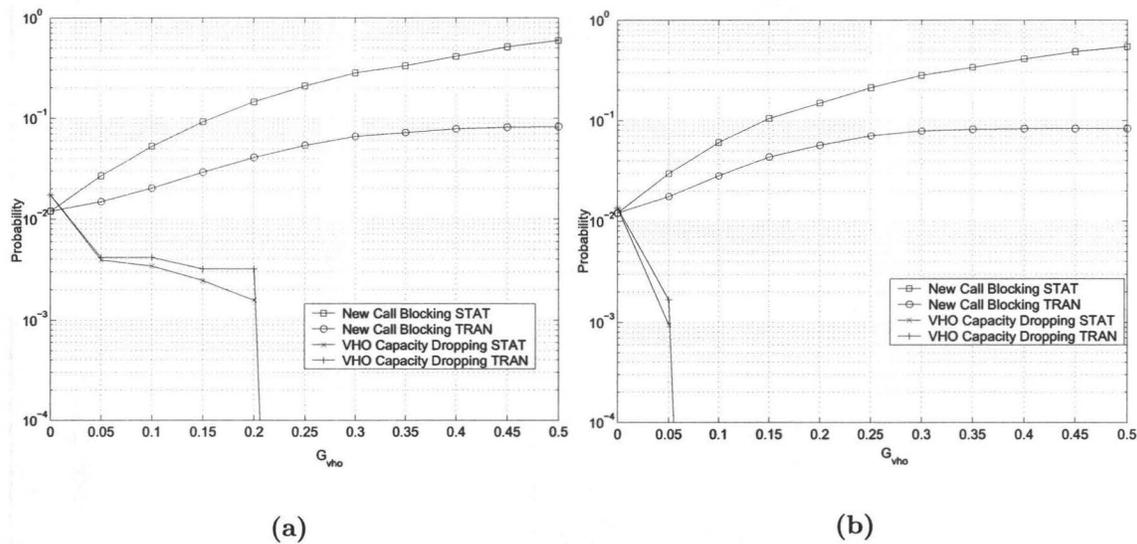


Figure 5.17: Performance Comparison of SR and TR-MFH Schemes for Different Voice Codecs: (a) G.711 With Header Compression and (b) G.723.1.

Figure 5.17 illustrates capacity-dropping versus the amount of bandwidth reservation for SR and TR-MFH when different codecs are used. The performance of G.711 with header compression (which reduces header size from 40 bytes to one), G.723.1 which is a low bitrate codec, and the original G.711 (Figure 5.18) are compared. It can be observed that G.711 requires the most reservation in both static and transient cases (slightly more than 30%) to achieve zero capacity-dropping. G.711 with header

²Note that here it is implicitly assume that some number of VHO calls will use the minimum bitrate. Otherwise the next lowest bitrate should be considered.

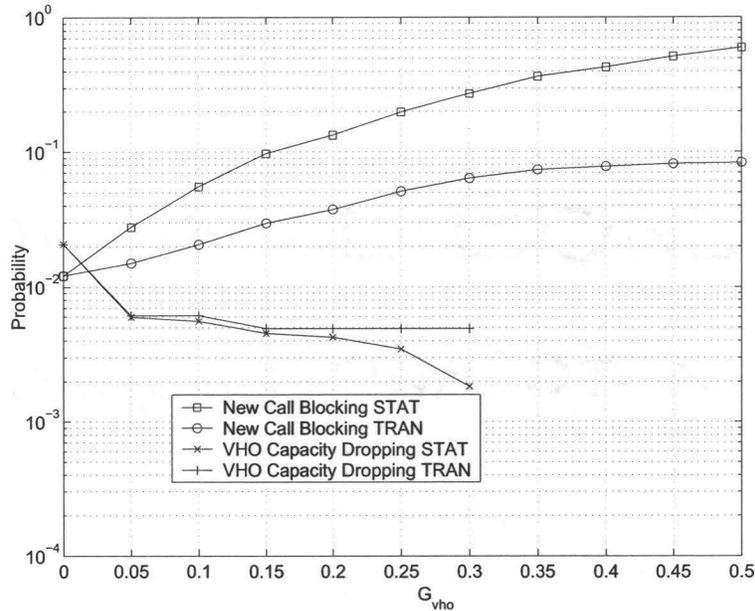


Figure 5.18: Performance of SR and TR-MFH Schemes for the G.711 Codec

compression requires slightly larger than 20%, and G.723.1 requires more than 5% reservation to achieve the same capacity-dropping. This can be explained by noting that the bandwidth used by a G.711 1Mbps call is 0.33 compared to 0.25 and 0.1 for G.711 with header compression and G.723.1, respectively.

The actual value of capacity-dropping when no bandwidth reservation is performed is 2% for G.711, which is larger than that for G.711 with header compression, i.e., 1.7% and G.723.1, i.e., 1.3%. This can be explained noting that for G.711 the bandwidth consumption ratio of a 1Mbps call compared to a 54Mbps call is equal to 18.3. This is larger than both the ratios for header compressed G.711 and G.723.1 codecs, which are 14.25 and 9.4, respectively.

As more efficient WLAN physical layer standards emerge, it is expected that less bandwidth will be needed by low bitrate calls. This also means that due to the small PHY layer overhead, the capacity will increase much sharper for higher bitrates which makes the ratio of low bitrate to high bitrate bandwidth consumption increase. This may result in a marginal increase in capacity-dropping. However, having more efficient

PHY layers, the APs are now able to support more calls and they could reach a point where the AP load for a reasonable range of new call arrival rate is small enough to eliminate capacity deficit as mentioned earlier in Section 5.4. Nevertheless the capacity deficit problem in general can apply to a WLAN deployment supporting any real-time application and is not limited to voice.

Chapter 6

Conclusion

This thesis has focused on the performance of a vertical handoff in loosely coupled architectures. The anchor-node based soft vertical handoff approach was adopted to provide lossless handoff from the WLAN to the cellular network. Loosely coupled architectures exhibit large VHO latencies, which in conjunction with the uncertain WLAN coverage at its boundaries can lead to loss of the WLAN link for the vertical handoff call. This can potentially render the soft handoff scheme useless.

The Vertical Handoff Support Node (VHSN) was proposed which reduces and potentially eliminates vertical handoff link loss. The VHSN is an entity which connects to both the WLAN infrastructure through a LAN connection and to the local cellular base station as a simple client node. When a WLAN link loss occurs, it quickly redirects the mobile station's connection through the cellular base station, while at the same time the MS is completing the soft VHO procedure and is establishing the cellular leg of the call. The VHSN approach allows for the independent operation of the WLAN and cellular networks and requires minimal changes to both. It potentially eliminates packet loss and the outage which results from a VHO. The approach is shown to be scalable in terms of the WLAN size that it can accommodate. The VHSN approach requires the addition of a node to the WLAN as well as requiring the cellular network to create a fast local redirection path.

Another issue that results in VHO dropping is the WLAN capacity deficit problem. More specifically, when a VHO call operates at a low bitrate (e.g., 1-2 Mbps), the

boundary AP serving the call may experience a bandwidth shortage, which leads to the VHO call being dropped. This is called capacity-dropping, in contrast to link-dropping.

Two solutions were proposed for the WLAN capacity deficit problem. The first is an adaptation of the well-known guard channel scheme used in cellular networks. This is referred to as the static reservation approach (SR), which reserves a fixed amount of bandwidth at the boundary AP in support of VHO calls. It is shown that this scheme eliminates VHO capacity-dropping, however, it significantly increases the new call blocking and horizontal handoff dropping rates at the WLAN boundary AP.

The second proposed solution is an efficient transient bandwidth reservation algorithm which reserves bandwidth at the boundary AP only when a VHO is triggered. The transient bandwidth reservation scheme was then augmented with a bandwidth securing mechanism when enough free bandwidth is not available at the boundary AP. This is performed by momentarily forcing ordinary calls within the boundary AP to horizontally handoff to one of the adjacent APs, until the VHO procedure is finished. This scheme is called Transient Reservation with Momentary Forced Horizontal handoff, TR-MFH.

Extensive simulations and analysis showed the superiority of TR-MFH to SR and that the new call blocking and HHO dropping rates are only slightly increased. Moreover, TR-MFH is able to achieve almost the same amount of improvement as SR in terms of VHO capacity-dropping. The effect of several different parameters was also considered such as, VHO latency, AP load, VHO arrival rate, wireless propagation parameters and different flow bitrates and codecs. The effect of adjacent AP coverage overlap on the performance of this approach was also studied. In addition, it was shown that for reasonably low VHO arrival rates the blocking rate for the TR-MFH scheme is always less than that for the SR approach.

This work can be extended in several ways. An interesting problem would be to investigate WLAN deployment techniques which reduce VHO link and capacity dropping. Such techniques could be based on providing a VHO call with graceful signal strength degradation as it moves away from the WLAN. In addition, the capacity-deficit problem can be taken into consideration by deploying enough APs at certain

locations. A direct extension of this work would be to consider outdoor WLAN deployments such as wireless mesh networks where the AP ranges are larger and users are more likely to experience low bitrates as they move out of coverage. Such networks have a potentially higher VHO rate and many APs may be boundary APs. This can impose a significant burden on the adjacent APs. Also the effect of WLAN capacity deficit on real-time applications such as video could be considered and the effectiveness of the TR-MFH scheme can be investigated for such cases.

Bibliography

- [Abr85] N. Abramson. “Development of the ALOHANET”. *IEEE Transactions on Information Theory*, March 1985.
- [ASKT06] S.V. Azhari, M. Smadi, V. Kezys, and T. D. Todd. “A method for WLAN data rate planning and usage during handover in multi-interface handsets”. Technical report, 2006.
- [BAA04] H. Badis and K. Al Agha. “Fast and efficient vertical handoffs in wireless overlay networks”. In *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications.*, volume 3, pages 1968–1972 Vol.3, 2004.
- [BBD03] N. Banerjee, K. Basu, and S.K. Das. “Hand-off delay analysis in SIP-based mobility management in wireless networks”. In *Proceedings of Parallel and Distributed Processing Symposium*, pages 22–26, 2003.
- [BCI04] M. Bernaschi, F. Cacace, and G. Iannello. “Vertical handoff performance in heterogeneous networks”. In *International Conference on Parallel Processing, Workshops Proceedings*, pages 100–107, 2004.
- [BDA05] N. Banerjee, S. K. Das, and A. Acharya. “SIP-based mobility architecture for next generation wireless networks”. In *Third IEEE International Conference on Pervasive Computing and Communications*, pages 181–190, 2005.
- [BHJ03] Hongyang Bing, Chen He, and Lingge Jiang. “Performance analysis of vertical handover in a UMTS-WLAN integrated network”. In *14th*

IEEE Proceedings on Personal, Indoor and Mobile Radio Communications, PIMRC, volume 1, pages 187–191 Vol.1, 2003.

- [BHPC04] C. Bettstetter, H. Hartenstein, and X. Perez-Costa. “Stochastic properties of the random waypoint mobility model”. *Wireless Network Journal*, 10(5):555–567, 2004.
- [Bic05] J. C. Bicket. “Bit-rate selection in wireless networks”. *MSc Thesis, Massachusetts Institute of Technology*, 2005.
- [BSK95] H. Balakrishnan, S. Seshan, and R. H. Katz. “Improving reliable transport and handoff performance in cellular wireless networks”. *Wireless Network Journal*, 1(4):469–481, 1995.
- [BSL05] D. Bultmann, M. Siebert, and M. Lott. “Performance evaluation of accurate trigger generation for vertical handover”. In *Proceedings of 16th Annual IEEE International Symposium on Personal Indoor and Mobile Radio Communications*, pages H–10–02, 2005.
- [Cel] Two Billion GSM Customers Worldwide, http://www.uwcc.org/english/news_room/, last checked december 2007.
- [Cisa] “Cisco 7920 wireless IP phone design and deployment guide”, <http://www.cisco.com/en/US/products/hw/phones>. Manual, last checked dec 2007, Cisco.
- [Cisb] Cisco aironet 1100 series access point data sheet, <http://www.cisco.com/en/US/products/hw/wireless/>, last checked December 2007.
- [Dua] Dual-mode WiFi/cellular VoIP Phones Sales Up, <http://www.cellular-news.com/story/21624.php>, last checked december 2007.

- [ea97] C.E. Perkins et al. "Route optimization in Mobile IP". Mobile IP working group, Internet draft- Work in progress, Nov 1997.
- [ea99] C.E. Perkins et al. "Optimized smooth handoffs in Mobile IP". In *IEEE International Symposium on Computers and Communications*, pages 340–346, July 1999.
- [ea01] K. El Malki et al. "Low latency handoff in Mobile IPv4". draft- work in progress, IETF, May 2001.
- [ea02a] J. Rosenberg et al. "SIP: Session initiation protocol". RFC 3261, 2002.
- [ea02b] Z. Jiang et al. "Seamless mobility management based on proxy servers". In *IEEE WCNC*, pages 563–568 vol.2, March 2002.
- [ea04a] D. Johnson et al. "Mobility Support in IPv6". RFC 3775, June 2004.
- [ea04b] H. Choi et. al. "A seamless handoff scheme for UMTS-WLAN interworking". In *GLOBECOM*, pages 1559–1564 vol. 3, Nov-Dec 2004.
- [ea04c] R. Chakravorty et al. "Performance issues with vertical handovers - experiences from GPRS cellular and WLAN hot-spots integration". In *PERCOM'04*, pages 155–164, March 2004.
- [ETS01] ETSI. "Requirements and architectures for interworking between HIPER-LAN/3 and 3rd generation cellular systems". Technical Report TR 101 957, ETSI, August 2001.
- [FT05] S. Fashandi and T. D. Todd. "Real-time handoff in solar/battery powered ESS mesh networks". In *16th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC 2005)*, 2005.
- [Goo99] B. Goodman. "Internet telephony and modem delay". *IEEE Network*, 13(3):8–16, May-June 1999.
- [Gue88] R. Guerin. "Queuing-blocking systems with two arrival streams and guarded channels". *IEEE Transactions on Communications*, 36, 1988.

- [Hil01] A. Hills. “Large-scale wireless LAN design”. *IEEE Communications Magazine*, 39:98–107, 2001.
- [Hot] Public Hotspot Deployments Show No Sign of Abating, <http://www.in-stat.com/press.asp?sku=in0602869wn&id=1830>, last checked december 2007.
- [HWT⁺05] M. He, X. Wang, T. Todd, D. Zhao, and V. Kezys. “Ad hoc assisted handoff in IEEE 802.11 infrastructure networks”. *International Journal on Computer and Communication Networks*, to appear, 2005.
- [JG97] C. Jain and D. J. Goodman. “General packet radio service in GSM”. *IEEE Communications Magazine*, October 1997.
- [KCC05] Rong-Jyh Kang, Hsung-Pin Chang, and Ruei-Chuan Chang. “A seamless vertical handoff scheme”. In *Proceedings of the First International Conference on Wireless Internet*, pages 64–71, 2005.
- [KT75] L. Kleinrock and F. Tobagi. “Random access techniques for data transmission over packet-switched radio channels”. In *AFIPS conference proceedings vol. 44*, 1975.
- [KTZ04] P. Khadivi, T. D. Todd, and Dongmei Zhao. “Handoff trigger nodes for hybrid IEEE 802.11 WLAN/cellular networks”. In *First International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks, QSHINE*, pages 164–170, 2004.
- [LJCG04] Benny Chen Venkatesh Rajendran Ling-Jyh Chen, Tony Sun and Mario Gerla. “A smart decision model for vertical handoff”. In *The 4th ANWIRE International Workshop on Wireless Internet and Reconfigurability (ANWIRE 2004, in conjunction with IFIP Networking’04)*, 2004.
- [LSB⁺04] M. Lott, M. Siebert, S. Bonjour, D. von Hugo, and M. Weckerle. “Interworking of WLAN and 3G systems”. *IEE Proceedings on Communications*, 151:507–513, 2004.
-

- [LZ05] Guming Liu and Chi Zhou. “HCRAS: A novel hybrid internetworking architecture between WLAN and UMTS cellular networks”. In *Second IEEE Consumer Communications and Networking Conference, CCNC*, pages 374–379, 2005.
- [MS98] R. Katz M. Stemm. “Vertical handoffs in wireless overlay networks”. *Mobile Networks and Applications 3*, pages 335–350, 1998.
- [MSA04] A. Mishra, M. Shin, and W. A. Arbaush. “Context caching using neighbor graphs for fast handoffs in a wireless network”. In *Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies INFOCOM*, volume 1, page 361, 2004.
- [Net00] Nortel Networks. “A comparison between GERAN packet-switched call setup using SIP and GSM circuit-switched call setup using RIL3-CC, RIL3-MM, RIL3-RR, and DTAP”. *3GPP TSG GERAN 2, GP-000508*, November 2000.
- [NKK05] S. Villette N. Katugampala, K. Al-Naimi and A. Kondo. “Real time end to end secure voice communications over GSM voice channel”. *13th European Signal Processing Conference*, September 2005.
- [PA04] Hole D. P. and Tobagi F. A. “Capacity of an IEEE 802.11b wireless LAN supporting VoIP”. In *IEEE International Conference on Communications*, volume 1, pages 196–201, 2004.
- [Per02] C. Perkins. “IP mobility support for IPv4”. RFC 3344, August 2002.
- [PKH⁺00] K. Pahlavan, P. Krishnamurthy, A. Hatami, M. Ylianttila, J. P. Makela, R. Pichna, and J. Vallstron. “Handoff in hybrid mobile data networks”. *IEEE Personal Communications*, [see also *IEEE Wireless Communications*], 7:34–47, 2000.
- [Pra00] N. R. Prasad. “IEEE 802.11 system design”. In *IEEE International Conference on Personal Wireless Communications*, pages 490–494, 2000.

- [QW00] C. Qiao and H. Wu. “iCAR: An integrated cellular and ad-hoc relay system”. In *Ninth International Conference on Computer Communications and Networks*, pages 154–161, 2000.
- [Rap96] T.S. Rappaport. “*Wireless communications : principles and practice*”. Upper Saddle River, N.J. : Prentice Hall PTR, 1996.
- [RRT96] R. Nagarajan R. Ramjee and D. Towsley. “On optimal call admission control in cellular networks”. *IEEE Infocom*, March 1996.
- [RS05] I. Ramani and S. Savage. “SyncScan: practical fast handoff for 802.11 infrastructure networks”. In *24th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM*, volume 1, pages 675–684 vol. 1, 2005.
- [SAT06] M. Smadi, S.V. Azhari, and T. D. Todd. “A measurement-based study of WLAN to cellular handover”. In *2nd IEEE International Workshop on Heterogeneous Multi Hop Wireless and Mobile Networks*, 2006.
- [SKK01] R. Shirdokar, J. Kabara, and P. Krishnamurthy. “A QoS-based indoor wireless data network design for VoIP applications”. In *Vehicular Technology Conference*, volume 4, pages 2594–2598 vol.4, 2001.
- [SMA04] Minho Shin, Arunesh Mishra, and William A. Arbaugh. “Improving the latency of 802.11 handoffs using neighbor graphs”. In *MobiSys '04: Proceedings of the 2nd international conference on Mobile systems, applications, and services*, pages 70–83, New York, NY, USA, 2004. ACM Press.
- [Sto02] I. Stojmenovic. “*Handbook of wireless networks and mobile computing*”. John Wiley & Sons, Inc., New York, NY, USA, 2002.
- [STZK06] M. Smadi, T.D. Todd, D. Zhao, and V. Kezys. “Dynamically anchored conferencing handoff for dual-mode cellular/WLAN handsets”. In *International Conference on Communications (ICC), Istanbul, Turkey*, June 2006.

- [UMA] UMA Today,<http://www.umatoday.com/>, last checked december 2007.
- [VK04] H. Velayos and G. Karlsson. “Techniques to reduce the IEEE 802.11b handoff time”. In *IEEE International Conference on Communications*, volume 7, pages 3844–3848 Vol.7, 2004.
- [WBBD05] Wei Wu, N. Banerjee, K. Basu, and S. K. Das. “SIP-based vertical handoff between WWANs and WLANs”. *IEEE Wireless Communications*, [see also *IEEE Personal Communications*], 12:66–72, 2005.
- [WCM00] X. Wu, S-H.G. Chan, and B. Mukherjee. “MADF: A novel approach to add an ad-hoc overlay on a fixed cellular infrastructure”. In *IEEE Wireless Communications and Networking Conference (WCNC)*, volume 2, pages 549–554, 2000.
- [WiF] Consumer Adoption of Wi-Fi Telephony Driving Up Wi-Fi Handset Sales,<http://www.tmcnet.com/planetpdamag/articles/>, last checked december 2007.
- [ZH96] N. Zhang and J.M. Holtzman. “Analysis of handoff algorithms using both absolute and relative measurements”. *IEEE Transactions on Vehicular Technology*, February 1996.
- [ZM04] Fang Zhu and J. McNair. “Optimizations for vertical handoff decision algorithms”. In *IEEE Wireless Communications and Networking Conference, WCNC*, volume 2, pages 867–872 Vol.2, 2004.
- [ZVP03] S. Zvanovec, M. Valek, and P. Pechac. “Results of indoor propagation measurement campaign for WLAN systems operating in 2.4 GHz ISM band”. In *Twelfth International Conference on Antennas and Propagation, ICAP*, volume 1, pages 63–66 vol.1, 2003.