WLAN TO WWAN VERTICAL HANDOVER IN LOOSELY-COUPLED NETWORKS

WLAN TO WWAN VERTICAL HANDOVERS IN LOOSELY-COUPLED NETWORKS

BY
MOHAMMED N. SMADI, M.A.SC, B.ENG AND MGMT.
APRIL 2008

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

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

McMaster University
Hamilton, Ontario

TITLE:	WLAN to WWAN Vertical Handovers in Loosely-Coupled Networks
AUTHOR:	Mohammed N. Smadi, M.A.Sc, B.Eng and Mgmt.
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:	xi,127

Abstract

Wireless handsets are becoming increasingly multi-modal, supporting both cellular and wireless LAN air interfaces. When such a handset roams from one network to another during an active voice call, the call must be re-routed into the new network. The process of re-routing calls in this manner is referred to as a vertical handover.

Achieving seamless WLAN-to-cellular handover can be very difficult due to the fact that WLAN coverage can often be lost long before a cellular call leg replacement can be established. A possible worse-case scenario occurs when mobile users walk from indoor building WLAN coverage to outdoors during voice connections. In this thesis, results from a measurement-based study of dual-mode handover are given. The presented results come from extensive IEEE 802.11 measurements that were made on the McMaster University campus. These results give important insights into the difficulty of this problem, and relate to various system parameters such as WiFi deployment type and link loss threshold. The results show that it is almost impossible to successfully complete the handover unless the WLAN deployment has been carefully engineered. In WLAN deployments which would ordinarily restrict the use of lower data rates, the results also suggest that a "limited data rate use" (LDU) algorithm can greatly improve the probability of seamless handover. This can be done without adversely affecting the capacity of the WLAN network.

Vertical handover can be performed using enterprise voice gateways which support basic conference bridging. This feature can be used to achieve soft handover, since a new leg of the call can be established to the conference bridge while the existing media stream path is active. Unfortunately this requires that all intra-enterprise calls be routed through the gateway when the call is established. In this work we consider a SIP based architecture to perform conferenced dual-mode handover and propose a much more scalable mechanism for short-delay environments, whereby active calls are handed off into the conference bridge prior to the initiation of the vertical handover. Results are presented which are taken from a dual-mode handset testbed, from analytic models, and from simulations which characterize the scalability of the proposed mechanism.

"رب أوزعني أن أشكر نعمتك التي
أنعمت علي وعلى والدي وأن أعمل
صالحاً ترضاه، وأدخلني برحمتك في عبادك
الصالحين"

Acknowledgments

I am grateful to my supervisor Dr. Todd for his support throughout this work. Terry has given me an opportunity to work on an interesting project with excellent people: an experience that changed how I view and appreciate things. I would like to thank Dr. Shirani and Dr. Zhao for being members of my committee and for their support. I also would like to thank Vytas Kezys from Research in Motion for his help and support. Many of the technical discussions we had with Vytas helped shape the overall picture of this work.

I would like to acknowledge the support of ECE's administrative and technical team. Special thanks are owed to Terry Greenlay, Cosmin and Cheryl for they were never bothered by my endless questions and requests.

I would like to express my appreciation to the enormous group of friends who never seized to believe in me. Out of this group, special recognition is payed to Amir, Ahmad, Amin, Enrique, and Shervan. In addition to this group, Vahid Azhari was a special team-mate who contributed significantly to my work through discussion and technical help. I truly enjoyed working with you guys, and wish for a life-long friendship.

To my family and fiancée, thanks for your support, belief and encouragement. It made all the difference.

To my parents, Nawaf and Lamah, it is because of your support and sacrifice I am able to write these words today. Jazakum Allah Khairn!

List of Acronyms

2G	Second Generation Cellular Networks
3G	Third Generation Cellular Networks
3PCC	Third Party Call Controller
AMPS	Advanced Mobile Phone System
AP	Access Point
B3G	Beyond Third Generation Cellular Networks
BSS	Basic Service Set
DACH	Dynamically Anchored Conferencing Handover
DM	Dual – Mode Handset
DS	Distribution System
EBSS	Extended Basic Service Set
ETSI	European Telecommunications Standards Institute
FCC	Federal Communications Commission
FDMA	Frequency Division Multiple Access
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
GW	Gateway
HHO	Horizontal Handover
HP	Hairpin Device
HTTP	Hypertext Transfer Protocol
iBSS	Independent Basic Service Set
IP	Internet Protocol

LAN	Local Area Network
LDU	Limited Data Rate Use Algorithm
MIME	Multipurpose Internet Mail Extensions
MIP	Mobile Internet Protocol Version4
MIPv6	Mobile Internet Protocol Version6
PBX	Private Branch Exhange
PLMN	Public Land Mobile Network
PoD	Point of Departure
PSTN	Public Switched Telephone Network
RA	Router Advertisements
RF	Radio Frequency
RSSI	Received Signal Strength Indicator
RTP	Real – Time Protocol
SDP	Session Description Protocol
SIP	Session Initiation Protocol
SMB	Small and Medium Size Businesses
TCP	Transport Control Protocol
UA	User Agent
UDP	User Datagram Protocol
URI	Uniform Resource Identifier
VHO	Vertical Handover
VoIP	Voice over Internet Protocol
WLAN	Wireless Local Area Network
WWAN	Wireless Wide Area Network

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

Wireless devices are becoming increasingly multi-modal, containing both wireless local area network (WLAN) and cellular air interfaces. This allows the device to access high speed WLAN coverage when it is available, and to communicate via the cellular network in other situations. In order to provide for full interoperability, real-time seamless handover between the two handset interfaces is very important. Since cellular coverage is typically much more widespread than that of WLAN, WLAN-to-cellular handover usually poses more strict time constraints than that in the cellular-to-WLAN direction.

The European Telecommunications Standards Institute (ETSI) has specified two approaches for WLAN-cellular interworking, namely, loose coupling and tight coupling [ETS01]. In tight coupling the WLAN is integrated into the service provider's cellular core network and an interworking gateway (GW) provides adaptation between the two systems. Tight coupling benefits from good handover delay and packet loss performance since it uses native cellular mobility management protocols. This approach is currently being standardized under the UMA activities [UMA08]. In contrast, loosely-coupled architectures connect the WLAN to the cellular network through an *anchor node* residing at an external Internet Protocol (IP) network, such as that found in the enterprise. This approach is less proprietary but real-time handover may be more difficult to achieve due to the long handover delays associated with cellular call establishment times [CVS+04, PKH+00]. The difficulty of this approach stems from two reasons. First, due to the vagaries of WLAN propagation, an active call over the WLAN interface may be lost long before the cellular call is established [SAT06]. Second, the scalability of the anchor node is heavily dependent on the mobility pattern of the user population [STZK06, STK+].

1.2 Motivation

The motivation behind this thesis is the need for ways to integrate loosely-coupled heterogeneous WLAN and cellular networks. Algorithms need to be developed which consider the abrupt WLAN signal strength degradation at the boundaries of WLAN hotzones. This abruptness of WLAN degradation makes performing a handover very difficult in light of the lengthy cellular call establishment times observed on today's Public Land Mobile Networks (PLMN) and Public Switched Telephone Networks (PSTN).

Cellular communication networks have been around for decades, and well established procedures have been developed for inter-cell user roaming. Similarly, well established procedures have been developed for user roaming within WLAN coverage. However, the use of WLANs has been gaining ground steadily and quickly. For example, McMaster University Technical Services (UTS) has reported a 400% increase in the use of its WLAN deployment over the past two years. This increase is fueled by many applications with real-time requirements such as Voice over Internet Protocol (VoIP) and media streaming. These applications are being integrated into devices with cellular and WLAN capabilities. Users actively engaged in media sessions while in the WLAN, expect to continue their sessions even after leaving WLAN coverage. A motivation of this thesis is to provide an enterprise-centric approach to allow users to seamlessly switch their voice calls from the WLAN to the cellular network and in the opposite direction.

Many businesses have adopted VoIP as their telephony technology. According to the Yankee Group 2004 Small and Medium size Businesses (SMB) Bundled Communications Survey, 70% of SMBs have mobile workers. The same report indicates that one-third of SMBs are interested in getting VoIP services. However, all VoIP-capable networks include gateways to continue communicating with the PSTN. An interesting feature of these gateways is their ability to perform conference bridging and to act as media stream mixers. The algorithms proposed in this thesis leverage this feature of existing enterprise PSTN gateways to perform seamless vertical handovers. Scalability issues which may arise under certain mobility conditions are resolved by using standard VoIP functionality.

At the physical level, the vagaries of WLAN propagation at the boundary of a WLAN deployment have a strong impact on the proposed algorithms. Another motivation of this thesis is to understand this impact. The conclusions arrived at provide insights for WLAN deployment engineers who are concerned about WLAN-to-cellular handovers.

1.3 Scope

Soft handovers have been extensively discussed in the literature. To highlight the difficulty of vertical handovers in loosely-coupled architectures we performed a measurement-based study of WLAN-to-cellular handover as mobile users roam out of a WLAN hotspot into cellular coverage. Our results are based on extensive WLAN measurements that were made on the McMaster University campus, and involved traversing many indoor-to-outdoor transition regions for a large number of campus building exits while monitoring WLAN access point (AP) coverage. The collected data was then processed to determine the probability of seamless handover using conventional threshold-based vertical handover algorithms. The results presented give important insights into the difficulty of this problem, and relate to issues such as WLAN deployment, handover triggering, and WLAN link loss threshold. In WLAN deployments restricting the use of lower data rates, a tradeoff is identified between WLAN-capacity and handover success probability. The results suggest that a Limited Data Rate Use (LDU) algorithm can improve the probability of seamless handover without severely affecting the WLAN capacity.

An architecture allowing for seamless roaming between circuit-switched cellular telephony and packet-switched WLAN telephony is proposed. This architecture is based on standard VoIP signaling, and leverages the conferencing/bridging abilities of standard PSTN Private Branch Exchanges (PBX) to anchor calls at establishment time through available conference room bridging. Anchoring the calls allows for seamlessly swapping the WLAN leg of the call by a newly established cellular leg upon vertical handover, or vice versa. The correctness and viability of the proposed architecture is verified by implementing it in a testbed and performing realtime handovers. The implementation of the architecture could reside on the PBX or on a separate entity that we refer to as a hairpin device (HP). We develop analytic models and show that in some situations, i.e., when the architecture is assumed to support large mobile populations, resource availability may be an issue. To address this situation, we propose Dynamically Anchored Conferencing Handover (DACH), which is a novel mechanism that exploits the low latency typically found in enterprise local area networks (LAN). In DACH a hard handover is used to move the call into the PBX/gateway anchor prior to a vertical handover. The scalability provided by DACH is useful in situations where there are large numbers of enterprise VoIP users, with heavy intra-enterprise call statistics. However the benefit of DACH may not be evident when the call arrival rate is small or the population is static.

1.4 Organization of Thesis

In Chapter 2 we describe WLANs and cellular communication networks as examples of heterogeneous wireless networks. This chapter also provides an overview of Voice over IP and inter-networking between a VoIP-capable enterprise network and the PSTN. The chapter also reviews the Session Initiation Protocol (SIP) and background needed for the subsequent chapters. Finally, the chapter defines the vertical handover problem with a comprehensive presentation of background literature on vertical handover algorithms and their performance.

Chapter 3 is devoted to discussing the relationship between WLAN propagation and the vertical handover success probability in loosely-coupled systems. The chapter introduces WLAN deployment parameters which are used to define the success probability of VHO. After this, a description of our data collection methodology is presented, followed by a discussion of the effect that different parameters have on VHO success probability. The idea of a Trigger Node is introduced and an assessment is provided.

Chapter 4 describes the system architecture and SIP signaling required to perform bi-directional soft handovers. Scalability issues are highlighted with the proposed architecture and an alternative, more scalable architecture is proposed. The chapter also provides an analytic characterization of the system performance, shows media traces of realtime handovers taken from our testbed, describes the simulation environment, and presents various performance results.

Chapter 5 identifies a dilemma resulting from the conflicting WLAN-capacity and VHO success requirements. The LDU algorithm is presented and assessed as a solution for the WLAN-capacity and VHO success probability tradeoff. The Building Exit model is developed to accurately calculate VHO success probability in common situations.

Chapter 6 presents conclusions drawn from this work. Areas for future work in WLAN-to-cellular handover research are outlined.

Chapter 2

Background

2.1 Overview

This chapter provides an introduction to cellular communication networks and WLANs. The building blocks of WLANs are described and the different operational modes of WLANs are explained. Voice over IP technology and the signaling methods of SIP are reviewed. Finally, a description of the VHO problem is provided with a comprehensive literature review.

2.2 Cellular Networks

Radio telephony can be described as a wireless and mobile interface to the wired PSTN [ZS05]. In 1921, the Detroit Michigan Police Department made the earliest significant vehicular use of mobile radio. Soon afterwards the channels became too congested due to limited bandwidth availability. In 1940 more channels were added to revitalize the use of mobile telephony by police departments and first respondent units. This system consisted of a single base-station that covered a large geographical area and was the first instance of Wireless Wide Area Networks (WWAN). The available bandwidth was divided into many channels, and each new call was assigned a channel for the duration of the call. This protocol allowed for multi-user access and was called Frequency Division Multiple Access (FDMA).

At this stage it became apparent that the single-cell system is not scalable due to the limited number of available channels. In 1970, the US Federal Communications Commission (FCC) invited proposals for solving the capacity problem. AT&T proposed a cellular approach where the geographic area is divided into smaller radio coverage cells, and the available frequencies are divided into non-overlapping sets. Each of these frequency sets is allocated to a given cell and is reused after a minimum separation distance which guarantees that the interference does not exceed a certain threshold. In Figure 2.1, cellular frequency reuse is shown with seven frequency sets (i.e., f_1, f_2, etc...) covering 21 cells. Note that no one frequency set is used in adjacent cells.

Mobile stations moving from one cell to another will have to change their operational frequency. This process is referred to as handover and normally occurs at

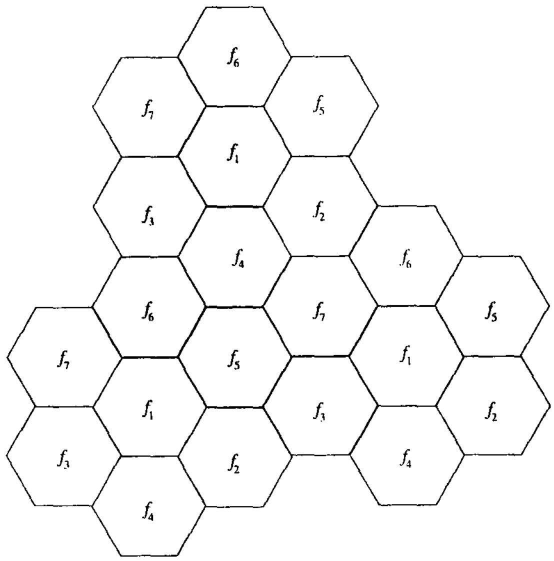


Figure 2.1: *Frequency Resuse in Cellular Systems*

the boundary between adjacent cells. This condition is detected when the Received Signal Strength Indicator (RSSI) from one cell becomes considerably lower than the RSSI of the adjacent cell, effectively triggering the inter-cell handover.

First generation cellular systems were based on analog technology such as the Advanced Mobile Phone System (AMPS) introduced by AT&T and the Nordic Mobile Telephone introduced by Ericsson. Second generation (2G) cellular systems witnessed the introduction of digital transmission technology and benefited from higher transmission rates, increased security and many additional services at lower cost than their analog counterparts. The European Global System for Mobile Communications (GSM), the American Digital AMPS and the Japanese Digital Cellular (JDC) system are all examples of 2G digital cellular systems [Kuc91]. Higher data rates of the 2G networks allowed for many data services. For example, the General Packet Radio Service (GPRS) that is the data counterpart of the voice-capable GSM network, provides services at data rates close to 80 kbps such as email-over-cellular which exists in many of today's cellular handsets.

Feasibility of third generation (3G) networks is being considered in many parts of the world. The first deployment took place in Japan in 2001. 3G networks are expected to support new media applications such as TV and music streaming. Currently, researchers are studying Beyond 3G (B3G) and fourth generation cellular networks [ZS05].

2.3 Wireless Local Area Networks

The use of WLANs has increased dramatically during the past decade. WLAN products have become commodity items used in professional and consumer products alike. The most common use for WLANs is the extension of wired networks, thereby giving stations equipped with wireless interfaces a higher degree of mobility. The two most common WLAN standards are the North American IEEE 802.11 and High Performance European Radio LAN. In the remainder of this section, IEEE 802.11 terminology is used to illustrate relevant concepts.

IEEE 802.11 (commonly branded as WiFi), is a physical layer specification and

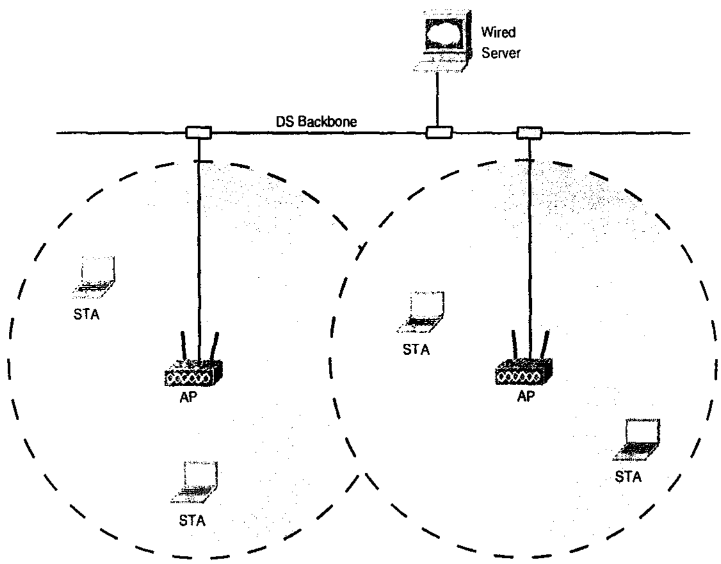


Figure 2.2: *WLAN with Two Basic Service Sets*

medium access control protocol designed for short range, high data rate wireless communications [Gas05]. Figure 2.2 shows the basic components of an IEEE 802.11 wireless network. The AP is the central entity of each coverage area with coordination functionality. Additionally, the AP acts as a bi-directional bridge between the wireless medium and the wired infrastructure (i.e., typically Ethernet). Stations (STA) are mobile devices equipped with IEEE 802.11 interfaces.

A Station must be associated with an AP in order for it to transmit and receive data from the wired infrastructure and to communicate with other stations in the same WLAN. The distribution system (DS) is the wired backbone connecting APs in one network and allowing the associated stations to access services available on the

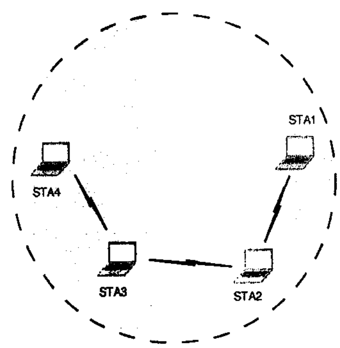


Figure 2.3: *WLAN Operating in Ad-Hoc Mode*

wired infrastructure. Communication between the AP and the associated stations occurs over the radio frequency (RF) that carries the data.

A Basic Service Set (BSS) is the term used to refer to an AP and its associated stations. In large WLANs where a single BSS is not sufficient to provide the desired coverage, multiple BSSs can be joined together. This setup of having two or more BSSs is referred to as an Extended Service Set (ESS). Figure 2.2 shows an ESS consisting of two BSSs. The two APs can communicate via the connecting DS, and they can bridge packets between any of their associated stations and the server connected to the wired infrastructure or any two stations residing in different BSSs. ESSs allow mobile stations to seamlessly handover between different BSSs.

In Figure 2.3 we show a less common mode for WLAN operation that is referred to as an independent BSS (iBSS) or more commonly as ad-hoc mode. In this mode, stations do not need an AP to communicate with each other, however, they have to be

Table 2.1: Range and Rate Comparison between WLAN and Cellular Networks

Network	**Rate**	**Indoor Range**	**Outdoor Range**
Legacy	2Mbps	35 meters	70 meters
802.11a/g	54Mbps	35 meters	150 meters
802.11n	250Mbps	100 meters	200 meters
GSM/EDGE	80kbps	N/A	35 km

within direct radio range. For example, `STA2` shown in Figure 2.3 can communicate with `STA1` and `STA3`. However, `STA1` and `STA4` have no direct way of communication. An iBSS can be used in situations where a temporary deployment is desired without the need to access information that is external to the network, a function that is essential to most WLANs.

WLANs have a much shorter range when compared to cellular networks, however they have data rates which are orders of magnitude greater. For example, IEEE 802.11 which was ratified in 1997 had a data rate of 2 Mbps, an indoor range of ~35 meters and an outdoor range of ~70 meters. Since 1997, IEEE 802.11 has seen many advancements with the introduction of IEEE 802.11a and IEEE 802.11g. Both standards have data rates of 54 Mbps and indoor range of ~35 meters and outdoor range of ~150 meters. Most recently, drafts of the new IEEE 802.11n standard have appeared with data rates reaching 250 Mbps and an indoor range of ~100 meters and an outdoor range of ~200 meters. In Table 2.1 we summarize the data rates of the IEEE 802.11 standards and compare them to those seen in a GSM/GPRS network (chosen as an example of cellular networks).

2.4 Voice over Internet Protocol

LANs were originally designed to carry data traffic for applications which are delay insensitive, such as emails, file transfers and web-browsing. As LAN data rates increased significantly, applications which require strict guarantees on bandwidth and delay could be admitted by booking the needed network resources. An example of such applications is VoIP, where voice calls are established over data networks with minimal cost since the infrastructure already exists, allowing for free voice communications.

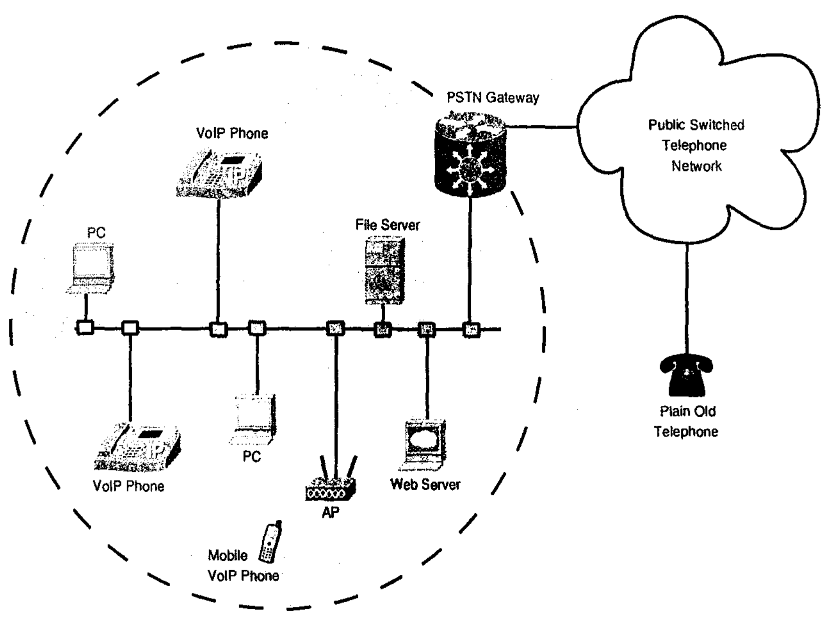


Figure 2.4: *A VoIP-Capable Network with a PSTN Gateway*

Many businesses have realized the advantages associated with having one network to carry voice and data traffic. Although VoIP-capable networks are quickly growing as a percentage of the communications market, they still represent a small percentage of the overall voice market. Therefore, these networks must be able to communicate with other types of networks especially the PSTN. The boundary entities performing the translation between the two networks are referred to as gateways. In Figure 2.4 an enterprise with one LAN carrying voice and data traffic is shown. PCs can access data off the Web and file storage servers for example, whereas the VoIP phones can communicate directly or can dial a plain old telephone that is residing on the PSTN. However, in the latter case, the call must be routed through the PSTN gateway to perform the necessary signaling and media conversion. Additionally, Figure 2.4 shows a mobile VoIP phone with a WLAN interface that is accessing the LAN through an AP. WLAN VoIP phones communicate in the same way that wired VoIP phones communicate. It remains to say that SIP is the de facto standard for establishing media and voice sessions in IP LANs. The next section provides a review of SIP.

2.5 Session Initiation Protocol

SIP [RSC+02, SJ01, SR99, SR00] is a signaling protocol that establishes, modifies, and terminates multimedia sessions. It can be used to invite new members to an existing session or to create brand new sessions [Cam01]. SIP requests and responses consist of a text header and Multipurpose Internet Mail Extensions (MIME) body.

This syntax is inherited from the Hypertext Transfer Protocol (HTTP) [SW00]. In [SJ01], the authors classify SIP entities as *user agents*, *servers* and *location servers.* A user agent (UA) is logically split into two sub-entities, namely: *user agent client* and *user agent server.* A user agent client residing within a user agent initiates *requests* and replies to the *responses* generated by the user agent server to these requests. Generally, user agents are the only elements of a SIP network where media and signaling converge [SW00]. A user agent can either be embedded in Internet hardware phones, or implemented in software running on handhelds, PCs or even PSTN gateways.

According to [SJ01], SIP servers can be further classified as *registrar*, *proxy* and *redirect* servers. A registrar server accepts registration requests from a user agent and acts as an interface to the location registrar. A proxy server can either statelessly forward requests received from user agents to their destination (or to a redirect server) or can maintain the state of the signaling sequence initiated by a user agent. Furthermore, a stateaware proxy server can fork requests to multiple destinations where a given user agent may have registered. This feature is useful in scenarios where the same user has two work locations [Zha02]. A redirect server helps locating user agents by providing alternative locations where the user may be found [Cam01].

The last entity of a SIP network is the location server. Location servers are not necessarily SIP entities per se since they do not communicate using SIP. Location servers are normally contacted by proxy servers where the interface between the proxy and location servers is not defined in [RSC+02] and it is normally dependent on the type of the database storing the information at the location server. The database normally includes information such as URIs, IP addresses, scripts and possible routing information [Cam01, SJ01].

2.5.1 SIP Signaling

SIP signals are either requests, known as *methods*, or *responses* to these requests. Upon initialization, a user agent registers with a registrar server which in turn informs the location server of the newly joining user agent. If the user agent changes its point of attachment to the network, then a new `REGISTER` request is sent to the registrar server which, in turn, updates the location server. To find the registrar, the user agent sends a multicast `REGISTER` request to a well known multicast address, contacting the home domain registrar [SW00] or using the Service Location Protocol [GPVD99]. Figure 2.5 shows a user registering initially from home, then after changing his point of attachment to the network, he re-registers from his office.

Now, let us consider an example that is similar to the one available in [SW00]. Alice and Bob are two users who are registered at two different SIP servers and Alice wishes to call Bob. Figure 2.6 shows different scenarios depending on the final outcome of the call. First; `alice@abc.net` sends an `INVITE` request to the local SIP proxy server (e.g., `sip.abc.net`) which - in turn - uses DNS SRV [GV96] to

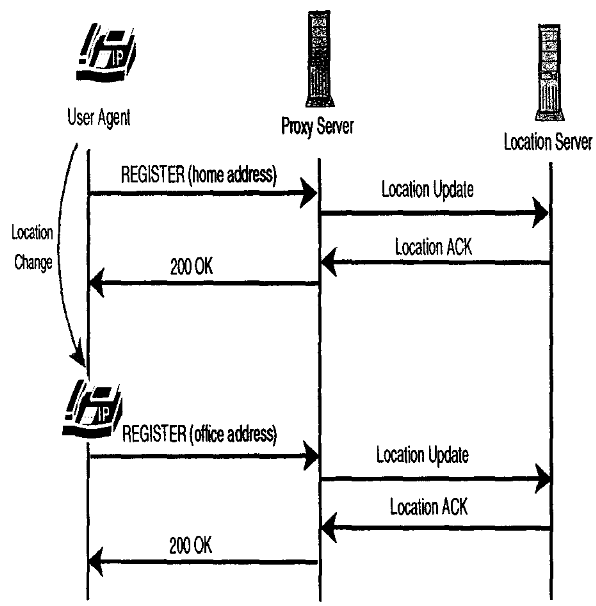


Figure 2.5: *SIP* ***REGISTER*** *Method*

locate the SIP proxy server at the `xyz.org` domain (e.g., `sip.xyz.org`) and forwards the `INVITE` request to it. Now, the `sip.xyz.org` forwards the `INVITE` request to bob@xyz.org based on the registration information available at the local location sever of the user Bob, namely: host `bob@xyz.org`. At this stage, several scenarios can happen. Figure 2.6 shows Bob accepting the call, where a `200 OK` response is sent to `sip.xyz.org` which is relayed back to `sip.abc.net` and in turn to `alice@abc.net`. Upon receiving this response, Alice `ACK`s the `200 OK` response directly (i.e., without going through the intermediate SIP servers) and the media stream flows end to end using the Real-Time Protocol (RTP) [SCFJ03]. However, if Bob is either busy or declines to accept the call then a `486 Busy Here` or `603 Decline` response is issued

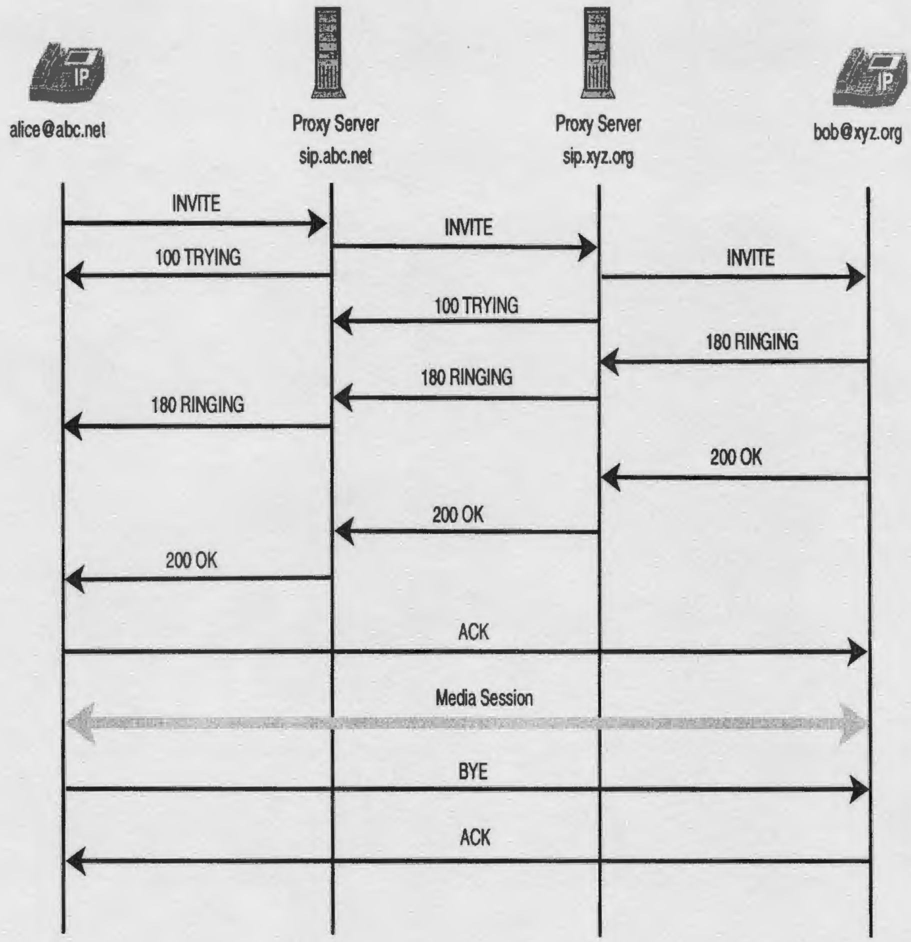


Figure 2.6: *Session Establishment Example with Intermediary Proxy Server*

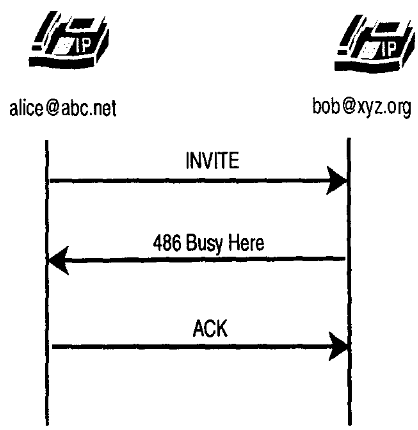


Figure 2.7: *Session Establishment with the Busy Receiving Party*

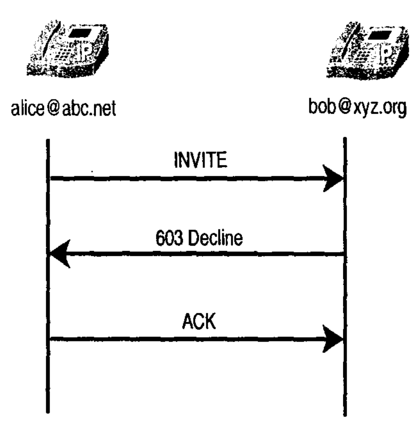


Figure 2.8: *Session Establishment where Receiving Party Declines Call*

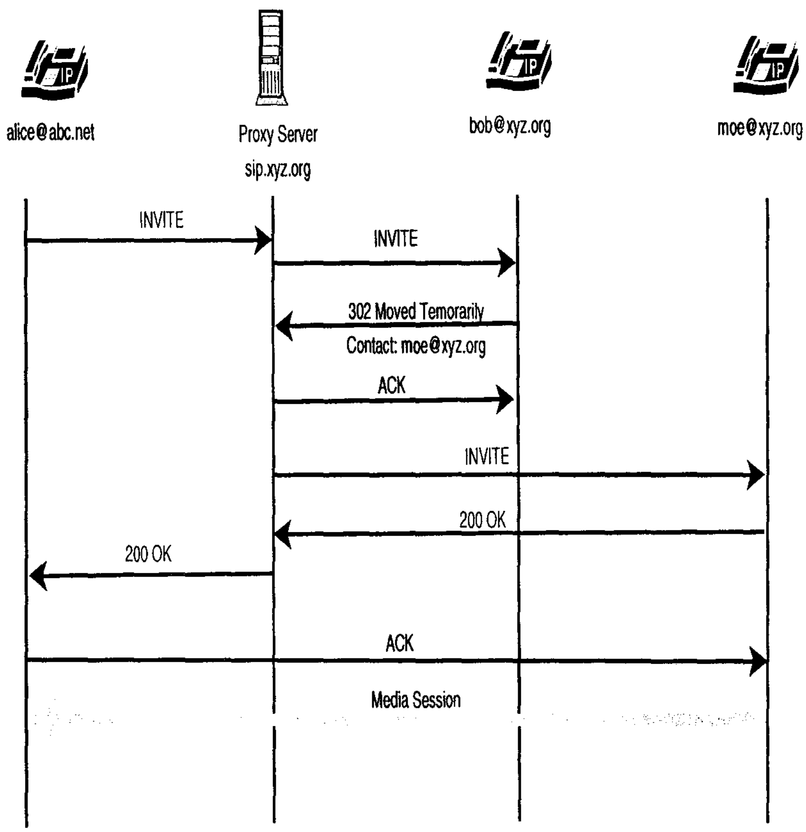


Figure 2.9: *Session Establishment with Third Party Call Forward*

and, in turn, ACK'ed by Alice. These two scenarios are shown in Figures 2.7 and 2.8 respectively with the intermediate SIP servers and provisional responses deleted for the sake of brevity . Bob also has the option of forwarding his incoming calls to moe@xyz.org by responding with a 302 Moved Temporarily message including Moe's address. The sip.xyz.org server will use Moe's address and perform a location look up by contacting the local location server, then it forwards the call to moe@xyz.org transparently without Alice's involvement as shown in figure 2.9.

The parameters characterizing the media session are described using the Session Description Protocol (SDP) [HJ98]. Tables 2.2 and 2.3 summarize the most commonly used SIP requests and responses respectively.

Table 2.2: SIP Methods

Method	**Function**
INVITE	Session setup
ACK	Acknowledgment of user agent server responses
BYE	Session termination
REGISTER	Location registration
CANCEL	Cancellation of a pending session
REFER	Session transfer
NOTIFY	Notification of subscribed events

Table 2.3: SIP Responses

Class Code	Function
1xx	Provisional or Informational: request still in progress
2xx	Success: Request completed successfully
3xx	Redirection: Request to be tried somewhere else
4xx	Client Error: Request was not completed due to error in request, can be retried when corrected
5xx	Server Error: Request was not completed due to error in recipient, can be retried when corrected
6xx	Global Failure: Request has failed and should not be retried again

2.5.2 Refer Method

As mentioned earlier, SIP signals are categorized as either methods (i.e., requests) or responses to these requests. Of special interest to us is the the REFER method [Spa03, SJP07]. The REFER method indicates that the recipient of the REFER request should contact a third party using the contact information provided in the request [Spa03]. The REFER request is a part of a continuous development effort to add call control features to SIP. Examples of such features include headers and mechanisms extending the capabilities of the REFER request and other requests [MBD04, Spa04], multiparty call control framework [Mah03] and conferencing [Ros06, JL04, RSL06].

As will be evident in later sections, the REFER request is instrumental in the proposed signaling sequence and in the design of the device coordinating the handover between the interfaces of the dual-mode handset. Assuming that Alice from the previous example decides to transfer Bob to Moe, then Alice sends a REFER request to Bob with Moe's contact Uniform Resource Identifier (URI) contained in the new Referred-to header which will be used by Bob to construct an INVITE request which will be sent to Moe. Figure 2.10 displays the signaling used to complete the REFER procedure. Notice that when Bob receives the REFER request, he will send a 202 ACCEPTED response before transmitting the INVITE request to Moe. Moreover, Bob uses a NOTIFY request to inform Alice of the outcome of the refer process. An optional header called the REFERRED-BY header was proposed in [Spa04] and it can be included in either of

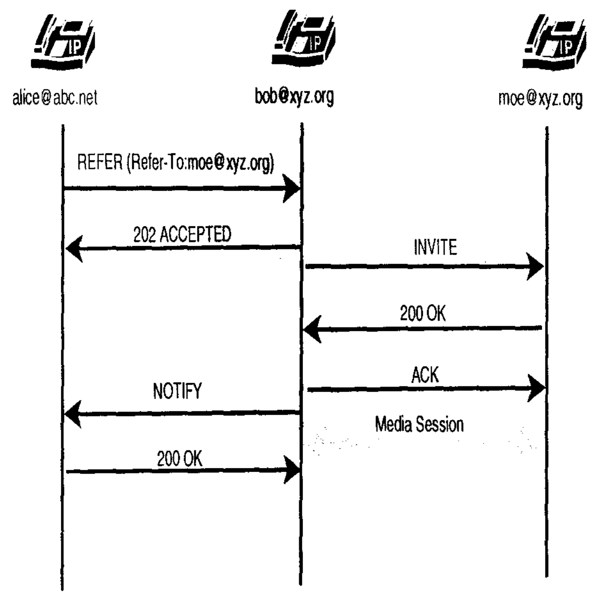


Figure 2.10: *Call Transfer using* REFER Request

the REFER or INVITE requests, giving Bob and Moe extra information which could be useful in processing these requests. Usually, but not necessarily always, after a dialog has been established between Bob and Moe, either Alice or Bob will send a BYE request to terminate the old session which is being placed on hold. In case that a BYE was not sent by either party and one of the parties would like to reactivate the old session (e.g., Moe declines to take the new call), then the session can be re-activated by sending a re-INVITE by either Alice or Bob provided that the receiving party responds with a 200 OK to the re-INVITE.

2.5.3 SIP Programmable Services

Enhancing the features set delivered by internet telephony services requires versatility and custom-programmability of IP-Telephony enabled networks. Several publicly available programming tools exist allowing for service creation and expanding the functionality of SIP enabled entities such as Call Processing Language [LWS04], SIP CGI [LSR01], VoiceXML [voi08] and JAIN [jai08]. Also some SIP device providers have their own SIP-programming tools such as SIP Express Router [ser08] and Asterisk Open Source PBX [Dig] with their Asterisk Gateway Interface (AGI) scripting tool.

Figure 2.11 borrowed from [RLS99] shows the basic model for providing logic for SIP services. In this specific example, the logic is augmenting the SIP server which is not necessarily always the case since similar logic can be placed at the terminal SIP devices. As will be shown later, we benefit from SIP's programmability by implementing some of the entities that coordinate the proposed signaling sequence.

2.6 Vertical Handover in Wireless Networks

Cellular networks support inter-cell roaming when a cellular user crosses the boundary of one cell and joins another cell. Similarly, WLAN ESSs provide the same functionality when a WLAN mobile user departs one BSS and joins another. In both instances,

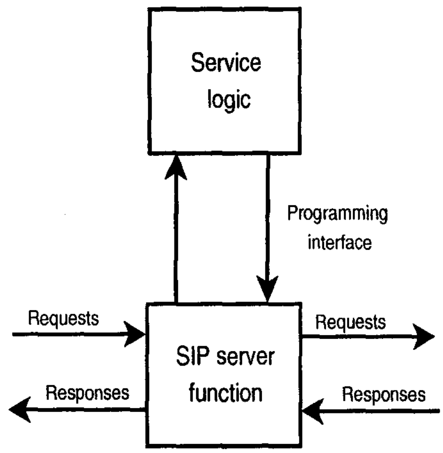


Figure 2.11: *SIP Server Programmability Model*

the transition happens seamlessly and without end-user involvement. This type of handover is referred to as a Horizontal Handover (HHO) since the original network and the destination network (i.e., before and after the handover) are operating with the same technology. However, handsets are emerging with multiple interfaces such as cellular and WLAN, to capture the high data rates whenever WLAN coverage is available, and falling back on the cellular network with lower data rates when the WLAN coverage is not available. The procedure of supporting a mobile user who is departing one network and joining the other network is referred to as a VHO since the technologies in both networks are different.

2.6.1 Vertical Handover Literature Review

Some of the first work on vertical handover (VHO) was done in [SK98]. The proposed scheme makes use of a multicast address in the mobile host which receives advertisements from potential access points in an overlay. Handover initiation relies on the detection of periodic beacons from the different networks, and using this approach it was shown that handover latencies can easily be as high as 3 seconds. Fast beaconing and packet/header double-casting can reduce this delay to 800 msec [SK98].

In [PKH+00], an investigation of VHO in hybrid wireless networks with non-real-time traffic is given. The paper classifies VHO issues into architectural choices and decision algorithms, and summarizes classical handover decision algorithms. The authors make a clear distinction between the WWAN-to-WLAN and WLAN-to-WWAN handover and indicate that WLAN-to-WWAN poses a stricter time constraint since very short warning is given due to severe WLAN degradation in transition regions.

In [BCI04] an experimental testbed is used to evaluate the performance of vertical handovers between GPRS, WLAN, and an Ethernet LAN. Mobile IPv6 (MIPv6) is adopted as the mobility management scheme and the performance of User Datagram Protocol (UDP) and Transport Control Protocol (TCP) flows have been evaluated. The authors show how differences in network link characteristics during vertical handovers can produce severe performance problems for TCP flows. Large vertical handover delays (4-5 seconds) were experienced when moving between WLAN and

GPRS networks. Early detection of network coverage loss using link layer events has reduced VHO delay to nearly 2 seconds. The results presented in [BCI04] show the performance inefficiency of VHO and the importance of utilizing link layer triggers to provide earlier triggering.

In [HHSDH04] a smooth VHO scheme using pre-authentication and pre-registration was proposed for WLANs tightly-coupled to a UMTS 3G cellular network. Pre-registration is a Mobile IP-based (MIP) fast handover scheme that triggers MIP handover before link layer handover, thereby limiting packet loss and handover delay to that caused by link layer handover [Mal07]. Moreover, by having the old AP buffer packets during handover, packet loss is further reduced. This forwarding mechanism is reasonable since the APs involved are typically separated by a small number of forwarding hops. However, this assumption is usually not valid in loosely-coupled WLAN/cellular networks. Another study [CVS+04] has experimented with a loosely-coupled MIPv6-based GPRS-WLAN testbed and has investigated the impact of VHO on TCP connections. Their experiments show a 3.8 seconds VHO delay, and that by using fast router advertisements (RA), RA Caching, and binding update simulcasting, it can be reduced to about 1.36 seconds. In [BDA05], the authors propose a SIP based architecture that supports soft handover for IP centric wireless networks while maintaining quality of service for packet loss and the end-to-end delay jitter.

Their approach may be more scalable in next generation networks as cellular IP centric wireless networks (e.g., GPRS) are incapable of supporting voice sessions at the current data rates. [CDM04, CVS+04, ZM04, BCI04] continue to address the issue of the handover between IP-based packet-switched over WWAN interface and WLAN interface. In [CDM04, ZM04] the authors propose a new approach in taking VHO decisions, which is not exclusively based on the knowledge of the available networks but also on higher level parameters such as user preferences which capture the trade-offs between cost and quality of service. [CVS+04] presents an experimental study of a loosely-coupled, MIPv6-based, GPRS-WLAN inter-network mobility, and analyze the effects of handovers on active TCP flows, and propose a number of network-layer handover optimization techniques, such as RA and RA caching.

Although many integrated WLAN and WWAN chipsets already exist, no standard mechanism exists today for real-time handover between the WWAN plain circuit-switched telephony and WLAN. The proposals presented so far address the handover between IP-based packet-switched over WWAN interface and WLAN interface which is fundamentally different from the problem we are tackling since the IP-based WWAN technology is incapable of supporting voice telephony at the current data rates and since plain circuit-switched telephony over WWAN is the current standard for WWAN voice telephony. Furthermore, the proposals for handover between IP-based packet-switched over WWAN interface and WLAN interface are not applicable to the handover between plain circuit-switched telephony WWAN interface and the WLAN interface because of the fundamental differences between circuit-switched and packet-switched WWAN technologies. IEEE 802.21 has recently developed a framework for inter-network handover which defines a media independent handover layer in the handset. This function coordinates multiple interfaces and provides a consistent handover view, regardless of the underlying physical layers.

Vertical handover triggering plays a very important role in the performance of any VHO procedure. The VHO process can be broken down into three phases: metric-collection, handover decision and handover execution. Some VHO triggering techniques, such as those described in [BHJ03, CSC+04, ZM04] formulate an optimization problem in terms of power consumption, outage, packet loss, load balancing and signal strength. Other VHO decision algorithms interpret the problem in terms of mobile user localization [KTZ04, LSB+04, BSL05]. In our work, we consider simple threshold based VHO triggering, where a VHO is triggered if the WLAN RSSI crosses a specified threshold S_{tp}. Threshold based VHO triggering is sufficient for us to illustrate the behavior of the WLAN signal in indoor-to-outdoor transition regions, and it remains by far the most commonly used method in practice.

Different WLAN deployment strategies have been discussed in the literature[Hil01, Pra00, SKK01]. Deployment strategies can try to optimize range or throughput, which are normally two conflicting requirements. In high capacity deployments, the

AP density is normally high and provides a strong average RSSI. In contrast, coverage deployments have sparse AP deployment structure with every AP covering a large service area and thus providing a low average RSSI at the receiver. In [Hil01] a combination of access point positioning, frequency assignments, and receiver threshold settings is used to develop campus-wide WLAN network design schemes. Another example is given in [SKK01] where WLAN support for VoIP traffic is considered. These deployment techniques are mostly based on well known propagation models such as the ones in [Rap96, ZVP03] and on studies assessing WLAN link variations such as [SKBMM06, GS06].

2.7 Conclusions

This chapter introduced cellular and wireless local area networks and provided an overview of today's wireless transmission technologies. An example gateway integrating a VoIP-capable enterprise network and the PSTN was used to illustrate the inter-networking of VoIP and SIP signaling. Finally, a literature review of the vertical handover problem was presented with a comparison to the horizontal handover problem.

Chapter 3

Vertical Handover Building Exit Data

3.1 Overview

In this chapter we report on a measurement-based study of WLAN-to-cellular handover. Our results are based on extensive IEEE 802.11 measurements that were made on the McMaster University campus. In Section 3.2 the procedure of performing vertical handover in loosely-coupled systems is described and parameters required to calculate vertical handover success probability are introduced. In Section 3.3 we describe our measurement methodology which involved traversing many indoor-to-outdoor paths for a large number of campus buildings and exits while monitoring multi-AP WLAN coverage. In Section 3.4 the collected data is processed to determine the probability of seamless handover using classical vertical handover algorithms. In Section 3.5 we propose a trigger node that extends the WLAN coverage to help the MS complete the handover and show that significant improvements are made possible using this approach.

3.2 Loosely-Coupled Vertical Handover

WLAN propagation is an important factor in determining the success probability of WLAN-to-cellular handover in the loosely-coupled integration model. In Figure 3.1 an example is shown of a dual-mode handset that is operating in a loosely-coupled system with interfaces that can access WLAN and cellular infrastructure. In the example shown the dual-mode handset initially establishes a call to a PSTN destination. This call is routed through an IP-based enterprise PSTN gateway/PBX, which establishes the call to the end destination. The media path is shown, consisting of two legs, numbered 1 and 2. Some time later the dual-mode handset moves out of the WLAN and into cellular coverage. At this time a new call leg is created between the cellular PLMN/PSTN and the GW, which replaces the original WLAN call leg 1. After the handover the new media path consists of call legs 2 and 3 as shown in Figure 3.1.

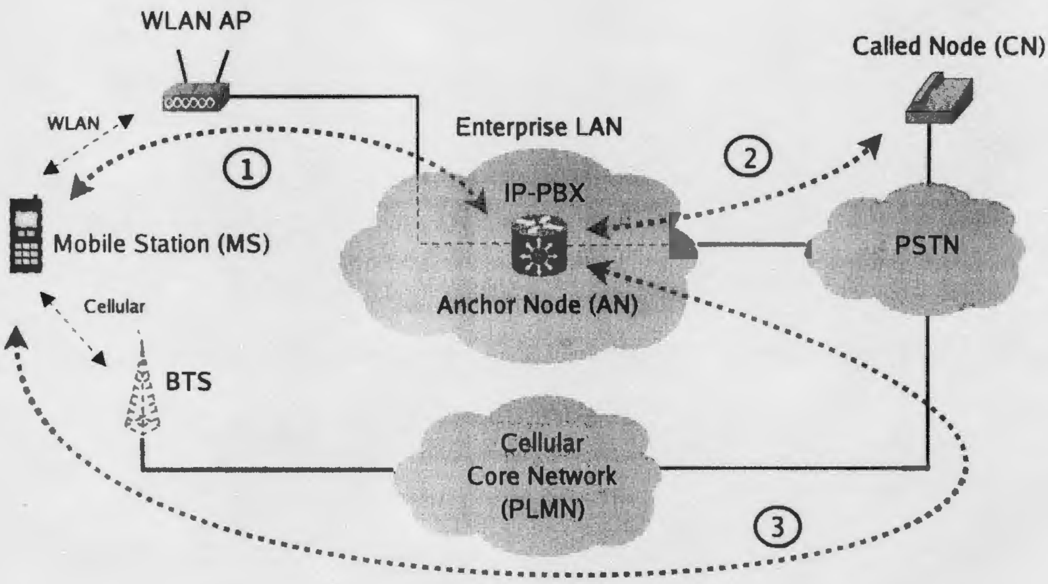


Figure 3.1: *Enterprise PSTN GW Anchored Vertical Handover*

The handover from call leg 1 to call leg 3 in Figure 3.1 must be completed before the WLAN link is lost. Due to the vagaries of WLAN propagation however, it is possible for the link loss to occur very quickly. This is exacerbated by the fact that the cellular call leg establishment can often take between 5-10 seconds or longer, which increases the frequency with which outage can occur [STZK06].

We define a successful seamless VHO as one where the triggering and establishment of the cellular call leg is completed before the WLAN link is lost, and hence the success probability of VHO is given by

$$P_{success} = P\{t_{margin} > t_{vho}\}. \tag{3.1}$$

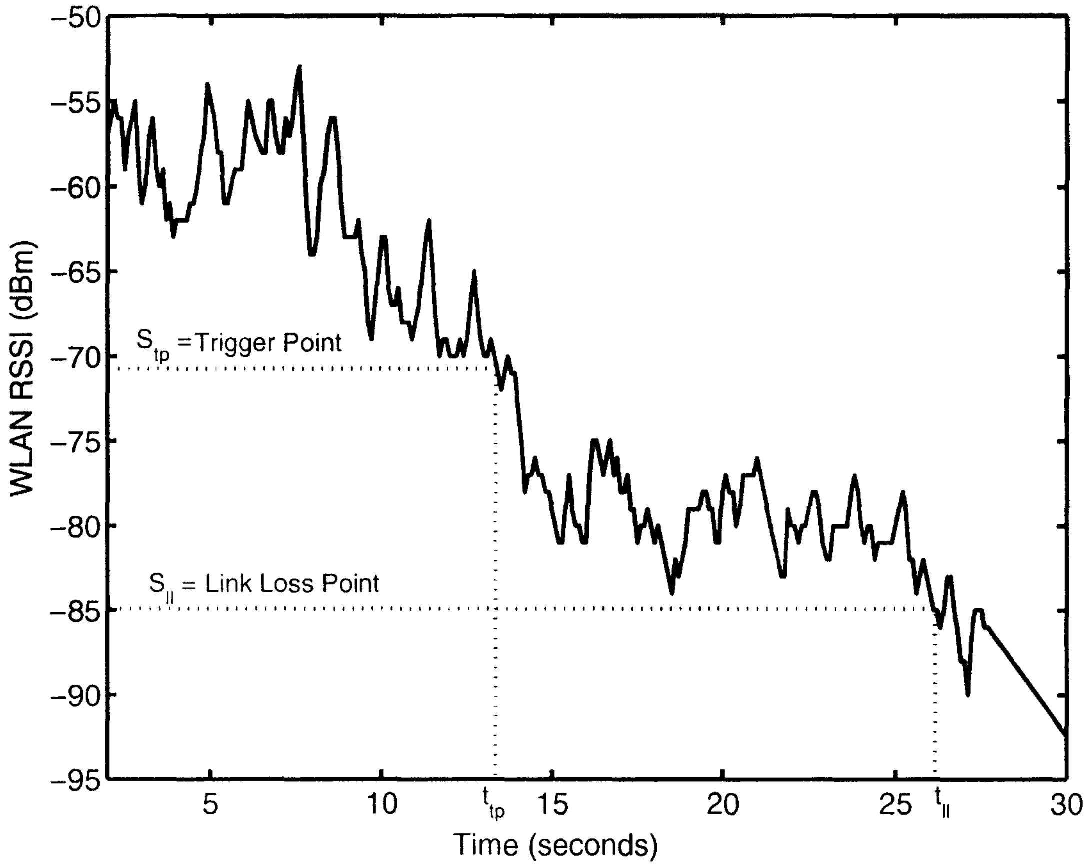


Figure 3.2: t_{margin} *Illustration*

In this expression, t_{vho} is the time required to perform the VHO and $t_{margin} = t_{ll} - t_{tp}$ is the time difference between the VHO trigger and the time at which the WLAN link is lost. An example of this is shown in Figure 3.2. This figure is a measured single path example of an indoor-to-outdoor mobility path taken from our results. The figure shows the smoothed (IEEE 802.11g) WLAN link quality (RSSI in dBm) as a function of time when the dual-mode Mobile Station (MS) is moving out of

WLAN coverage at a constant speed. At $t = t_{tp}$ we assume that the smoothed RSSI has dropped below a trigger threshold, S_{tp}, and the VHO is triggered. At $t = t_{ll}$ the WLAN link has then dropped to a value, S_{ll}, that is insufficient to reliably carry the connection.

Threshold-based handover triggering remains by far the most common mechanism for initiating VHO, and we will assume this in the remainder of this thesis. It should be noted that although more efficient triggering algorithms may be theoretically possible [BHJ03, CSC$^+$04, ZM04], they are rarely used in practice due to their complexity.

In a loosely-coupled system the user has no control over cellular call establishment latency, and thus to improve $P_{success}$ one must increase t_{margin} by either triggering the handover earlier, or by holding the link to lower RSSI values. In the former case this may be done by using a higher signal strength triggering threshold, S_{tp}. Unfortunately, there are strict limitations to this approach which are dictated by the nature of the WLAN deployment.

WLAN coverage is normally deployed based on a target minimum guaranteed signal strength, S_{min}. This means that throughout the coverage area of interest, the link quality will exceed S_{min} with a high probability. Due to the automatic link rate-adjusting in IEEE 802.11, this value of S_{min} translates into a minimum link rate for which the system was designed. When APs are sparsely deployed, then the value of

Table 3.1: Handover Path Parameter Definitions

t_{vho}	Vertical handover execution time.
S_{ll}	Signal strength at WLAN link loss point (dBm). Occurs at t_{ll}.
S_{tp}	Signal strength at VHO trigger point (dBm). Occurs at t_{tp}.
t_{margin}	Maximum time available to execute VHO. ($= t_{ll} - t_{tp}$)
S_{min}	Minimum guaranteed WLAN signal strength (dBm).

S_{min} will tend to be low, resulting in low average link rates. Conversely, when APs are densely deployed, S_{min} will tend to be much higher resulting in higher link rates and much higher network capacity.

When a VHO is to be triggered, the RSSI trigger threshold, S_{tp}, must be strictly less than S_{min}, i.e., $S_{tp} < S_{min}$, otherwise the *false handover* rate may be unacceptable. False handover occurs when VHO is triggered unnecessarily, resulting in a cellular call leg while WLAN coverage is still sufficient. Obviously, if S_{tp} exceeds S_{min} then this threshold may be crossed frequently while the MS is inside the WLAN hotspot, and the false handover rate may be unacceptable. From the above discussion it is clear that a dense AP deployment will permit higher values of S_{tp}, thus improving the probability of successful handover.

The other way to improve $P_{success}$ is to allow the mobile station to hold the WLAN link as long as possible. When this is done in IEEE 802.11, the link will be rate-adjusted to lower bit rates as the link quality drops. But allowing stations to drop to low bit rates can have a negative impact on system capacity, and is especially undesirable in high capacity WiFi deployments. For this reason procedures are now being used to prevent links from dropping below some preset rate limit [IEE97]. This is done by having the AP advertise restricted IEEE 802.11 basic service set rates and requiring MSs to handover rather than to drop below the set limits, otherwise the link is terminated. These mechanisms create an artificially high value of S_U which makes VHO much more difficult.

3.3 WLAN Measurement Methodology

The results to be presented are based on data which was collected across a total of 7 buildings and 17 building exits on the McMaster University campus. The sampled buildings included those constructed over a period of about 100 years. For each building the team placed multiple Cisco Aironet 802.11b/g APs in different ground-floor locations after setting all of them to operate on the same frequency. No specific algorithm was used to place the WLAN access points, however care was taken to distribute the APs as uniformly as possible across each building floor. In the absence of accurate characterization of the propagation environment and the distribution of end-client locations, uniformly distributing the APs removes any bias in the WLAN signal strength over the observed paths. A point was chosen well inside each building as a starting point for paths which extended to points outside the building using the available exit doors. Upon leaving a building exit, different paths were taken at different radial angles moving away from the exit door.

Figure 3.3 shows the paths associated with one of the building exits (the ITB Annex Building) where data was collected. One of the starting points for the paths is marked in the figure as the point labeled "**A**". Nine different paths emanating from this point and leaving the building exit are shown labeled with their path number at the top of the figure. When a measurement run was taken a laptop would be carried along a given path while recording RSSI measurements from all of the APs. Since the APs were all locked to the same frequency, joint AP RSSI values were obtained while traversing each path based on receiving 100 msec beacons, and were later smoothed using a 5 sample running average. To calibrate the measurements, distances were marked on the ground at 2 meter intervals and whenever one of these markers was reached the data collected was time stamped. This allowed us to correlate the physical location with the recorded measurements. The interval markings for the paths in Figure 3.3 are shown as **X**'s in the figure. Each path was traced 5 times during the data collection phase. The measurements were taken with IBM Thinkpad laptops equipped with Cisco Aironet 802.11b/g cards running in monitor mode. It is important to note that our measurements were made at the output of the client WLAN cards, and thus the reported RSSI is subject to any smoothing such as that obtained via antenna diversity. This is desired since it is indicative of the signal strength that is used to set the link data rate. Also we identify a point called the

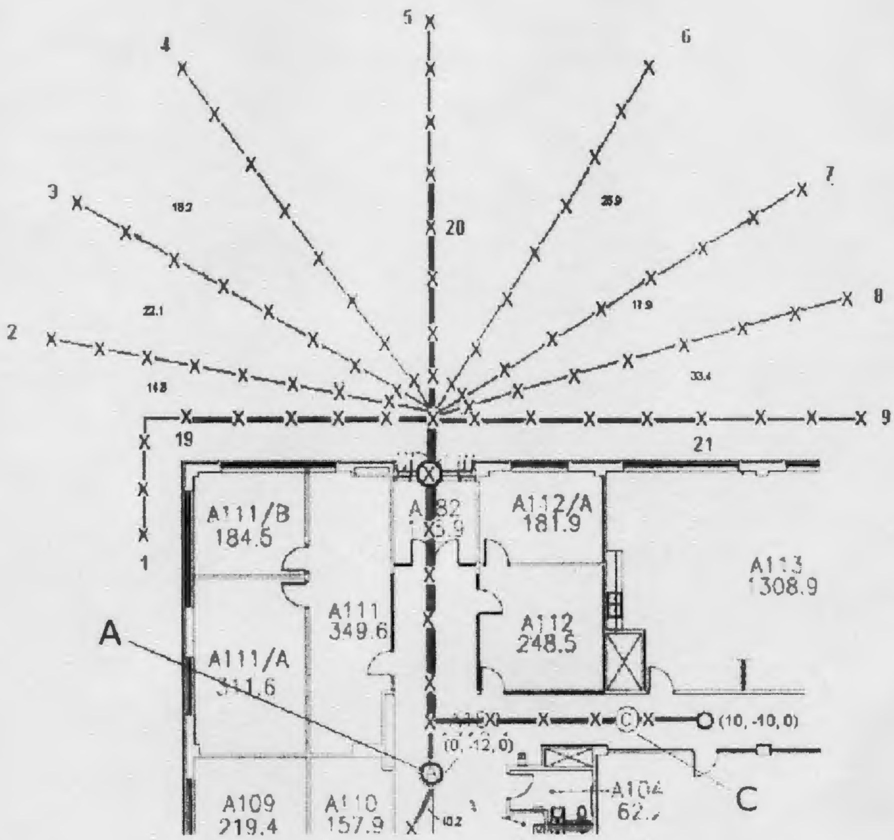


Figure 3.3: *Example Path Map for one ITB Building Annex Exit*

Point of Departure, or PoD, just inside the exit doorway. In Figure 3.3 the PoD is shown at the exit as an X enclosed by a circle.

3.4 Vertical Handover Results and Discussion

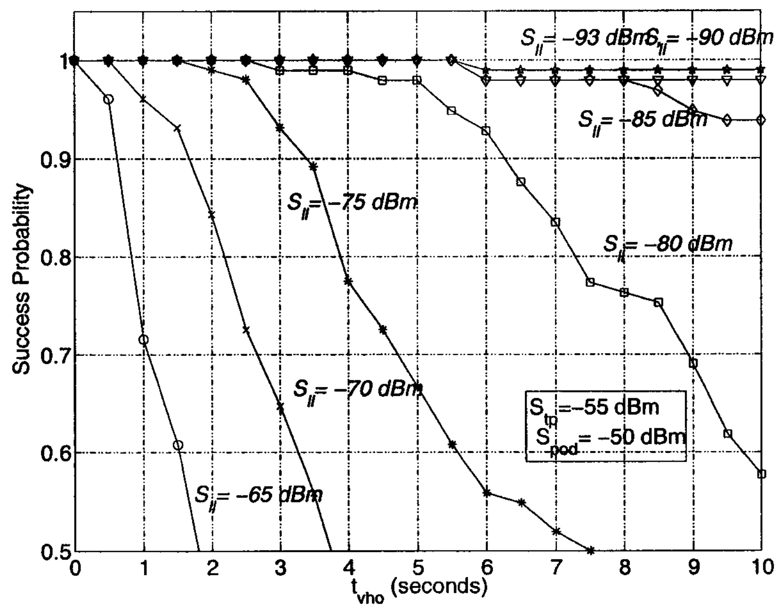
Using the collected data we have performed many experiments to characterize the potential handover performance in this difficult situation. In this section we will show representative examples of the results that we have taken.

When an MS has a voice connection in WLAN coverage, the station continuously measures the quality of the downlink whenever voice packets are received. In addition it may periodically test for better candidate APs which would result in an improved link. This searching may be performed proactively or reactively based on a preset scan threshold. Once the handover threshold has been reached the MS would aggressively search for new candidate APs and if found, would attempt a horizontal handover (i.e., an AP-to-AP handover) as defined in IEEE 802.11. It is normally the case that the MS applies some hysteresis before executing a HHO to prevent ping-pong effects, i.e., continuously performing HHO between APs with similar link quality. After the above procedures, if a candidate AP is not found the MS would then initiate a vertical handover.

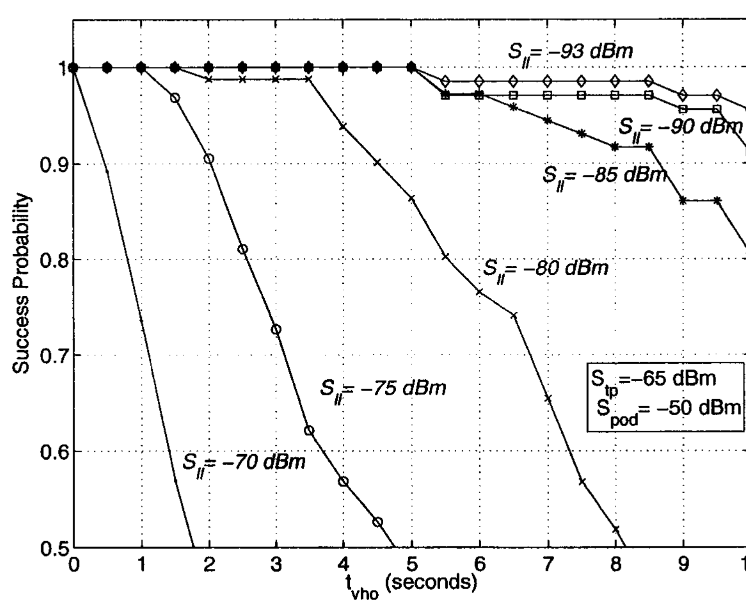
It is clear that the longer the HHO process is, the shorter the time that is left to perform VHO, however with proactive scanning and neighbor lists, this time can be reduced to very short values [SMA04, MSA04]. For this reason we will make the assumption that the HHO attempt occurs very quickly compared with the time needed to execute the vertical handover. The results will then provide an indication of what can be achieved using state of the art scanning procedures. As previously discussed we assume that the MS uses a sliding window average over the received RSSI measurements. Once this drops below the trigger threshold, S_{tp}, the VHO is initiated. If the WLAN RSSI drops below S_{ll} or MAX_MISSED_BEACONS or more consecutive beacons have been lost, before the cellular call leg has been established, the VHO is assumed to have failed.

3.4.1 Call Establishment Time and VHO Success Probability

In the first set of results, we examine the vertical handover success probability assuming that the MS moves at a constant speed from inside the building to the outside along our collected data paths. Rather than presenting aggregate results for all building exits, it is more useful to partition our set of paths into sets corresponding to the same link quality at the point where the MS leaves the building. At the PoD we also define the link quality for a particular path run, denoted by S_{pod}, i.e., S_{pod} is the link signal strength (in dBm) at the point where the MS is just about to leave the building. In our results we then sort all collected paths into bins with similar values of S_{pod} (within 5 dBm). For example, all the curves with S_{pod} in the range of $[-61, -65]$ dBm were classified as $S_{pod} = -65$ dBm. In this set of results we also assume various values for vertical handover time, t_{vho}. The testbed reported on in Chapter 4 and [STK+] collected data for cellular call establishment times based on results collected using the Canadian Rogers PLMN (GSM/GPRS) network and show that the latency is almost deterministic when calls are being made to cells within the same area as the PSTN destination.



(a)

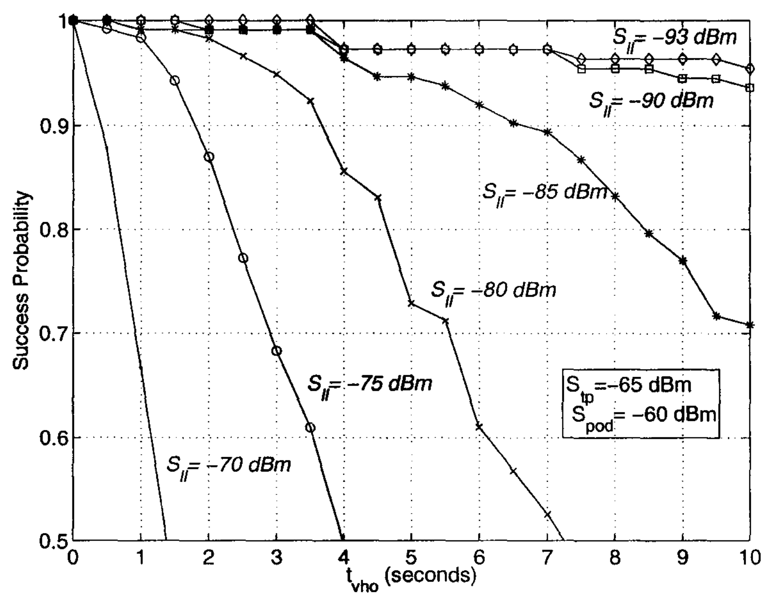


(b)

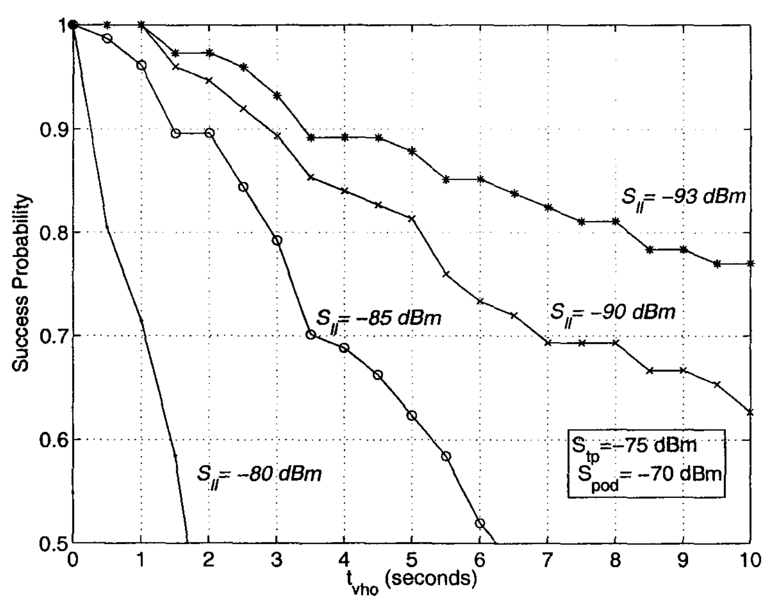
Figure 3.4: *$P_{success}$ as a Function of t_{vho} for $S_{pod} = -50$ dBm*

The graphs in Figure 3.4 show how $P_{success}$ varies with t_{vho} for different values of link loss threshold, S_{ll}. These curves correspond to an S_{pod} of -50 dBm. Figure 3.4(a) shows the case where the trigger threshold, S_{tp}, is set to -55 dBm and Figure 3.4(b) corresponds to the -65 dBm case. An S_{pod} of -50 dBm corresponds to a strong signal strength at the point of departure, and the results show that reasonable handover success can be obtained. A striking feature of the curve however, is that even for this high S_{pod}, the only way to ensure good VHO performance is to permit the MS to hold the link down to very low levels of signal strength, i.e., S_{ll}. Figure 3.4(a) shows that an installation that restricts the use of data rates to those corresponding to a S_{ll} of -75 dBm for example, cannot achieve a VHO success probability greater than 90% when t_{vho} is about 6.5 seconds. This range of cellular establishment time is typical of the results taken in [STK$^+$]. The situation is far worse in Figure 3.4(b) which shows results where the trigger threshold is 10 dB lower. This corresponds to a situation where S_{min} is smaller than in the previous case, thus preventing a higher handover trigger threshold. It can be seen that in this case the only hope for high VHO success (using the same t_{vho}) is to allow the link to drop to very low values of SNR, e.g., in the -90 dBm range. Permitting stations in general to hold the link to these values of SNR (by including them in the supported rate set) can result in a severe loss of WLAN capacity.

Now we consider Figure 3.5 where we have used data corresponding to paths with



(a)



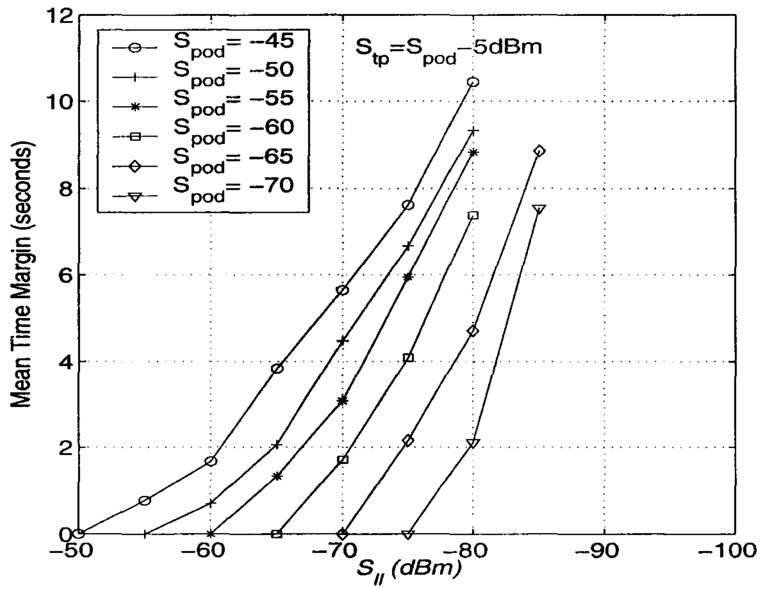
(b)

Figure 3.5: $P_{success}$ as a Function of t_{vho} for $S_{pod} = -60$ dBm, and -70 dBm

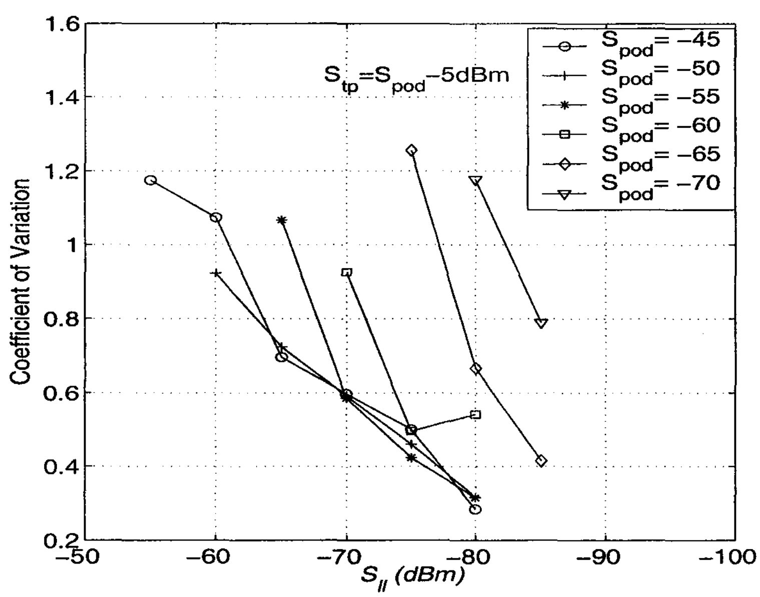
even lower values of S_{pod}, i.e., -60 dBm and -70 dBm. These results consider cases which are more typical in sparse AP deployments, where S_{pod} is likely to be much lower in general. Figure 3.5(a) shows the case where S_{tp} is set to -65 dBm and Figure 3.5(b) shows the -75 dBm case. In the former case a VHO requiring more than about 4 seconds to complete will suffer failure rates exceeding about 5% even if the WLAN link is held close to the noise floor. In the latter case nearly all VHOs requiring more than one second to complete would fail regardless of the S_{ll} value. For the cellular latencies found in [STK$^+$], deployments corresponding to these graphs make successful VHO virtually impossible.

In the graphs shown in Figures 3.4 and 3.5, the slope of each curve shows the sensitivity of $P_{success}$ to t_{vho} for the corresponding value of S_{ll}. As S_{ll} is chosen to be larger and closer to S_{pod}, $P_{success}$ becomes more sensitive to t_{vho} agreeing with the conclusions drawn from Figure 3.6(b) which shows that the coefficient of variation for t_{margin} decreasing as we move toward lower values of S_{ll}.

Interestingly, $P_{success}$ stays relatively constant up to a certain *pinch-off* value of t_{vho}, after which it starts to quickly decrease. The pinch-off point increases as S_{ll} is chosen to be smaller and further away from S_{pod}, effectively increasing the distance that the mobile user has to travel before losing the link. For instance, Figure 3.4(a) shows that at $S_{ll} = -80$ dBm, the success probability stays almost constant and very close to 100% for vertical handover latencies of up to 5 seconds compared to



(a)



(b)

Figure 3.6: *Mean and Coefficient of Variation for t_{margin} for Different S_{pod}*

8 seconds for $S_{ll} = -85$ dBm. One can think of the pinch-off point as where the tightest guarantees on $P_{success}$ can be given. These observations regarding $P_{success}$ can be explained by restating that the pdf of t_{margin} corresponding to larger S_{ll} values has a larger coefficient of variance and a smaller mean than the pdf's corresponding to smaller S_{ll} values. The smaller mean of t_{margin} associated with larger S_{ll} values explains the smaller pinch-off value. The higher rate of decrease of $P_{success}$ associated with larger S_{ll} values is due to the higher coefficient of variance at these values of S_{ll}.

3.4.2 Link Quality in Indoor/Outdoor Transition Areas

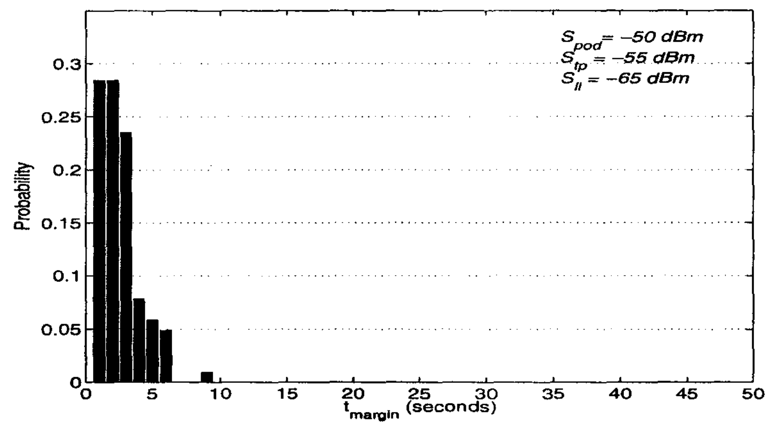
VHO success probability is dependent on the distribution of t_{margin} as discussed in Chapter 1. Figure 3.6(a) shows the dependency of t_{margin} on the link loss point for different values of S_{pod} in the case when S_{tp} is set to be 5 dB below S_{pod}. This corresponds to the most optimum situation from a triggering viewpoint since the trigger threshold is very close to S_{pod}, thus giving the most time for executing the VHO.

The slope of each line segment indicates the sensitivity of t_{margin} to S_{ll}. We see that the first 5 dB segment has the steepest decay for all of the S_{pod}'s when compared to the following two segments. This steep drop corresponds to an abrupt signal strength degradation in the first 5 dB segment after exiting the building. The following two segments have progressively decreasing decay rates implying that the signal strength degradation continues however at a rate that is less than that in the first 5 dB segment. This evident decrease in the signal strength degradation rate becomes more subtle after dropping 15 dBs below the S_{tp}. These observations can be explained using simple propagation models such as exponential path loss models. If we consider the case where the propagation exponent is 2 (i.e., free-space propagation), a 6 dB loss between point A and point B implies that the distance from point B to the transmitter is twice that of point A. Similarly, if point C is two times further away from the transmitter than point B (i.e., four times further than point A), then point C will be at 6 dB loss when compared to point B. So at constant pedestrian speed, going through the first 6 dB loss from point A to point B will take half the time it would take a pedestrian to experience the same loss while walking from point B to point C. To understand this behavior of the signal in the context of VHO, we define the window $S_{diff} = S_{tp} - S_{ll}$. The placement of S_{diff} (i.e., choosing the position of S_{tp}) is very important to the performance of VHO, since having it where the signal degradation rate is high, implies that it will take the pedestrian a shorter amount of time to experience a loss of S_{diff} dBs than if the same width S_{diff} has been placed in a region with lower signal degradation rate.

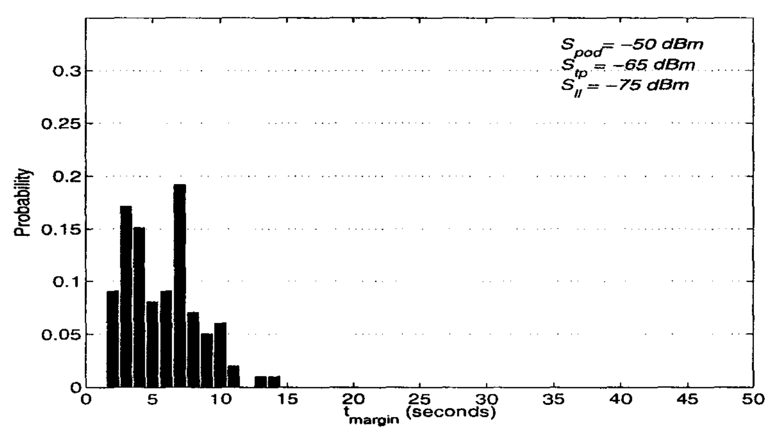
In Figure 3.6(b) we show the coefficient of variation of t_{margin} which is the ratio of the standard deviation to the mean. For wider S_{diff} windows associated with S_{ll} values set much smaller than S_{tp}, the mean of t_{margin} increases. Also as the

S_{diff} window size is increased by setting S_{ll} to smaller values which are further away from the PoD, the WLAN signal propagates over longer distances, experiencing more shadowing and consequently leading to t_{margin} having higher variance for these smaller S_{ll} values. However, this increase in t_{margin} variance is offset by the increase in its mean, explaining the decrease in the coefficient of variation as S_{ll} is chosen smaller and further away from the PoD as observed in Figure 3.6(b).

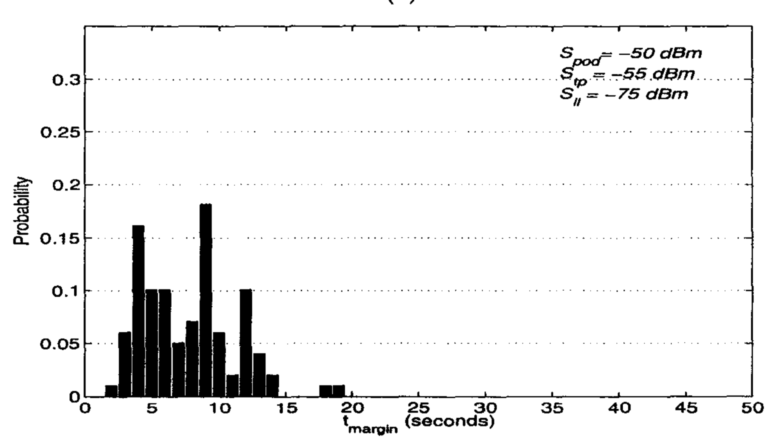
In order to gain more insight into the actual behavior of different RSSI paths the pdf of t_{margin} is considered for a small yet representative set of S_{pod}, S_{tp}, and S_{ll} values. Figure 3.7 shows the t_{margin} distribution for the three pairs ($S_{tp} = -55$ dBm, $\mathrm{S_{ll}} = -65$ dBm), ($S_{tp} = -65$ dBm, $\mathrm{S_{ll}} = -75$ dBm) and ($S_{tp} = -55$ dBm, $\mathrm{S_{ll}} = -75$ dBm) when $S_{pod} = -50$ dBm which may correspond to a dense AP deployment or the AP being located close to the PoD. Figure 3.7(a) where $S_{diff} = 10$ dB and S_{tp} is set 5 dBs below S_{pod}, suggests that there is a large number of paths that experience very quick signal drop in this region. More specifically, 55% of the paths drop from -55 dBm to -65 dBm in less than three seconds, leading to unacceptable $P_{success}$ even though the average t_{vho} interval is typically much longer than three seconds. Furthermore, by comparing Figure 3.7(b) to Figure 3.7(a), it is observed that setting S_{tp} 15 dBs lower than S_{pod} decreases the chances of having very abrupt signal degradation significantly. For instance, merely 9% of the paths drop from -65 dBm to -75 dBm in less than three seconds comparing to the 55% in the previous case. Overall, the t_{margin}



(a)



(b)



(c)

Figure 3.7: *t_{margin} Probability Distribution Function for S_{pod}=-50. Number of observations is 102 paths*

distribution shifts to the right as the 10 dB S_{diff} window is moved farther away from the S_{pod}. This behavior of the pdf can be explained by noting that as the distance to the transmitter increases, the rate at which the signal degrades decreases. Thus, experiencing a constant loss takes a longer period of time at points much further away from the transmitter than points which are close to the transmitter. These longer intervals corresponding to S_{tp} values which are set much lower than S_{pod} yield longer t_{margin} values explaining the shift to the right of the distribution. Also, from the same two figures we see that the variance increases when the 10 dB S_{diff} window is moved farther away from the S_{pod} because the WLAN signal experiences more shadowing over the longer distance it has to traverse.

The behavior of increased mean and variance of t_{margin} continues in Figure 3.7(c) where we increase S_{diff} from 10 dBs to 20 dBs, resulting in a wider t_{margin} pdf with a greater mean and consequently a higher $P_{success}$. For example, only 2% of the curves experience a drop from -55 dBm to -75 dBm in less than three seconds. Nevertheless, even for this window size, (i.e., $S_{diff} = 20$ dB), many paths can experience an unsuccessful handover since a significant portion of the distribution continues to span low values of t_{margin}. Therefore, performing a successful handover may require allowing the station to hold the value of S_{ll} very close to the noise floor for a short period of time (i.e., until the completion of VHO).

Figure 3.8 shows the t_{margin} distribution for different positions and values of S_{diff}

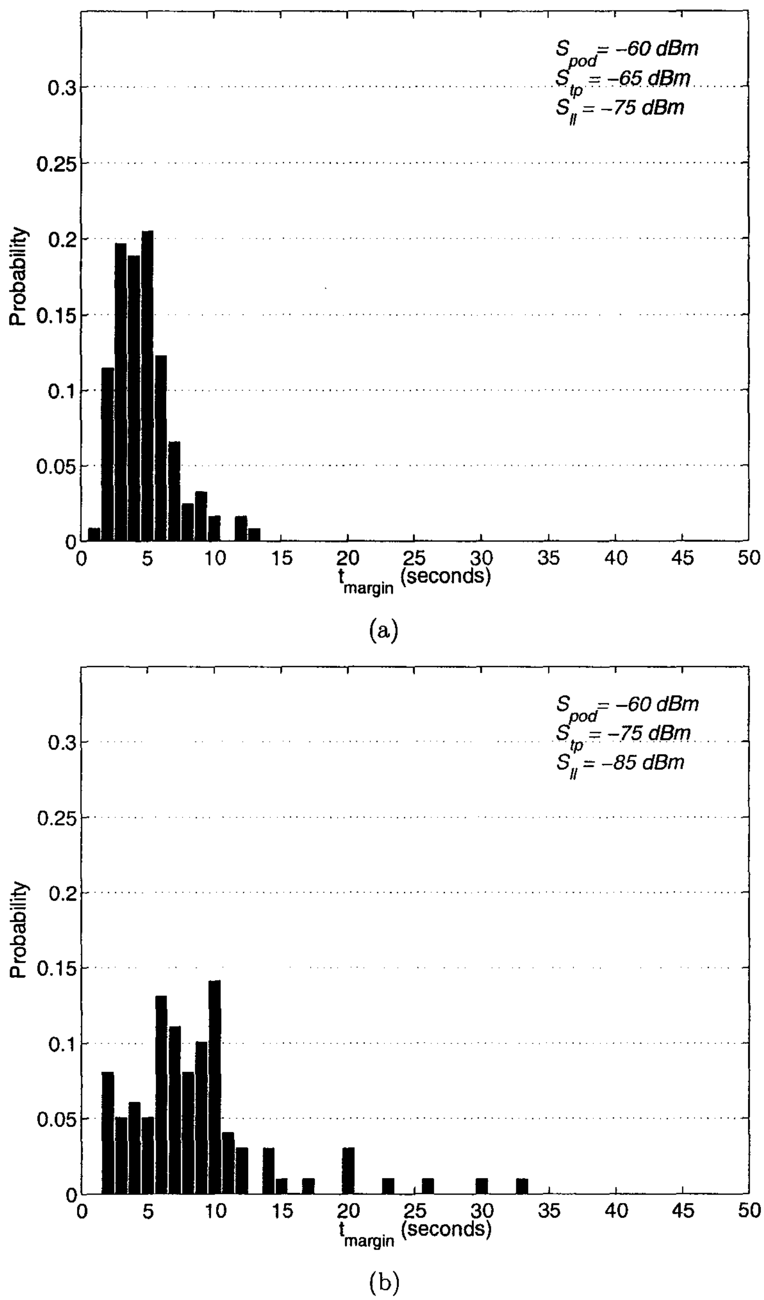


Figure 3.8: t_{margin} *Probability Distribution Function for* $S_{pod} = -60$ *dBm. Number of observations is 121 paths*

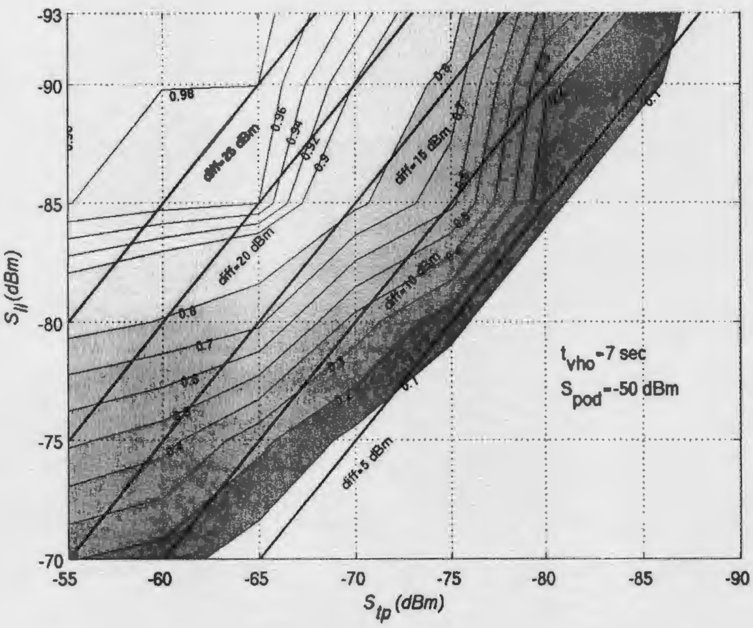
when $S_{pod} = -60$ dBm which corresponds to a sparse deployment with emphasis on range. If we compare Figure 3.7(a) to Figure 3.8(a) we observe that having a smaller S_{pod} value while maintaining the same S_{diff} window size and relative distance between S_{pod} and S_{tp} results in a decreasing signal degradation rate (i.e., larger t_{margin} mean) and widens the pdf of t_{margin} (i.e., higher variance). This is explainable by noting that a smaller S_{pod} value is a result of having the transmitter being far away from the PoD and consequently distances traveled after crossing the PoD to experience a fixed S_{diff} dBs of loss will be longer for smaller S_{pod} values than for larger ones. Also, if we compare Figure 3.7(b) where $S_{pod} = -50$ dBm to Figure 3.8(a) where $S_{pod} = -60$ dBm, for the same trigger and link loss points, namely ($S_{tp} = -65$ dBm, $\mathrm{S_{ll}} = -75$ dBm), we notice that t_{margin} corresponding to $S_{pod} = -50$ dBm has a pdf with larger mean and variance than $S_{pod} = -60$ dBm. This result may seem counterintuitive, however, setting S_{tp} 15 dBs below $S_{pod} = -50$ dBm results in the WLAN signal traveling a longer distance than the one traveled when $S_{pod} = -60$ dBm with S_{tp} being set 5 dBs below the S_{pod}. This longer distance is a result of the outdoor loss exponent being smaller than the indoor one. Therefore, we may conclude that even at high S_{pod} values, the designer may choose to trigger further away from the PoD to avoid areas with higher degradation rate when restricted to a fixed S_{diff} window size.

As expected, when Figure 3.8(b) is compared to Figure 3.8(a) we see that by moving the beginning of the fixed $S_{diff} = 10$ dB window farther away from S_{pod}, abrupt signal degradation decreases significantly. Interestingly, setting the S_{diff} window 15 dBm below the S_{pod} results in larger increase in the mean and variance of t_{margin} for $S_{pod} = -60$ dBm than for $S_{pod} = -50$ dBm. Again, this is attributed to the distance between the points $S_{tp} = -75$ dBm and $S_{ll} = -85$ dBm for $S_{pod} = -60$ dBm being larger than that between the points $S_{tp} = -65$ dBm and $S_{ll} = -75$ dBm for $S_{pod} = -50$ dBm due to the $S_{pod} = -60$ dBm being further away from the transmitter than $S_{pod} = -50$ dBm.

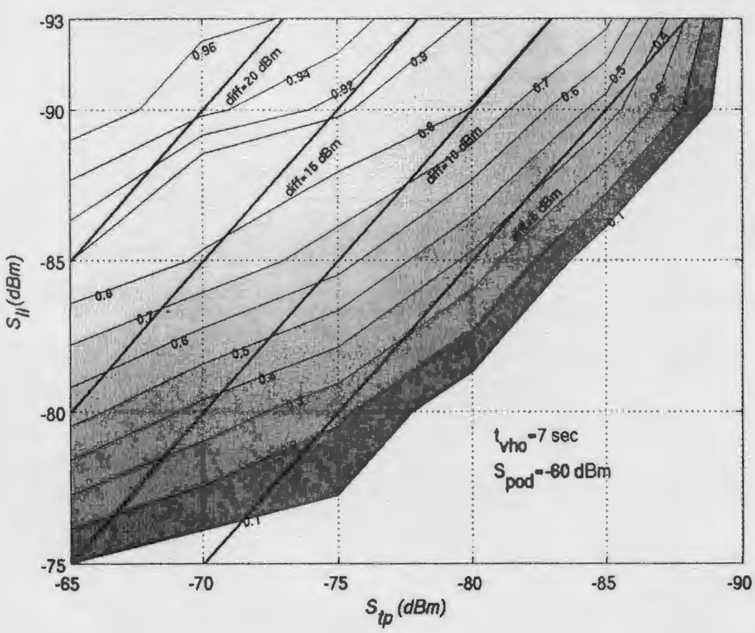
3.4.3 Impact of Trigger and Link Loss Points on VHO Success

The performance of a vertical handover algorithm is measured by its ability to achieve acceptable $P_{success}$. Figure 3.9 shows the $P_{success}$ for different (S_{tp},S_{ll}) combinations. The vertical handover latency is seven seconds.

In a handover algorithm, S_{ll} could be a fixed value that is selected based on the type of WLAN deployment. For example, if $P_{success} = 98\%$ is desired for ($t_{vho} = 7$ seconds, $\mathrm{S_{pod}} = -50$ dBm), then from Figure 3.9(a) it can be observed that ($S_{tp} = -55$ dBm,$S_{ll} = -85$ dBm) and ($S_{tp} = -65$ dBm,$S_{ll} = -90$ dBm) both provide the desired $P_{success}$. If having $S_{pod} = -50$ dBm is the result of a dense AP deployment where the indoor RSSI is greater than S_{pod}, then setting



(a)



(b)

Figure 3.9: *Different* (S_{tp}, S_{ll}) *Pairs Success Probability for* $t_{vho} = 7$ *seconds*

(S_{tp} = −55 dBm,S_{ll} = −85 dBm) is the optimal choice. This (S_{tp},S_{ll}) combination guarantees minimal false triggering and does not impact the WLAN capacity. On the other hand, S_{pod} = −50 dBm could result from an AP being located close to the PoD in a sparse deployment with indoor RSSI being lower than S_{pod}. In this case, setting (S_{tp} = −65 dBm,S_{ll} = −90 dBm) severely impacts the WLAN capacity in lieu of better false triggering probability. Fortunately, WLAN capacity is not emphasized in sparse deployments where range extension is emphasized. Another design approach is to fix the S_{diff} window size and find the highest achieveable $P_{success}$ for a given (S_{pod},t_{vho}) pair. For example, from Figure 3.9(b) it can be seen that for ($S_{pod} = -60$ dBm, $\mathrm{t_{vho}} = 7$ seconds, $\mathrm{S_{diff}} = 15$ dB) the highest achievable $P_{success} = 92\%$ and requires ($S_{tp} = -78$ dBm, $\mathrm{S_{ll}} = -93$ dBm).

Decreasing the value of S_{tp} implies moving toward the upper right corner along the lines marking fixed S_{diff} window sizes on the subfigures of Figure 3.9. Increasing the difference between S_{pod} and S_{tp} results in longer times to experience S_{diff} dBs of signal loss and consequently higher $P_{success}$. However, if capacity is of concern, then there will be a tendency to move in the opposite direction, namely the lower left corner along any of these lines to increase S_{ll} and consequently preserve the WLAN capacity. These conflicting movements along fixed S_{diff} lines imply that achieving high VHO success probability requires allowing the mobile user to maintain the WLAN link at very low RSSI values, a requirement that contradicts WLAN capacity optimization.

This tradeoff between WLAN capacity and vertical handover success probability is studied in greater detail in Chapter 5.

Another important property of success probability can be observed by carefully considering the distance of adjacent contour level lines (i.e., where $P_{success}$ is constant). The level lines are generally much denser in the vertical direction than in the horizontal direction, suggesting that success probability is more sensitive to changes in S_{ll} than in S_{tp}. This sensitivity is due to the physical location of the S_{ll} point being further away from the transmitter than S_{tp}. For example, if δ is chosen to be an arbitrarily small positive value, then the point at which the signal strength is $S_{ll} + \delta$ dBm will be further away from S_{ll} than the point $S_{tp} + \delta$ dBm from S_{tp}. This sensitivity to S_{ll} can prove useful in some instances. For example, in a sparsely covered WLAN deployment with a strong signal at the PoD (e.g., $S_{pod} = -50$ dBm), the designer is forced to set S_{tp} to be much lower than S_{pod} to avoid a high false triggering probability, thus lowering $P_{success}$. Due to the high $P_{success}$ sensitivity to S_{ll}, one can compensate for the decrease in success probability by slightly decreasing S_{ll} without sacrificing WLAN capacity. In other words, a small decrease in S_{ll} may not necessarily cause the AP to rate-adjust, thus increasing $P_{success}$ without impacting the WLAN capacity. An example of $P_{success}$ sensitivity to S_{ll} can be seen by looking at the level line corresponding to $P_{success} = 96\%$ in Figure 3.9(a) and observing that ($S_{tp} = -55$ dBm, $\mathrm{S_{ll}} = -84$ dBm) and ($S_{tp} = -65$ dBm, $\mathrm{S_{ll}} = -85$ dBm) both are located at the same level line. Therefore, a one dB decrease in S_{ll} has allowed decreasing S_{tp} by 10 dBs while maintaining the same success probability.

The level lines in Figure 3.9 are more concentrated along the horizontal direction for low values of S_{tp} (i.e., less than -85 dBm for $S_{pod} = -50$ dBm and less than -90 dBm for the other two cases $S_{pod}=-60$ dBm and -70 dBm). In other words, increasing S_{tp} improves $P_{success}$ more than decreasing S_{ll}. By carefully considering the data corresponding to this region, it is found that a significant number of paths experience lost links due to consecutive missed beacons. This condition occurs at distances far from the exit point where significant shadowing causes these beacons to be missed at the receiver, even if the average signal strength is still at relatively acceptable values (-85 dBm to -90 dBm). When missed beacons occur, holding the WLAN link to low RSSI values does not improve $P_{success}$. However, triggering earlier increases $P_{success}$ since it allows the mobile user more time to complete the VHO before starting to experience missed beacons.

The study of individual level lines representing a fixed $P_{success}$ provides insight on changing S_{tp} and S_{ll} to maintain a desired success probability. These lines show that almost everywhere $P_{success}$ is more sensitive to S_{ll} than to S_{tp}. This is the same result we arrived at before by looking at the placement of level lines with respect to each other and by looking at the pdf of t_{margin}.

From Figure 3.9, it can be observed that the regions corresponding to high $P_{success}$

(e.g., 98% and 96%) are small leaving the handover designer with limited choices for S_{tp} and S_{ll} especially if lower values of S_{ll} are not acceptable due to capacity constraints. All cases which result in high success probability require the mobile user to hold the WLAN link even if the RSSI values become very close to the noise floor (i.e., -95 dBm). Conditions become even worse if the designer has to pick S_{tp} much smaller than S_{pod} due to a sparse deployment which would prevent achieving a high success probability. This conclusion highlights the fundamental difference between VHO and handover in homogeneous cellular networks. Achieving WLAN-to-WWAN success probabilities similar to those found in homogeneous WWAN handovers is very difficult and may require the use of non-conventional handover techniques (i.e., not threshold-based or tightly-coupled) [KTZ04, LSB+04, BSL05].

3.5 Trigger Node

Vertical handover success probability is a strong function of S_{pod}. One solution to improve the success probability is to set S_{pod} as high as possible. In this section we control the value of S_{pod} by positioning APs at known distances to the PoD. A new set of measurements is obtained using the procedure described in Section 3.3. Our objective is to study the effect of placing an AP close to the PoD specifically for the purpose of aiding VHO.

In Figure 3.10 , the pdf of the S_{pod} is shown for APs which are placed in close proximity to the PoD. There is a wide range of RSSI values in the presented pdf with a minimum observed value that is close to -70 dBm and a maximum value that is close to -35 dBm. Due to this wide range we consider the t_{margin} pdf with respect to specific APs which are positioned at different distances from the PoD.

In Figure 3.11, the t_{margin} pdf of APs with -60 dBm $\leq S_{pod} < -50$ dBm is presented. The value of $S_{tp} = -70$ dBm is chosen to satisfy the false triggering constraint for a deployment with $S_{min} = -67$ dBm, which is the recommended value for VoIP-capable deployments [Sys05]. In Figure 3.11(a), the value of $S_{ll} = -92$ dBm is chosen to satisfy a minimum advertised data rate of 6 Mbps. In this case, 3.5% of the VHO events will fail if $t_{vho} \geq 7$ seconds. This failure probability is also similar to the one obtained from Figure 3.11(b) where the rate is allowed to drop to 1 Mbps. This is explainable by noting that many links are terminated early due to lost beacons which occur at a higher rate as the RSSI approaches the noise floor (i.e., $S_{ll} = -95$ dBm).

In Figure 3.12, the t_{margin} pdf of APs with -50 dBm$\leq S_{pod} < -40$ dBm is presented. Figure 3.12(a) shows 2% of VHO events fail when $S_{tp} = -70$ dBm and $S_{ll} = -92$ dBm when $t_{vho} = 7$ seconds. A similar percentage of VHO events fail when $S_{ll} = -95$ dBm as shown in Figure 3.12(b) . Again, early-termination of the links at RSSI values close to the noise floor due to missed beacons explains the similarity between the success probability when $S_{ll} = -92$ dBm and $S_{ll} = -95$ dBm. Having the S_{pod} change from −60 dBm to −50 dBm results in a 1.5% improvement in failure probability. This increase in S_{pod} value is a direct result of positioning APs closer to the PoD.

Only 7% of the observations obtained (i.e. 24 paths) have their $S_{pod} \geq -40$ dBm. In all of these cases, 100% VHO success probability is achievable for $t_{vho} \leq 13$ seconds even when the link is terminated at $S_{ll} = -92$ dBm. These values are the highest cellular call establishment times observed in the industry. Access points with $S_{pod} \geq -50$ dBm are those placed either right at the PoD or in its immediate vicinity. Having an AP at the PoD provides a line-of-sight link with the MS and thus a longer period of time before the link is lost. In VoIP-capable deployments, placing an additional AP at the PoD can be considered as a viable solution to satisfy the false triggering requirement while achieving acceptable success probabilities. This AP is referred to as a Trigger Node (TN).

3.6 Conclusions

In this chapter we have considered the impact of WLAN propagation on the vertical handover in the loosely coupled WLAN/cellular integration model. Vertical handover was characterized using a measurement-based study of WLAN signal behavior in building exit transition regions. The presented results were obtained from detailed IEEE 802.11 measurements that were made on the McMaster University campus.

This data was collected by traversing many indoor-to-outdoor paths for a large number of campus buildings and exits while measuring multi-AP WiFi coverage. The performance of threshold-based vertical handover was examined and it was found that it is virtually impossible to successfully complete this type of handover unless the WLAN deployment has been very carefully engineered. Our results show for example, that even when minimum indoor WLAN coverage exceeds -67 dBm, vertical handover success probabilities above 90% are impossible to achieve in loosely-coupled systems while using threshold-based triggering. In a practical system this situation is compounded by the fact that many deployments now place a lower limit on the WLAN data rate that can be used. This is done to ensure that high data rate links are used whenever possible. A trigger node was proposed for the purpose of aiding vertical handovers when crossing a WLAN boundary. The trigger node was found to be a viable option to satisfy the false triggering requirement while achieving acceptable success probability.

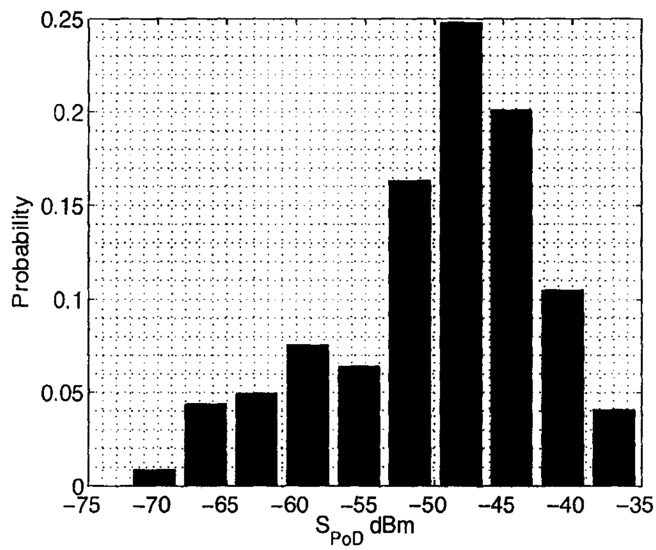


Figure 3.10: *Pdf of RSSI values at the PoD*

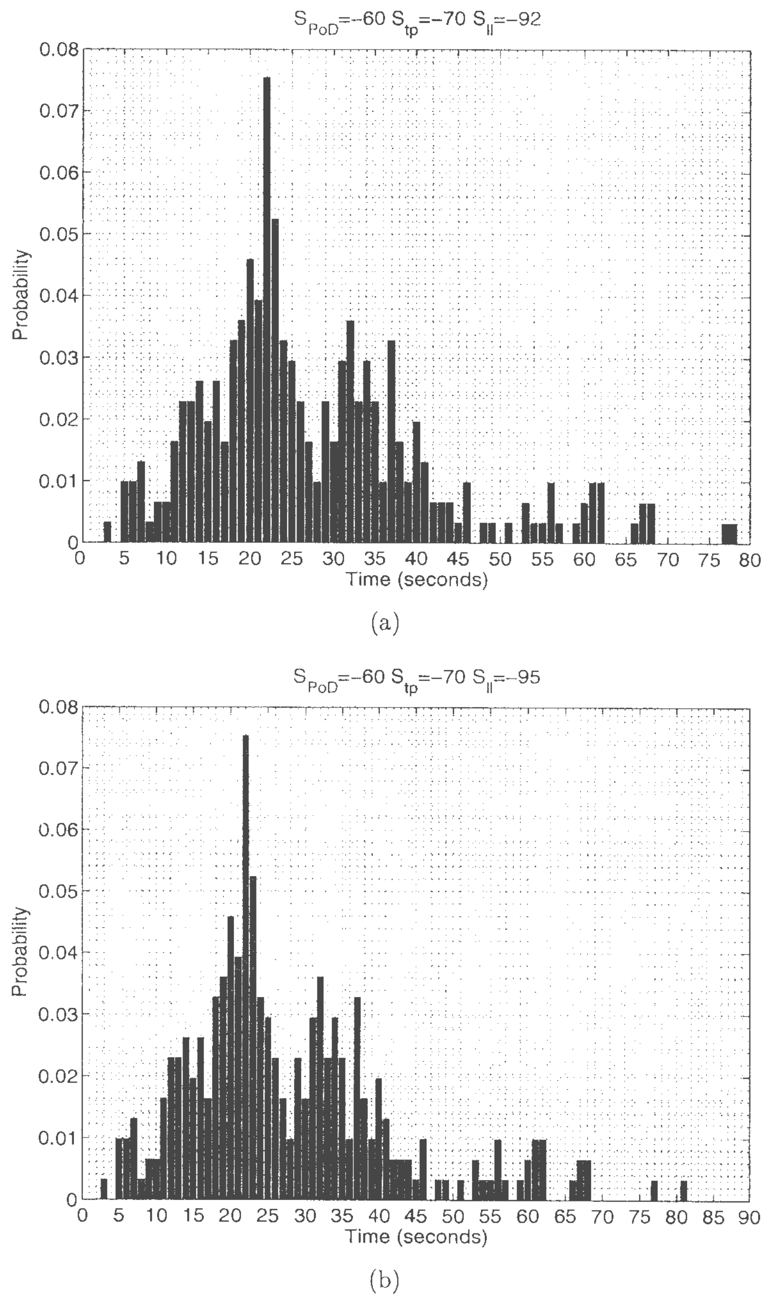
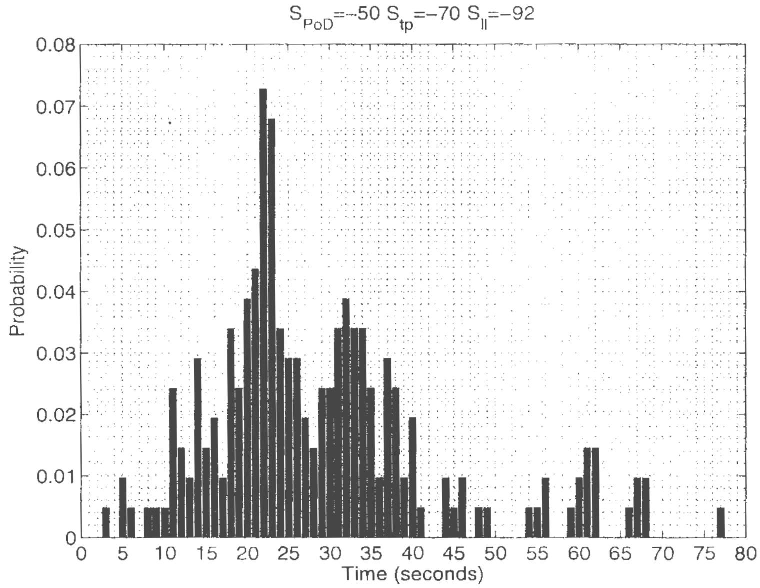
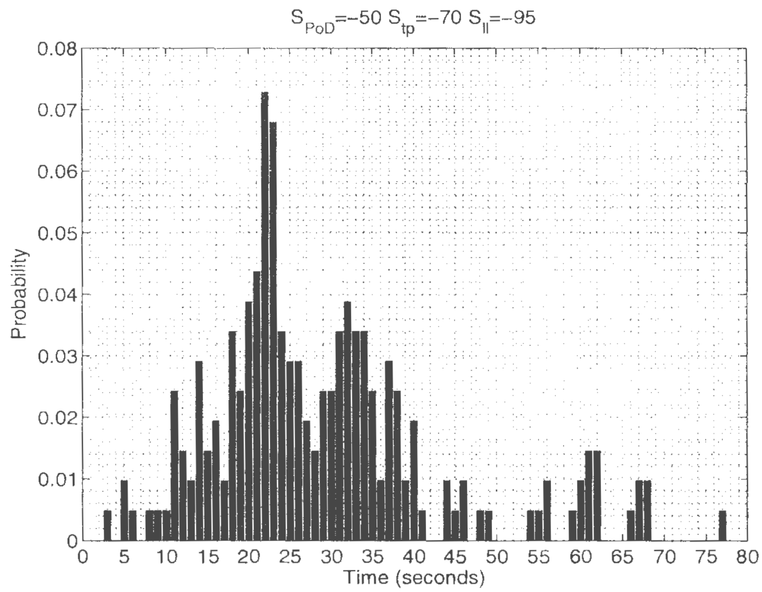


Figure 3.11: t_{margin} *pdf for APs with* $S_{pod} = -60$ *dBm. Number of observations is 99 paths*



(a)



(b)

Figure 3.12: *t_{margin} pdf for APs with $S_{pod} = -50$ dBm. Number of observations is 182 paths*

Chapter 4

Anchored Conferencing Handover

4.1 Overview

In today's world, most enterprises operate their own PSTN gateways which support conference bridging. A loosely-coupled system can benefit from the gateway's conferencing ability to act as a media mixer between the interfaces of a dual-mode handset during a handover. In Figure 4.1 (reproduced here from Chapter 3 for reader's convenience), the initial two legs of the call are placed through a conference room at the GW. When VHO is imminent, the third leg of the call is established between the cellular interface and the GW conference room. At this point all three legs of the call are active, and leg 1 can then be dropped. This mechanism provides a soft

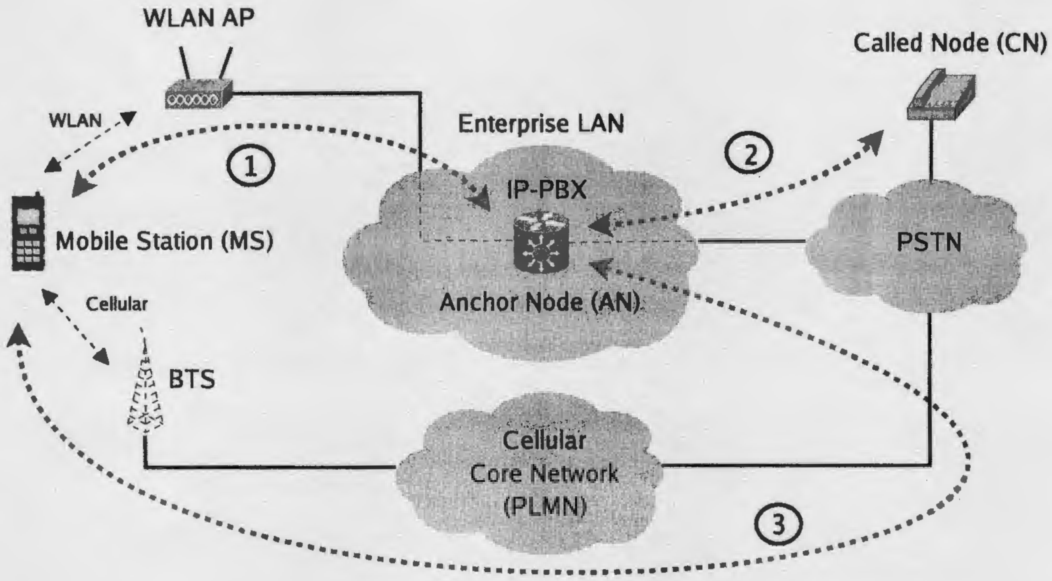


Figure 4.1: *Enterprise PSTN GW Anchored Vertical Handover*

make-before-break handover using a PSTN GW that already exists within the enterprise. The only added components involve the signaling required to accomplish the soft handover.

The conferencing approach to soft handover requires that all calls be routed through the gateway when the call is being established, even if the origin and destination of the call are residing within the enterprise for the duration of the call. The conference room involvement is needed in the event that at some later time the dual-mode handset transitions out of WLAN coverage. In this case, resource availability may be an issue if the enterprise has a large mobile population. In Section 4.3 a method - referred to as DACH - is proposed that leverages short latencies observed in enterprise networks to dynamically anchor calls at the gateway only when a handover is imminent. DACH solves scalability problems associated with the previously proposed schemes. In Section 4.4 we describe a testbed that was constructed to evaluate the proposed mechanisms. In Section 4.5 analytic models are proposed for calculating the new call blocking and handover dropping probabilities. A simulation environment is described which is used to present various performance results.

4.2 Soft Handover Using a Conferencing Gateway

In the PSTN GW soft handover approach, the GW is used as an enterprise anchor, i.e., all calls are routed through the GW. A call from the WiFi interface to the PSTN destination consists of two legs, the first is from the MS's WiFi interface to the GW and the second leg from the GW to the PSTN destination as shown in Figure 4.1. When performing a handover, the first call leg is replaced by a newly established leg from the MS's GSM interface to the PSTN interface of the GW (i.e., leg 3), rather than directly to the PSTN destination. A third party call controller (3PCC) [RPSC04] attached to the enterprise LAN is used to coordinate the three legs of the call. Alternatively a hairpin device can be used to automate the functionality of the third party controller if this is supported by the PSTN GW.

Figure 4.2 shows the signaling carried out by the 3PCC/HP device to perform a WiFi-to-GSM handover. Initially the WiFi interface dials a PSTN destination (F1)

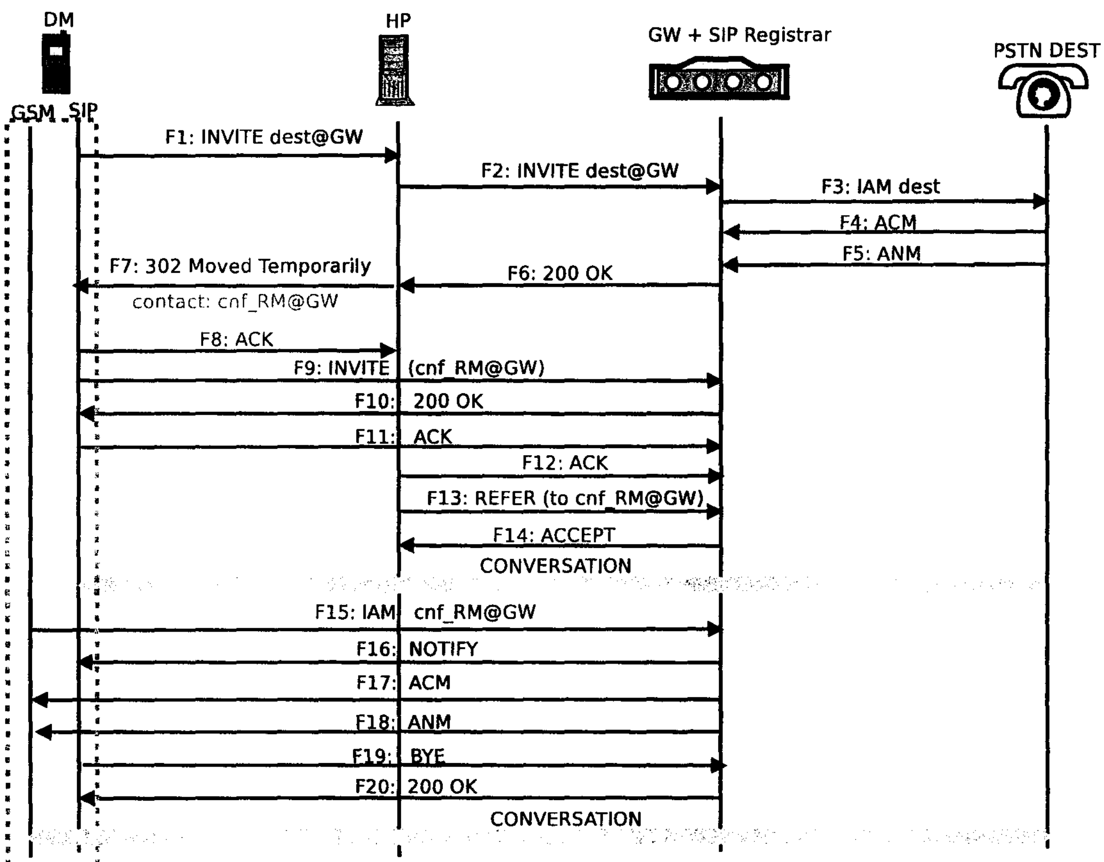


Figure 4.2: 3PCC/Hairpin Signaling for WiFi-to-GSM Handoff

and based on the MS's dial plan the invitation is forwarded to the HP, which in turn dials the PSTN destination through the GW (F2). Once the PSTN signaling in response to F2 arrives to the GW, that is F3: Initial Address Message (IAM), F4: Address Complete Message (ACM) , and F5: ANswer Message (ANM) is relayed back to the HP (F6), the HP delays sending the `ACK` (i.e., F12) until it redirects the WiFi interface to the conference room on the gateway (F7-F11). When the WiFi leg has been established to the conference room, the leg from the HP to the PSTN destination is acknowledged (F12) and immediately referred into the conference room (F13-F14). This referral is transparent to the PSTN destination since the conference room resides on the GW itself; thus the signaling is shielded from the PSTN destination. At this stage the media stream is flowing through the conference room and the two legs to the conference room have been successfully established. Once a handover trigger occurs, the MS's GSM interface establishes a connection with the conference room at the GW (F15, F17, F18) and the GW notifies the WiFi interface of the GSM's successful leg establishment (F16), at which time the old leg from the WiFi interface to the conference room is terminated (F19, F20). The amount of time required to perform a WiFi-to-GSM handover, T_{WiFi_GSM} is given by

$$T_{WiFi\_GSM} = t_{GSM} + t_{BYE}, \tag{4.1}$$

where t_{GSM} is the GSM call establishment time and t_{BYE} is the time required to terminate the SIP leg from the MS to the gateway (i.e., F19 and F20).

Now consider the case where the MS is transitioning from cellular to WLAN, shown in Figure 4.3. Initially the MS's GSM interface dials the PSTN destination through the GW rather than directly to the PSTN destination (F1-F3). Based on the MS's dial plan, the HP dials the PSTN destination through the gateway (F4-F7) and once the leg is established from the HP to the PSTN destination (F8, F9), it is referred into the conference room (F10, F11) transparently to the PSTN destination since the signaling is shielded by the GW. If a handover trigger is detected, the MS's WiFi interface establishes a leg to the conference room (F12-F14) and then terminates the GSM leg to the conference room (F15, F16). The amount of time required for a GSM-to-WiFi handover, T_{GSM_WiFi}, is given by

$$T_{GSM\_WiFi} = t_{INVITE} + t_{END}, \tag{4.2}$$

where t_{INVITE} is the amount of time required to create the SIP leg from the MS to the conference room and t_{END} is the amount of time required to terminate the MS's GSM leg to the conference room.

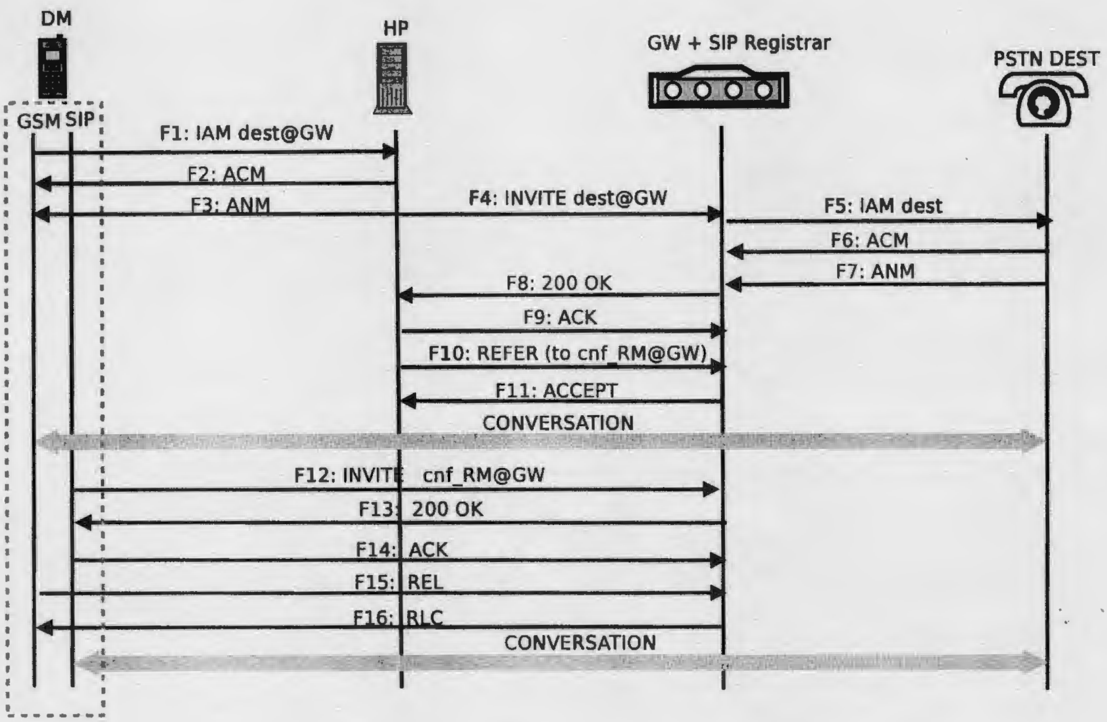


Figure 4.3: 3PCC/Hairpin Signaling for GSM-to-WiFi Handover

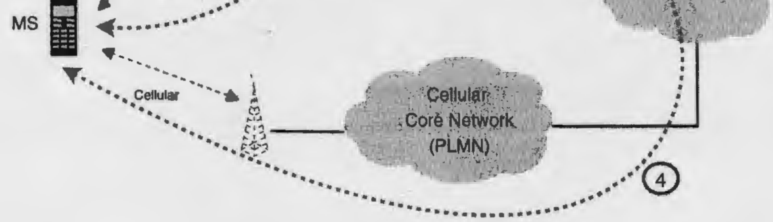


Figure 4.4: *DACH Handover*

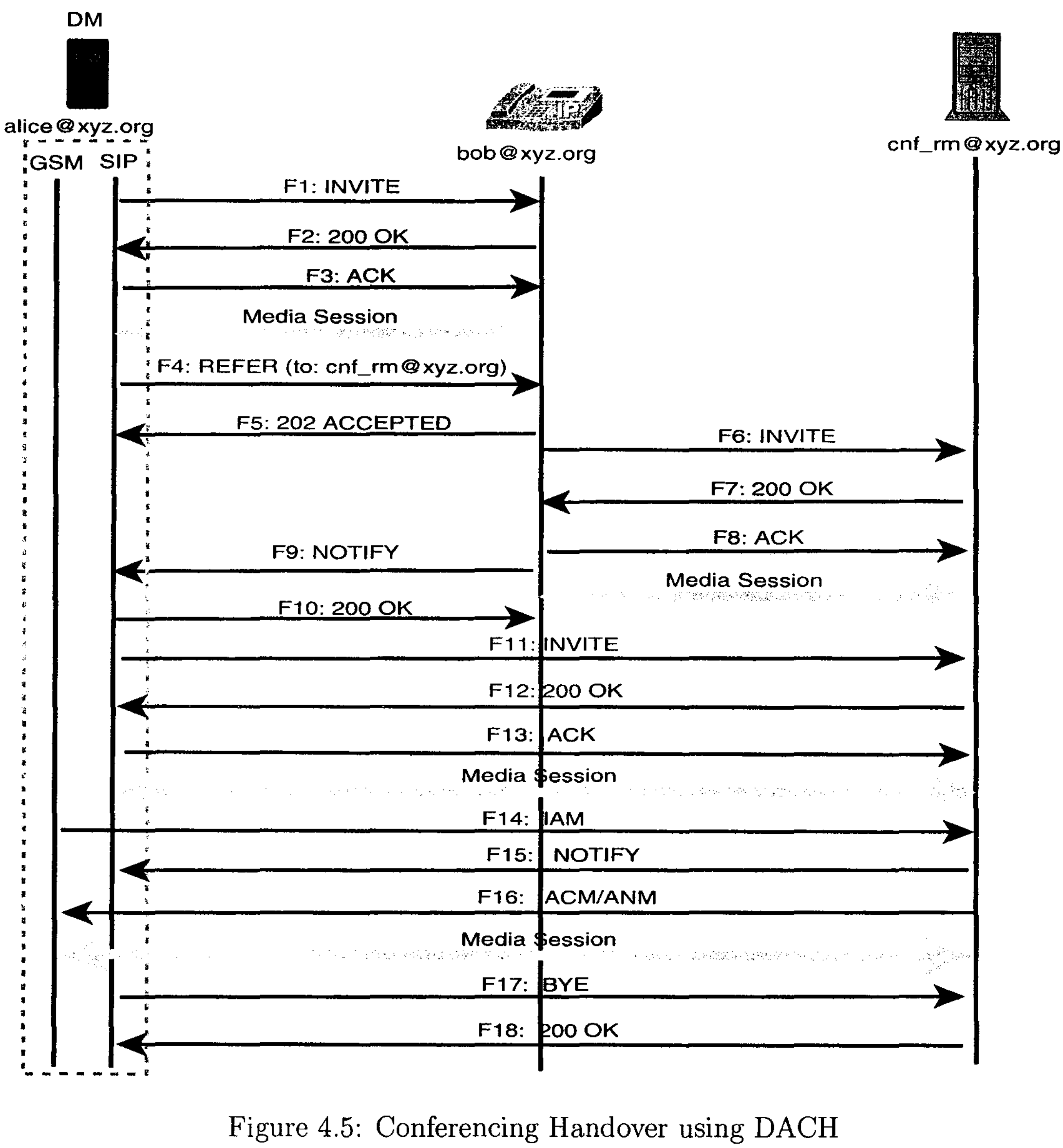


Figure 4.5: Conferencing Handover using DACH

4.3 Dynamically Anchored Conferencing Handover

A major disadvantage of the soft conferencing handover described previously is that all intra-enterprise calls must be established through a conference room. In an enterprise setting this may very easily require significant GW resources compared with that needed to support normal conferencing activities. However, intra-enterprise UA-UA SIP calls are contained solely within the enterprise and normally incur very short end-to-end latencies. The scheme proposed in this section exploits these short network delays by establishing the call directly (i.e., without involvement from the PSTN GW) and then dynamically anchoring the call to a conference room only when VHO is imminent. This mechanism is referred to as Dynamically Anchored Conferencing Handover (DACH).

In Figure 4.4 an example is shown of a MS that is using DACH to perform VHO. In the example shown the MS is initially residing within the enterprise LAN and establishes a call to a stationary IP phone that is also residing within the enterprise LAN. This call is established as a UA-to-UA call without the media stream being routed through the IP PBX. The media path is shown as an end-to-end leg 1. Some time later the MS decides to perform a VHO. At this time a leg 1 is broken into two legs, numbered 2 and 3. These two new legs are established by referring the media streams from the MS and the stationary IP phone into the conference room at the IP PBX. At this time a new call leg 4 is created between the cellular PLMN/PSTN and the GW, which replaces the WLAN call leg 2. After the handover the new media path consists of call legs 3 and 4 as shown in Figure 4.4. The splitting of the media path in leg 1 into legs 2 and 3 is performed using a hard handover. However, the short latency of the enterprise LAN makes this hard handover unnoticeable. The handover from call leg 2 to call leg 4 in Figure 4.4 still has to be completed before the WLAN link is lost.

The signaling required for this is shown in Figure 4.5. The WiFi interface of the MS (i.e., alice@xyz.org in this case) sends an `INVITE` to the SIP destination residing on the same LAN (i.e., bob@xyz.org). In response to the `INVITE`, an `OK/ACK` sequence is exchanged between Alice and Bob and the RTP media stream flows end to end without going through the conferencing gateway, reducing the load on the GW (F1-F3). Upon detecting the handover trigger, Alice refers Bob to the conference room on the GW using a `REFER` request (F4-F8) and simultaneously establishes a leg from herself to the conference room (F11-F13). This step is possible because of the short latencies incurred during a direct UA-to-UA call over an enterprise LAN, and is not noticeable to the end user. The GSM interface is then instructed by the MS to establish a leg to the conference room and the WiFi-to-GSM handover occurs in the same way as previously described in Section 4.2. However, to maintain an acceptable level of voice quality during the referral procedure of the original SIP-to-SIP call into the conference room, the establishment of the media streams from both SIP UAs to the conference room should not experience a total delay longer than D_{MAX} which is defined as 150ms [Uni]. This requirement imposes an upper limit on the allowable network delay. To calculate the upper limit on the allowable one-way network delay such that DACH can be used, we examine the flow of the media stream and note that it is put on hold once the `REFER` request has been issued, F4, as shown in Figure 4.5, and it only resumes when both `INVITE/OK/ACK` sequences (F6, F7, F8 and F11, F12, F13 respectively) have completed. However, the two `INVITE/OK/ACK` sequences do not necessarily have to be performed sequentially, and the same applies for establishing the GSM leg, as these activities can be performed in parallel. Therefore the network delay D_{LAN} must abide by

$$4 \times D_{LAN} \leq D_{MAX} - t_{proc}, \tag{4.3}$$

where t_{proc} is the time to process the SIP requests at higher layers.

4.4 Realtime Handover

A diagram of our testbed is shown in Figure 4.6. The MS handset was built using an IBM Thinkpad T41 laptop with a Linksys WPC11 802.11b WLAN card based on the PrismII chipset and an Option GlobeTrotter GPRS/GSM cellular card. The choice of WLAN card was necessary to allow the collection of sniffed traces using network analyzers such as Ethereal. We interfaced to the GPRS/GSM card via serial over PCMCIA interface. The WLAN card was connected to the network via a Cisco Aironet 1120 series AP in some instances and Linksys WRT54GS in some other instances. The cellular card accessed the Roger's PLMN network. We used two different gateways, namely a Cisco 1760V and an open source Asterisk running on a Linux PC, connected through their Foreign Exchange Office (FXO) interfaces to Bell Canada's PSTN network. We used different SIP destinations including soft and hard phones such as Kphone, Snom 200 and Snom 260 when collecting signaling and media traces. We also used different open source implementations of the SIP proxy including SIP Express Router and the Asterisk implementation. The third party controller was implemented at the Asterisk gateway using the Asterisk Gateway Interface programming language. Our choice was based purely on ease of implementation as it could have been implemented as an external device on the network.

Many handover experiments have been implemented using the testbed. Figure 4.7 shows a sample trace of the media stream while flowing through a conference room during a WiFi-to-GSM handover. In the figure, time is shown on the horizontal access and RTP packet sequence numbers on the vertical axis. The packets shown as x's are those arriving on the WLAN interface, and those shown as □'s are those arriving on the cellular interface.

The handover trigger event occurs at time 8.88 seconds. It takes the GSM interface

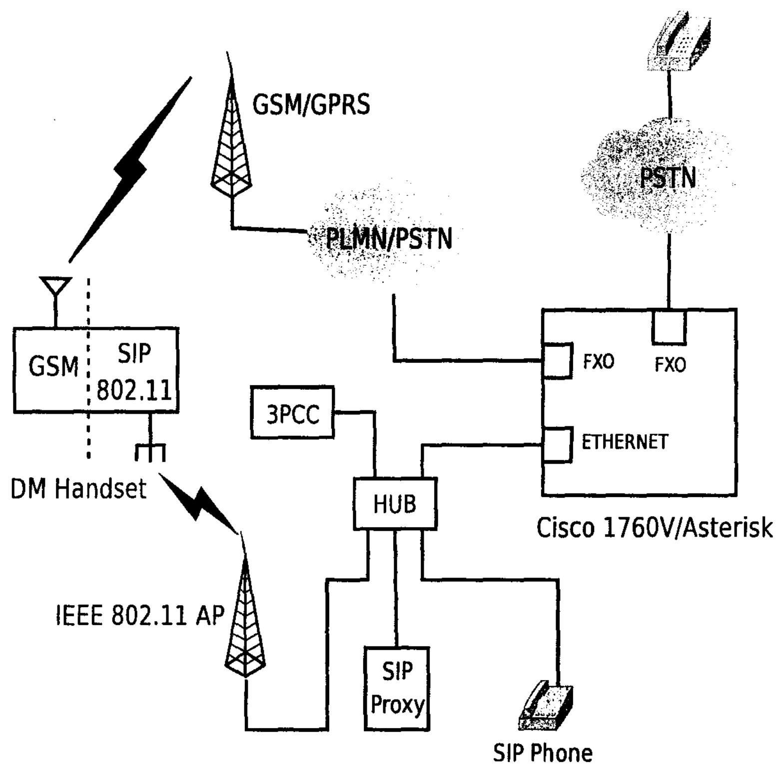


Figure 4.6: Dual-Mode Handset Testbed

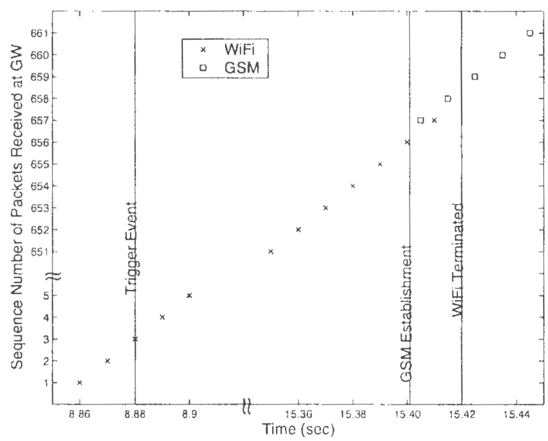


Figure 4.7: Media stream During WiFi-to-GSM Handover

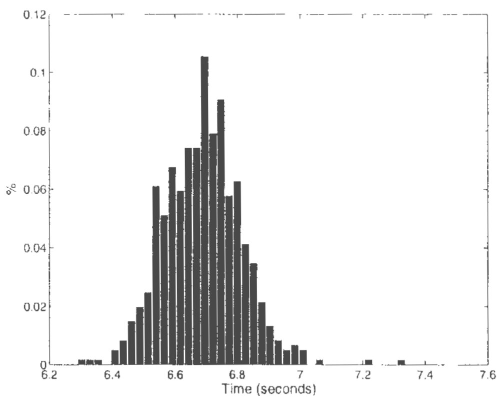


Figure 4.8: GSM Call Establishment Time

approximately 6.40 seconds to create the cellular leg to the conference room after the handover trigger has occurred. Once the GSM leg to the conference room is created, the MS terminates the WiFi leg. The handover procedure is completed at time 15.42 seconds, 20 msec after the cellular leg reaches the conference room. The handover is soft in that the two legs arriving from the MS to the conference room overlap for approximately 20ms. From Figure 4.7, we can see the importance of the time it takes to establish the GSM leg, t_{GSM}, since we have to trigger at least t_{GSM} seconds before the WLAN link becomes non-usable. The probability distribution function for t_{GSM} was estimated by establishing GSM calls every 15 minutes for two continuous weeks and capturing the time it took to establish the call leg. It was found that the average value of t_{GSM} was 6.68 seconds with a standard deviation of 0.123 seconds and is shown in Figure 4.8. The mean GSM call establishment time is quite long, which makes early handover triggering very important. Interestingly, the standard deviation of GSM call latency is quite low for this particular cellular operator. When DACH is used, we expect to see the same behavior for the media stream as shown in Figure 4.7 in addition to the initial signaling required to re-direct the end-to-end SIP call to the conference room. This re-direction step was also repeated many times using our gateway. The mean time to perform this function was found to be 56.04 msec with a standard deviation of 12.17ms. This time is sufficiently short that it is not noticeable during the call.

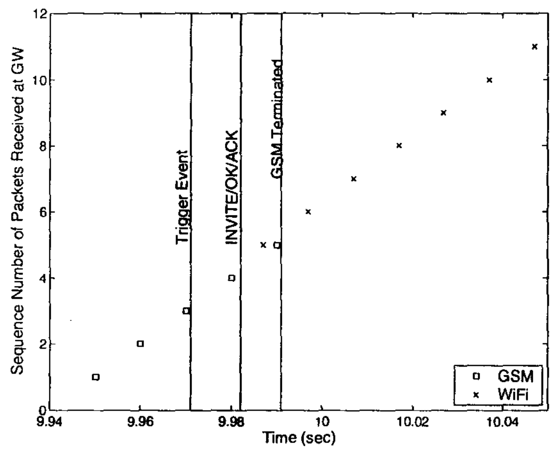


Figure 4.9: Media Stream During GSM-to-WiFi Handover

Figure 4.9 shows the opposite scenario where a GSM-to-WiFi handover takes place after a handover trigger occurs at time 9.97 seconds. The WiFi leg to the conference room is completed at time 9.98 seconds. The 10 msec required to establish the WiFi leg to the conference room is typical of enterprise LANs. At time 9.99 seconds the MS terminates the GSM leg to the conference room and completes the handover procedure. As mentioned earlier, this case poses a looser time constraint since the GSM enjoys more ubiquitous coverage and the call establishment over the WiFi interface takes, on average, a shorter period of time than the GSM call establishment.

4.5 Call Blocking and Handover Behavior

In this section, we assume an enterprise setting where some users are equipped with dual-mode handsets while others are equipped with stationary SIP phones. We allow for two different types of call arrivals, namely, intra-enterprise and inter-enterprise calls. Intra-enterprise calls are established between two users who reside on the enterprise LAN and are referred to as `IN_IN` calls . An `IN_IN` call requires one conference room in the conventional soft handover approach and does not require any conference rooms when using DACH. Inter-enterprise calls always require a conference room and a PSTN channel and are referred to as `IN_OUT` . We assume that the conferencing gateway has M conference rooms and N PSTN channels and that DACH is capable of supporting a maximum of Z concurrent `IN_IN` calls where typically $Z \gg M$.

We first develop analytical models for new call blocking and handover dropping probabilities, assuming the case where all new calls are intra-enterprise. All new calls involve one dual-mode handset user and one stationary user, however, both parties are assumed to be initially within the enterprise LAN. If the dual-mode handset departs the enterprise WLAN, it will then require an additional PSTN channel. We assume that if the call hands over to cellular, it is completed using the cellular network. We do not consider the case of cellular-to-WLAN handover, since the ubiquitous cellular coverage eliminates time constraints on performing the cellular-to-WLAN handover. For example, if on a given attempt a cellular-to-WLAN handover fails due to any

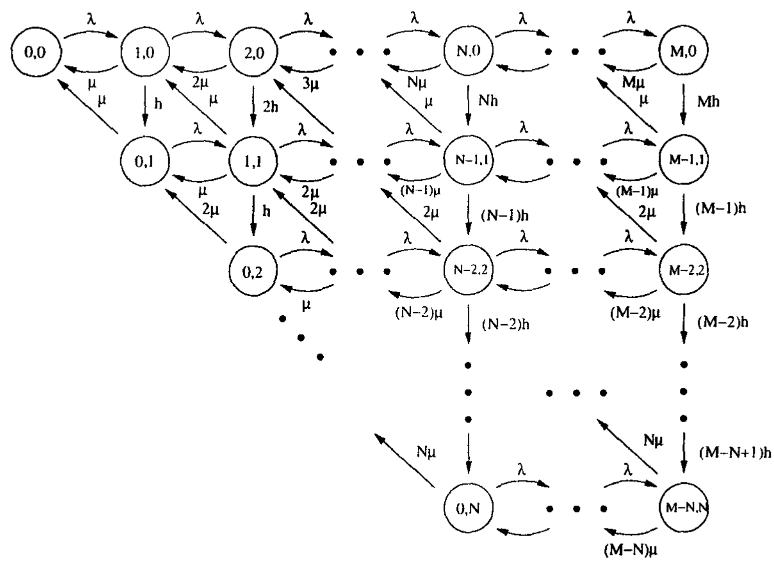


Figure 4.10: Markov Process for Conferencing GW, $M > N$

reason, then another attempt can be repeated without losing the original call since it will continue to be carried over the GSM leg.

The above case can be modeled as a Markov process. We assume that new call arrivals are Poisson distributed with a rate equal to λ, call durations are exponentially distributed with mean, $1/\mu$, and we also assume that an exponentially distributed time elapses before a handover, with rate h [FCL98, Eve94, YY95, Gu87, HR86]. Since our triggering scheme deals with a false handover in the same way that true handover, h indiscriminately includes false handover attempts. The new call blocking and handover dropping probabilities can be regarded as upper bounds since these probabilities will only improve as the triggering scheme becomes better at distinguishing between false and true handovers. Figure 4.10 shows the Markov process describing the system when $M > N$ where the state $s(i,j)$ indicates i direct intra-enterprise calls, and j is the number of calls that have handed off and are using a cellular leg, such that $i = 0, \dots M, j = 0, \dots N$. During state $s(i,j)$, any of the i intra-enterprise calls or j already handed off calls could terminate with rate μ. Moreover, any of the i intra-enterprise calls could handover and change the state of the system to $s(i-1, j+1)$. However, the call arrival rate λ is constant and independent of the current system state. When the system reaches $s(M-N, N)$, it becomes incapable of admitting any new calls or allowing handovers. Thus, the new call blocking probability for the system when $M > N$ is equivalent to the probability of the system being in any of the boundary states on the right edge of the Markov process because at that boundary no new calls are admitted. The new call blocking probability is given by

$$B_{new} = \sum_{i=0}^{N} Pr\{s(M-i, i)\}. \tag{4.4}$$

In a similar way, the handover dropping probability is equivalent to the probability of the system being in any of the boundary states on the lower edge of the Markov process because when the system state is on that edge, all conference rooms have been expended and no more handovers are accepted. The handover dropping probability

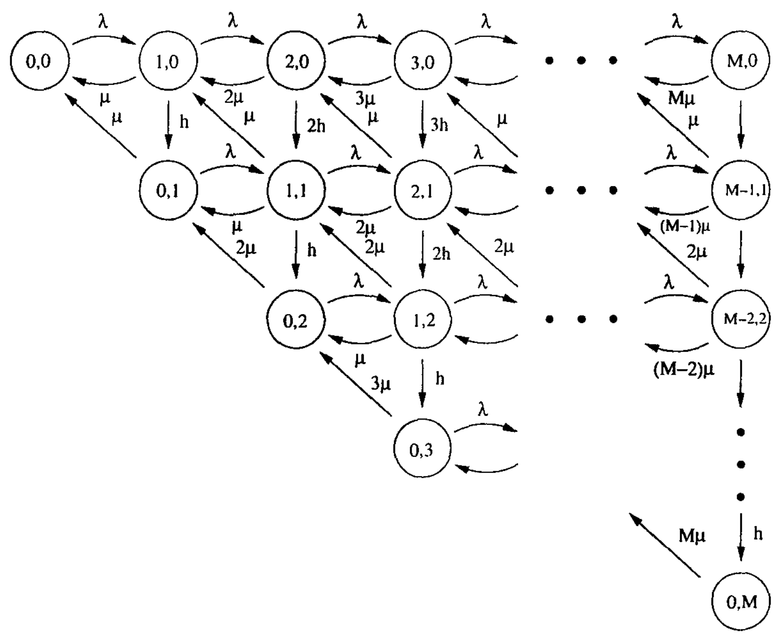


Figure 4.11: Markov Process for Conferencing GW, $N > M$

is thus given by

$$B_{handover} = \frac{\sum_{i=0}^{M-N} Pr\{s(i,N)\}ih}{\sum_{j=0}^{N}\left(\sum_{i=0}^{M-j} Pr\{s(i,j)\}ih\right)}. \tag{4.5}$$

Figure 4.11 describes the system with $N > M$ where $i, j = 0 \ldots M$. The interpretation of the system state and the transitions into and out of any state are as in the previous case. However in this case, if a new call is admitted then it is guaranteed a PSTN channel in the case of a handover because the number of conference rooms is less than the number of PSTN channels. Thus, the handover dropping probability is $B_{handover} = 0$. Since we are focusing on gateway performance, we have assumed a negligible cellular blocking probability. In this case, the new call blocking probability

Table 4.1: Simulation Parameters

Parameter	Value
N	23
M	25
Z	100
μ	0.0056 sec^{-1}
h	2.92×10^{-3} sec^{-1}

is equivalent to the probability of the system being in any of the boundary states on the right edge of the Markov process, since no more new calls can be admitted while the state of the system is residing on that edge. The new call blocking probability is therefore given by

$$B_{new} = \sum_{i=0}^{M} Pr\{s(M-i,i)\}. \tag{4.6}$$

In Figure 4.12 we give the corresponding Markov Process for DACH, where Z refers to the maximum allowable number of concurrent intra-enterprise calls. Also we define $X = \min(M, N)$ where M and N have the same meaning as before. Again, the interpretation of the system state and the transitions into and out of the state are as in the previous two cases. The DACH new call blocking probability is equivalent to the probability of the system being on any of the boundary states on the right edge of the Markov process and is given by

$$B_{new} = \sum_{i=0}^{X} Pr\{s(Z-i,X)\}. \tag{4.7}$$

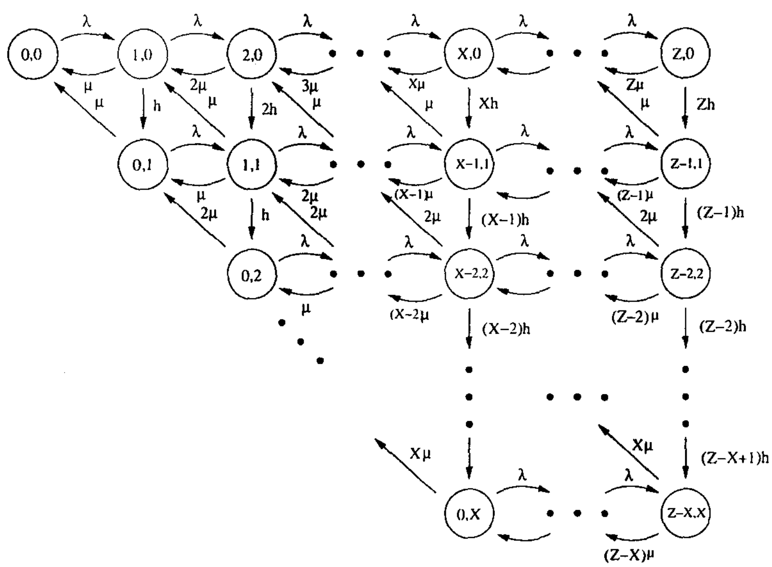


Figure 4.12: Markov Process for DACH

The handover blocking probability is equivalent to the probability of the system being in any of the boundary states on the lower edge of the Markov process because when the system state is at that edge, no more calls can handover since the lesser of the conference rooms and the PSTN channels has been totally consumed by calls which are already handed off. Interestingly, the handover blocking probability will never be zero in this case, since Z is always assumed to be larger than M and N. Therefore, the handover dropping probability is given by

$$B_{handover} = \frac{\sum_{i=0}^{Z-X} Pr\{s(i,X)\}ih}{\sum_{j=0}^{X}\left(\sum_{i=0}^{Z-j} Pr\{s(i,j)\}ih\right)}. \tag{4.8}$$

In Figure 4.13 the above analytic models are used to assess the performance of

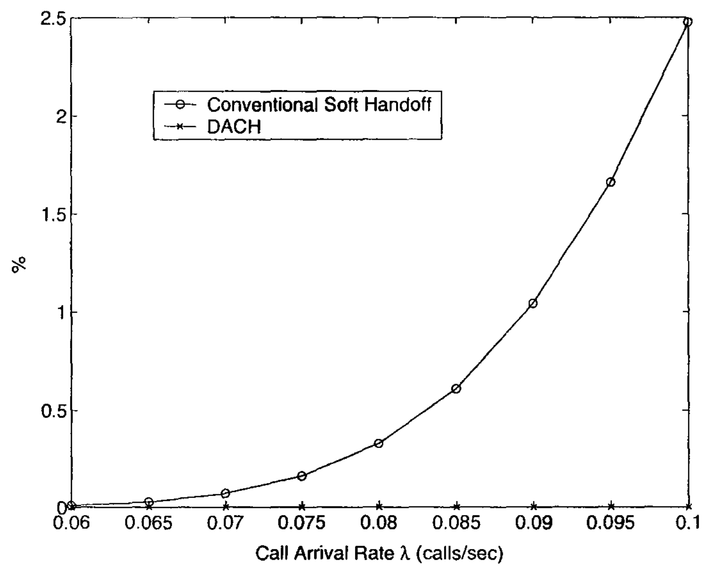


Figure 4.13: New Call Blocking Probability

DACH compared to conventional soft handover where the call duration and the handover rate are given in Table 4.1. In the figure, the superior call blocking probability of DACH is shown, whereas DACH and conventional soft handover both have zero handover dropping probabilities under the assumed conditions.

In the analytic models developed so far, all new call arrivals have been assumed to be intraenterprise calls. New call blocking and handover dropping probabilities are expected to depend heavily on the call composition. We now present discrete-event simulation results which study the effect of call composition on call blocking and dropping probabilities. All other assumptions regarding the new arrivals being Possion distributed, and exponentially distributed call durations are maintained. We

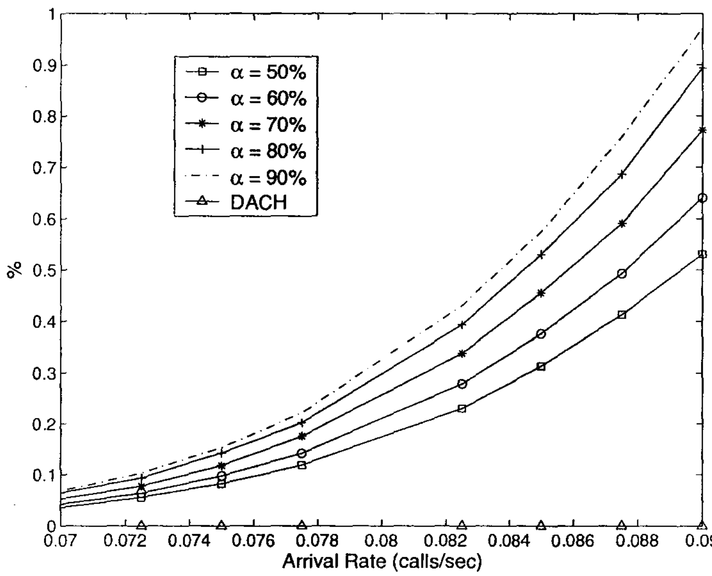


Figure 4.14: IN_IN Call Blocking Probability

assume that a fraction, α, of calls are intra-enterprise versus inter-enterprise. We allow IN_IN calls to handover with rate, h, while IN_OUT are not allowed to handover since we can safely assume full cellular coverage. Both call arrivals are assumed to be exponentially distributed. The parameters given in Table 4.1 are again used.

Figure 4.14 compares the IN_IN call blocking probability for different call compositions. Since the IN_IN call arrival rate is a function of the total call arrival rate λ and the IN_IN percentage of total calls α, the IN_IN call blocking probability increases with the values of λ and α. In the same figure, the IN_IN call blocking probability when DACH is used is shown to be zero regardless of the call composition because

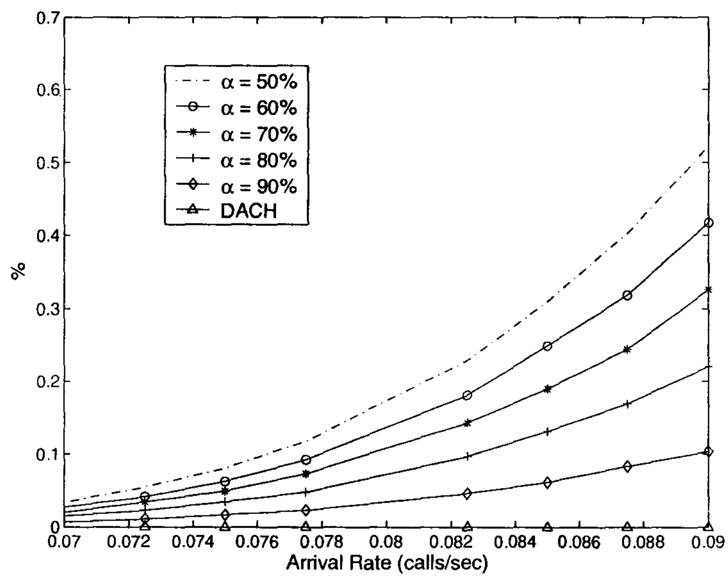


Figure 4.15: IN_OUT Call Blocking Probability

the sensitivity of IN_IN call blocking to the parameters α and λ is diluted due to our ability to admit $Z \gg M$ IN_IN calls. Figure 4.15 shows the IN_OUT call blocking probability for different call compositions. In the conventional soft handover scheme, new IN_OUT calls and IN_IN calls attempting to handover compete for the available PSTN channels. Therefore, as α increases, the IN_OUT percentage of total calls decreases, balancing out the increased percentage of IN_IN calls attempting to handover. When DACH is used, IN_IN calls do not compete with IN_OUT calls for conference rooms, leading to a near zero new IN_OUT call blocking. The handover dropping probability is found to be zero and is independent of the call composition at the assumed simulation parameter values regardless of whether DACH is used. However, it was found that at higher call arrival rates, DACH experiences a slightly higher handover dropping probability than the conventional scheme since the number of IN_IN calls admitted is much higher. We can conclude that at the assumed level of channel provisioning, DACH has gained the conferencing gateway a much better performance in terms of IN_IN and IN_OUT blocking probabilities without any evident tradeoff in terms of handover dropping probabilities.

We now study call blocking and handover dropping probabilities as functions of handover rates and consider the advantages of using DACH. We fix the call composition at $\alpha = 50\%$, vary the percentage of calls which perform a handover, ν, by

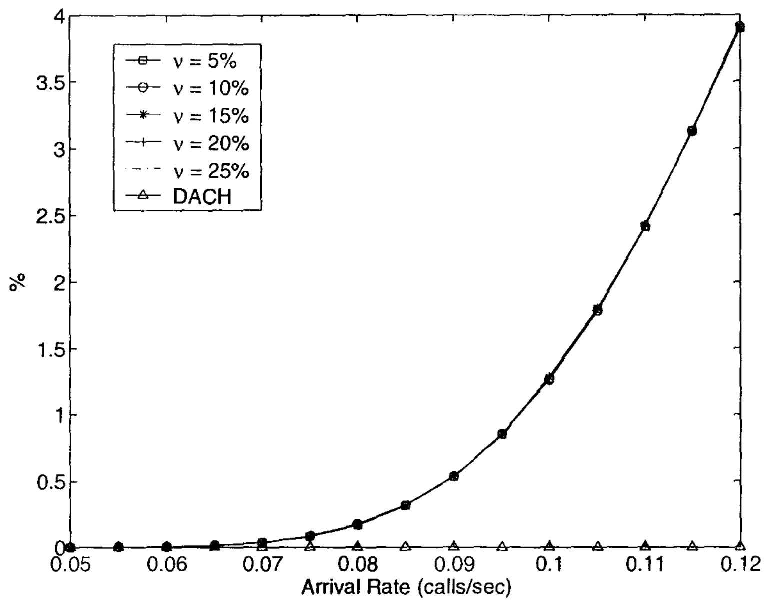


Figure 4.16: IN_IN Blocking Probability

varying the handover rate h, maintain the rest of the parameters unchanged as previously described and only allow IN_IN calls to handover. Figures 4.16 and 4.17 show the IN_IN and IN_OUT call blocking probabilities respectively. From the figures we can observe that the IN_IN and IN_OUT call blocking probabilities are identical for all the arrival rates, and are independent of the percentage of calls attempting to handover, ν. This result is easily explained by viewing the PBX as a two-stage concatenation of conference rooms followed by PSTN channels. Newly arriving IN_IN calls never require PSTN lines, whereas IN_OUT calls require a conference room followed by a PSTN channel, therefore, the conference rooms will be the bottleneck of the system as viewed by the IN_IN and IN_OUT call flows, explaining the identical blocking probability for both. In the case where ν is set to zero, the blocking probability for both flows can be found using the Erlang-B formula where the number of channels is equal to the number of conference rooms. Furthermore, even at our highest choice of ν (i.e., 25%), the conference rooms continue to be the bottleneck of the system, since the combined `IN_OUT` calls and the calls attempting to handover are still much smaller than the calls arriving at the conference rooms, explaining the independence of the new call blocking probability from the handover rate. The DACH `IN_IN` and DACH `IN_OUT` call blocking probabilities are also displayed on the same figure and are shown to be consistently zero implying that DACH has effectively decoupled the `IN_IN` and `IN_OUT` blocking probabilities from the arrival rates (at least for our choice of arrival rates) and from the handover rates. Finally, our simulation results show that the handover dropping probability varies in the negligible range of 0.0005 % and .0025 % for the different handover rates regardless of whether DACH was used. This negligible handover dropping probability affirms our previous explaination regarding the conference rooms being the bottleneck of the system and not the PSTN channels. In other words, if an `IN_IN` call attempts to handover it is most likely going to find an available PSTN channel and conduct a successful handover.

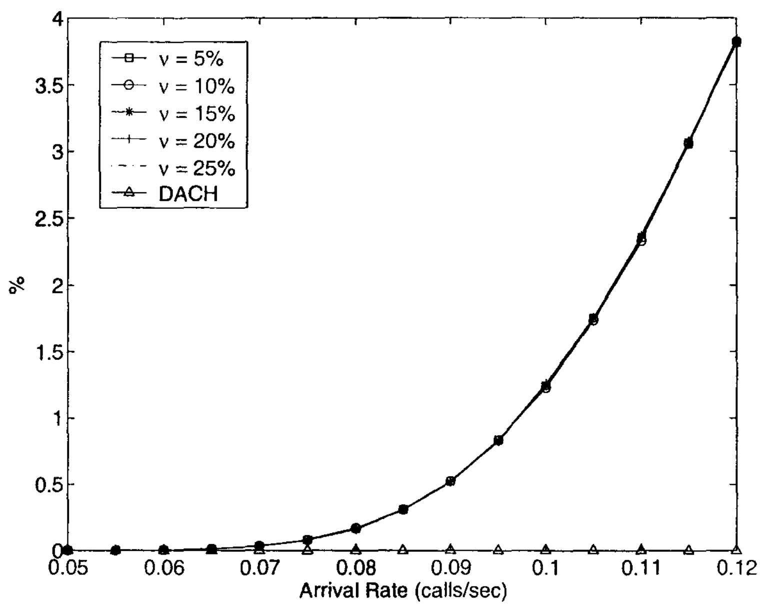


Figure 4.17: IN_OUT Blocking Probability

4.6 Conclusions

In this chapter we have considered soft vertical handover for dual-mode cellular/WLAN handsets where the call is anchored in the enterprise using a PSTN gateway/PBX. A dual-mode handset testbed was used to show that GSM-to-WiFi and WiFi-to-GSM vertical handovers can be performed seamlessly via the use of a hairpin device that performs the required SIP signaling. In both cases, we have obtained a 20 msec controllable overlap between the two legs from the DM to the conference room before the handover was completed. Conventional conferenced handover requires that all intra-enterprise calls be routed through the gateway when the call is established. We proposed a more scalable mechanism whereby active calls are handed off into the conference bridge just prior to the initiation of the vertical handover. An analytic characterization of the scalability of the improvement mechanism was given and was shown to reduce the mean new call blocking and handover dropping probabilities by orders of magnitude when compared to the original conferencing mechanism. Furthermore, we have shown that DACH effectively decouples the handover rate from the call blocking and dropping probabilities, and maintains these probabilities at least two orders of magnitude less than that of the original conferencing mechanism.

Chapter 5

Limited WLAN Data Rate Usage

5.1 Overview

WLAN deployments carrying traffic which has strict capacity requirements such as VoIP must ensure that a minimum guaranteed data rate is seen across the deployment. This is done by restricting the basic service set rates advertised by the AP and forcing the MS to handover rather than to drop below the set limits, otherwise the link is terminated. These mechanisms create an artificially high value of S_{ll} which makes VHO much more difficult. In many situations achieving acceptable VHO success probability requires maintaining the WLAN link at S_{ll} values close to the noise floor. This conflict between WLAN capacity and VHO performance creates a dilemma where either capacity or VHO success probability has to be sacrificed.

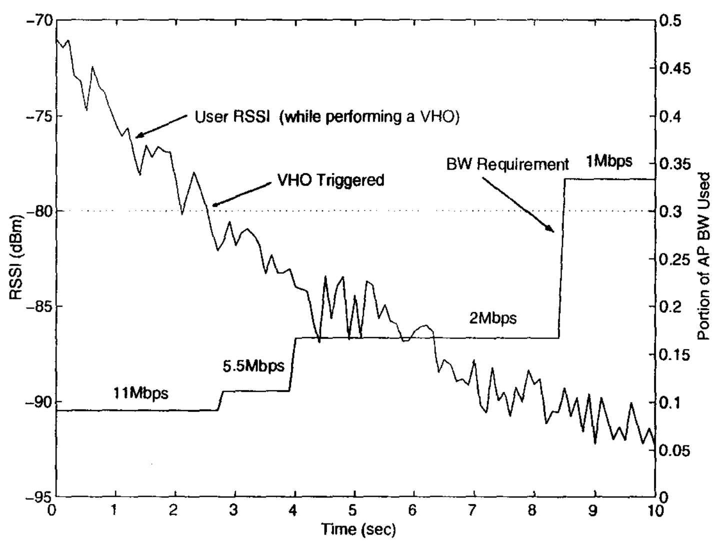


Figure 5.1: WLAN Link Rate-Adjusting during VHO Procedure

Figure 5.1 illustrates how the bandwidth requirement for a VoIP call changes as the link signal strength (i.e., RSSI) decreases during a VHO. The left-hand axis shows an example of the smoothed RSSI as a user moves out of the WLAN. The right-hand axis displays the corresponding fraction of total AP bandwidth that is required to support the call. For simplicity we have shown the IEEE 802.11b case. It can be seen that once the link has rate-adjusted to 1 Mbps, it requires over 1/3 of the (IEEE 802.11b/g) total AP bandwidth [HT04]. Although this is a huge overhead, it is only required during the VHO execution time which is small compared with typical call holding times. In this chapter we propose and assess the performance of a Limited Data Rate Use (LDU) algorithm where WLAN APs restrict their advertised rates and only allow stations performing VHO to rate-adjust to rates lower than those advertised. This algorithm allows mobile stations to perform successful vertical handover by holding the WLAN link at low RSSI values without adversely affecting the WLAN capacity.

5.2 The LDU Algorithm

The LDU algorithm is shown in Figure 5.2. In a conventional system the AP will terminate an MS's association if the link quality drops below S_{ll}. In an LDU-compliant AP however, the AP-MS link quality is permitted to drop to a value as low as $S_{disassociate}$ ($\leq S_{ll}$) for up to a time period denoted by T_{LDU}. In the LDU algorithm, a timer is kept in the AP for each associated mobile station, $\mathtt{MS}_i$, and is denoted by $\mathtt{MS}_i\mathtt{.Timer}$. If $\mathtt{MS}_i$'s link quality, $\mathtt{MS}_i\mathtt{.RSSI}$, drops below S_{ll}, the timer begins running until it reaches a value of T_{LDU}. Note that if the link recovers before the timer reaches T_{LDU} then the timer is reset and the process continues. However, if the link quality remains below S_{ll} for this time period or if the link quality ever drops below $S_{disassociate}$, then $\mathtt{MS}_i$ is disassociated. The intent behind the LDU Algorithm is to permit mobile stations to hold the WLAN link at values less than S_{ll} for a limited period of time, so that vertical handover can be successfully accomplished. Enforcing a fixed time period prevents stations from holding the link at low data rates for too long which might adversely affect the capacity of the AP. The value of T_{LDU} must be

```
while (MS_i is associated)
  begin
    Set MS_i.Timer = 0
    while (MS_i.RSSI < S_ll)
    begin
      Run MS_i.Timer
      if (MS_i.Timer > T_LDU || MS_i.RSSI < S_disassociate)
      begin
        Disassociate MS_i.
      end
    end
  end
```

Figure 5.2: WLAN Limited Data Rate Use (LDU) Algorithm

greater than the time needed for VHO, t_{vho}, with a high probability.

5.2.1 LDU Algorithm Performance Analysis

In this section we compare the new call blocking probability of an AP that is running the LDU algorithm to that of an AP that is not running the LDU algorithm. We first present Markovian models that compare an LDU-compliant AP to an AP that does not implement the LDU algorithm in a simple situation. Following this we present detailed simulation results using our collected propagation data.

The capacity of the AP is discretized such that a new call consumes a single "channel" at the highest data rate. The AP is assumed to handle at most N concurrent calls at this data rate. An active call is allowed to roam outside the original designated coverage area and to rate-adjust to lower data rates. An AP that is not implementing the LDU algorithm will restrict the advertised data rates to maintain capacity, whereas an LDU-compliant AP will allow stations which have associated with the AP at high data rates to maintain the link when an active session starts to experience lower data rates. To model this behavior, we consider a simple example where the non-LDU AP is assumed to advertise two data rates, R_1 and R_2. An LDU-compliant AP is assumed to advertise a third rate, R_3, but this is only permitted for MSs that have initiated a VHO. We assume that the VHO is triggered when the MS link drops from R_1 to R_2. If the MS cannot maintain the data rate at R_2 before the call has completed or the VHO has succeeded, then the call is forcefully terminated. In contrast, the AP with the LDU algorithm allows the MS to rate adjust down to R_3 and to maintain the link until either the call is completed or the VHO has succeeded.

Figure 5.3 shows the Markov process describing the non-LDU AP. New calls arrive according to a Poisson process with mean arrival rate, λ. A given call may complete after an exponentially distributed time $1/\mu$. An active call may also initiate a VHO at a rate of h. Immediately after initiating the VHO, the MS starts to experience R_2 and the call starts to consume X channels instead of one channel that it consumed at R_1. VHO events are assumed to be rare in comparison with new call arrivals, thus only a single VHO event can be outstanding at any given time. A call that has shifted to a lower data rate can complete with a rate of μ, finish the VHO successfully with rate ν or be dropped with rate ϵ if it requires rate adjusting to R_1. In Figure 5.3, the state is described by $s(i, j)$ where i is the number of new calls operating at R_1.

Additionally, if $j = 0$ then there is no call that is actively handing over, otherwise there is a single call that is attempting a handover.

From Figure 5.3, the new call blocking probability is found by summing over the probabilities of being in states $s(N, 0)$ and $s(N-X, X)$ since no more new calls can be admitted to the system, i.e.,

$$B_{NewCall}^{\overline{LDU}} = Pr\{s(N,0)\} + Pr\{s(N-X,X)\}. \tag{5.1}$$

Handover dropping occurs when one of two events occur. First, when the MS operating at R_2 needs to switch to R_3, which is not supported by the non-LDU AP. Second, when the MS is experiencing R_1 and needs to rate adjust to R_2, but there are no available channels. Thus, this handover dropping probability is given by

$$B_{handover}^{\overline{LDU}} = \frac{\sum_{i=N-X+2}^{N} Pr\{s(i,0)\}ih + \sum_{i=0}^{N-X} Pr\{s(i,X)\}\epsilon}{\sum_{i=0}^{N} Pr\{s(i,0)\}ih}. \tag{5.2}$$

In Figure 5.4 the Markov process is modified to describe an LDU-compliant AP. An additional row is added to allow for MSs to rate adjust to R_3. Therefore, a MS that is experiencing R_2 and needs to rate adjust is allowed to operate at R_3 and the call starts to consume Y channels instead of the X channels it consumed at R_2.

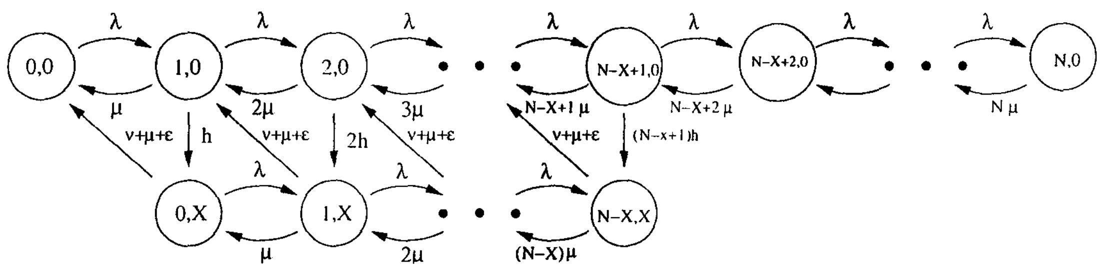


Figure 5.3: Markov Process without the LDU Algorithm

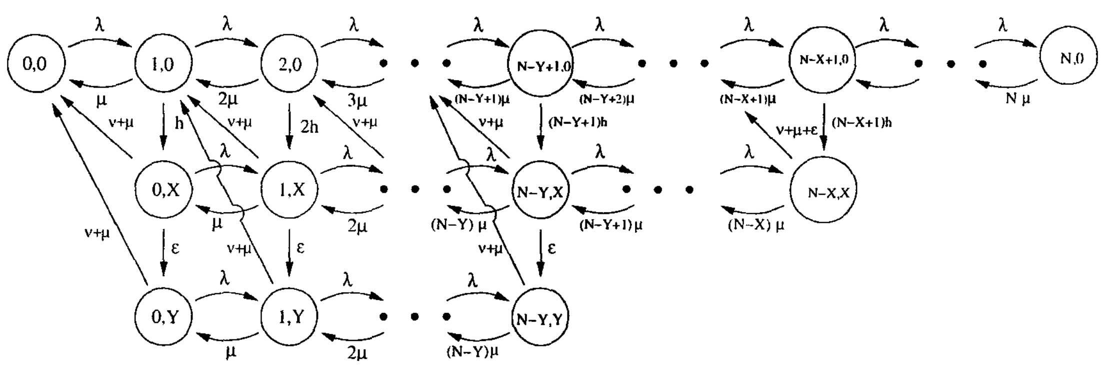


Figure 5.4: Markov Process for the LDU Algorithm

Therefore, the new call blocking probability becomes

$$B_{NewCall}^{LDU} = Pr\{s(N,0)\} + Pr\{s(N-X,X)\} + Pr\{s(N-Y,Y)\}. \tag{5.3}$$

Since there is no rate restriction in this case (i.e., a MS can operate at any of the three available rates), handover dropping only occurs when a call attempts to rate adjust and there are no more channels available. Thus the handover dropping probability can be written as

$$B_{handover}^{LDU} = \frac{\sum_{i=N-X+2}^{N} Pr\{s(i,0)\}ih + \sum_{i=N-Y+1}^{N-X} Pr\{s(i,X)\}\epsilon}{\sum_{i=0}^{N} Pr\{s(i,0)\}ih}. \tag{5.4}$$

In Table 5.1 the values of the parameters used in evaluating the above probabilities are shown. The three rates chosen are a subset of the rates available on a standard IEEE 802.11g AP. We assume that at R_1 the AP can support a maximum of 54 calls (thus the 1 channel/call assumption), 27 calls at R_2 and 3 calls at R_3. The values of X and Y are the ratios of R_1 to R_2 and R_1 to R_3 respectively. The value of ν is taken as the inverse of a chosen t_{vho} of 10 seconds. The value of T_{LDU} is assumed to be longer than the amount of time required by any call to complete its VHO.

In Figure 5.5 the new call blocking probability for an LDU-compliant AP is compared to that of a non-LDU AP for two different handover rates, h. The AP implementing the LDU algorithm has a higher new call blocking probability than the

Table 5.1: LDU Algorithm Markov Parameters

Parameter	**Value**
R_1	54 Mbps
R_2	9 Mbps
R_3	1 Mbps
N	54 Channels
X	2 Channels
Y	18 Channels
μ	0.0056 sec^{-1}
ν	0.1 sec^{-1}
ϵ	0.033 sec^{-1}

non-LDU AP. This difference becomes larger as the handover rate increases since a larger fraction of the LDU-compliant AP channels are consumed at these higher rates. However, the absolute value of this difference is nominally small. For example, when $h = 7.31 \times 10^{-5}s^{-1}$ and $\lambda = 0.22$ calls/second, the difference between the LDU-compliant AP and the non-LDU AP in new call blocking probability is less than 0.05%. When $h = 5.5 \times 10^{-4}s^{-1}$ and λ is still the same, the difference is 0.25%. It is interesting to observe that the two LDU algorithm curves corresponding to the two different handover rates cross at $\lambda = 0.218$ calls/second. This is explained by noting that at higher handover rates, the system starts to experience higher capacity due to the early termination of handover calls. This additional capacity results in a lower new call blocking probability.

In Figure 5.6 the handover dropping probability for an LDU-compliant AP is compared to that of a non-LDU AP for two different handover rates, h. At call arrival rates less than 0.14 calls/second, the LDU-compliant AP has more than two orders

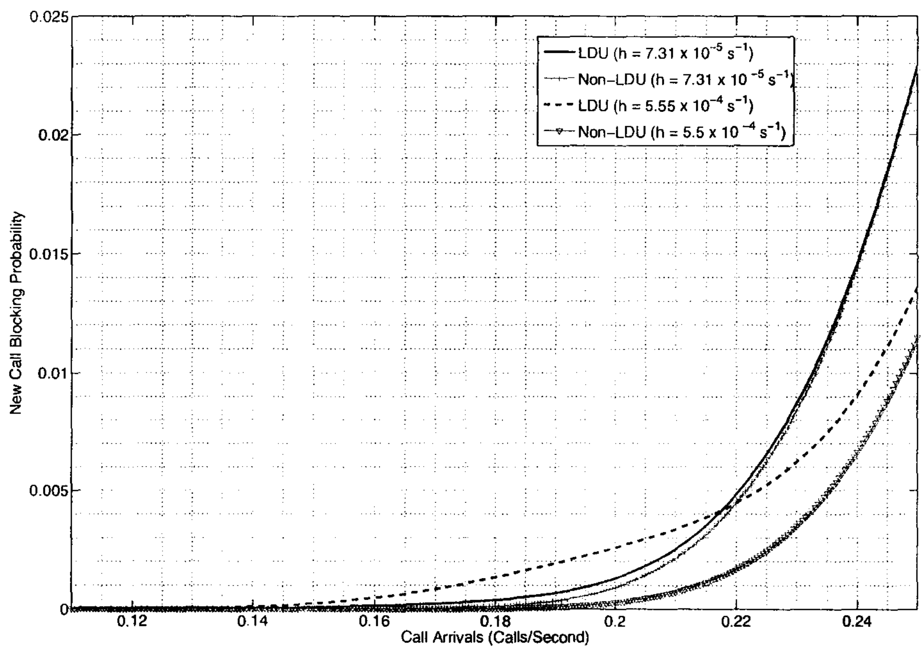


Figure 5.5: New Call Blocking Probability

of magnitude improvement when compared to the non-LDU AP. This is typically the region of interest since an acceptable handover dropping is in the order of 0.1%. However, even at higher call arrival rates, the LDU-compliant AP consistently shows a better handover dropping probability than the non-LDU case. At higher handover rates and similar to the case of new call blocking probability, the LDU-compliant AP experiences better handover dropping probability due to the early termination of handover calls.

To further evaluate the performance effect of the proposed algorithm we now

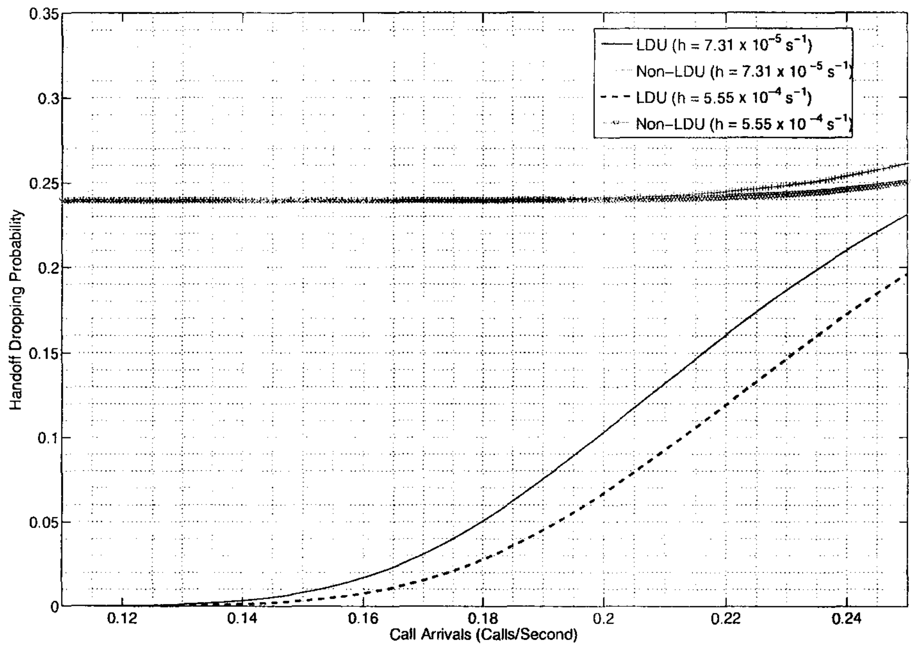


Figure 5.6: Handover Dropping Probability

present detailed simulation results using our collected propagation data. We assume that a newly admitted $\texttt{MS}_i$ with $\texttt{MS}_i\texttt{.RSSI} > S_{min}$ consumes one normalized channel during a VoIP connection. When its RSSI drops below S_{min} then it will consume more bandwidth due to physical layer rate adjusting. New calls arrive according to a Poisson process with an average rate of $\lambda_{new} = 0.094$ calls per second. The call duration is assumed to be exponentially distributed with an average of 180 seconds and VHO latency, t_{vho}, is chosen as a fixed value of 7 seconds [STK+]. We use the RSSI data collected in our measurements to evaluate the performance of the system.

Table 5.2 shows the result of our evaluation for four different S_{pod} values. These S_{pod} values were chosen since they result in acceptable dropping and blocking probabilities. The results show the VHO dropping probability derived experimentally from the collected traces. It is clear that by allowing the users to hold the link to the lowest supportable bit rate for IEEE 802.11b/g, (i.e., 1 Mbps), it is possible to reduce VHO dropping by roughly an order of magnitude. This significant improvement becomes even more valuable when looking at the penalty imposed on the new call blocking rate which is an increase by less than a factor of two. For example, for the $S_{pod} = -50$ dBm case when $S_{tp} = -55$ dBm, as shown in Table 5.1(a), the VHO dropping rate without the LDU algorithm (i.e., $S_{disassociate} = -80$ dBm corresponding to a minimum bit rate of 11 Mbps for the case of our measurement equipment), results in a VHO dropping rate of 15.7%. When the LDU algorithm is used (corresponding to $S_{disassociate} = -93$ dBm), one can obtain about a 1% VHO dropping probability while new call blocking does not change.

When S_{pod} is lowered to -60 dBm and the trigger point is set to -65 dBm, the improvement is even more significant. The VHO dropping probability goes from about 46% to roughly 2.5%. In this case the new call blocking probability is affected and increases from 1% to about 2.2%. A similar LDU algorithm result is seen for the $S_{PoD} = -65$ dBm case for VHO dropping although the non-LDU case is about 72%. The slight increases seen in blocking probability are far more acceptable than the

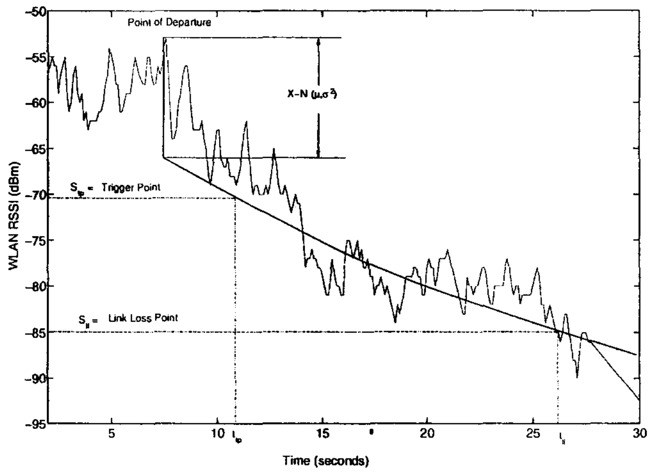


Figure 5.7: *Building Exit RSSI Model Example*

increase one could expect if users were allowed to operate at 1 Mbps (i.e., −93 dBm) under no time limitation.

5.3 Building Exit Model and Parameters

In this section and using the collected data, we extract some simple model parameters that can be used to accurately compute vertical handover behavior in certain common situations. At a building exit, the mobile user typically experiences a sudden drop in received signal strength immediately after exiting through the PoD due to attenuation associated with the exit. An example of this is shown in Figure 5.7 which illustrates the decrease in RSSI as a station leaves the building. The PoD is marked at −53 dBm (i.e., $S_{pod} = -53$ dBm) and an abrupt drop in average RSSI is seen at that point. Following that drop, our results tend to follow a lognormal path loss model as would normally be the case [Rap95]. This is also seen in the figure as the gradual loss in RSSI as we move further to the right. In the figure we also show the vertical handover trigger value, S_{tp} and the link loss point, S_{ll}, as previously discussed. In Figure 5.7, S_{tp} is shown to be below the abrupt exit drop and hence the handover is triggered at some point *after* the station has moved further away from the building. In some cases, however, S_{tp} is crossed when passing through the building exit. In this case the handover is triggered immediately at the PoD.

Our observations have shown that the abrupt attenuation suffered when passing through the exit is well approximated as a random variable that is normally distributed. Furthermore, since the classical logdistance path loss model has a zero mean normally distributed component, we lump together the lognormal shadowing and PoD attenuation components as one random variable, X, which is normally distributed (in dB), i.e., $N(\mu, \sigma^2)$. Since we are considering the propagation segment starting after the PoD we can approximate the received signal power as follows,

$$S_{rx}(d) = P_{tx} - P_0 - 10n\log(d_{exit} + d) - X \quad \text{for} \quad d > 0, \tag{5.5}$$

where P_{tx} is the access point transmission power, P_0 is the path loss at the reference distance of $1m$, and n is the propagation exponent (approximately 4.5 in our measurements). d_{exit} is the distance from the access point to the PoD and d is distance the receiver has traveled along the outdoor segment. The relationship between d_{exit} and S_{pod} is given by

$$S_{Pod} = P_{tx} - P_0 - 10n\log(d_{exit}). \tag{5.6}$$

Note that the attenuation due to exiting through the PoD only applies to distances larger than d_{exit}. We assume that the mobile user crosses the PoD at $t = 0$ with a constant speed v, thus $d = vt$. We use the propagation model parameters derived above to derive an analytic expression for the VHO success probability, which is used in Section 5.2.1. At time $t = 0^+$ (i.e., immediately after exiting the PoD), the signal strength can be found using Equation 5.5. Depending on the drop caused by X (since $d \approx 0$ at $t = 0^+$), we define $\mathcal{T}_{PoD}$ to be the event that the VHO is triggered at the PoD. Then,

$$\begin{aligned} &S_{tp} \geq S_{pod} - X, \quad \text{if } \mathcal{T}_{PoD}, \text{ and,} \\ &S_{tp} < S_{pod} - X, \quad \text{otherwise.} \end{aligned} \tag{5.7}$$

The success probability of a VHO event can then be defined as

$$P_{success} = Pr\{\mathcal{T}_{PoD} \cap (t_{margin} \geq t_{vho})\} + Pr\{\overline{\mathcal{T}_{PoD}} \cap (t_{margin} \geq t_{vho})\}. \tag{5.8}$$

To evaluate the first term in Equation 5.8 we note that $\mathcal{T}_{PoD}$ corresponds to VHO triggering immediately upon exiting, thus $t_{margin} = t_{ll}$ (i.e., $t_{tp} = 0$). From Equation 5.5,

$$t_{ll} = \frac{10^{\frac{P_{tx}-S_{ll}-P_0-X}{10n}}}{v} - \frac{d_{exit}}{v}. \tag{5.9}$$

If we choose $\Psi = t_{ll} + \frac{d_{exit}}{v}$ and $\kappa = \frac{1}{v}10^{\frac{P_{tx}-P_0-S_{ll}}{10n}}$ then

$$\Psi = \kappa 10^{\frac{-X}{10n}},$$
$$\log(\Psi) = \log(\kappa) + \frac{-X}{10n}. \tag{5.10}$$

However since $X \sim N(\mu, \sigma^2)$ then

$$\log(\Psi) \sim N(\log(\kappa) - \frac{\mu}{10n}, \frac{\sigma^2}{100n^2}). \tag{5.11}$$

From Equation 5.7 we know that $\frac{-X}{10n} \leq \frac{S_{tp}-S_{pod}}{10n}$, which then gives

$$\Psi \leq \kappa 10^{\frac{S_{tp}-S_{pod}}{10n}}. \tag{5.12}$$

Therefore the first term in Equation 5.8 can be rewritten as

$$Pr\{\Psi \leq \kappa 10^{\frac{S_{tp}-S_{pod}}{10n}} \cap \Psi \geq t_{vho} + \frac{d_{exit}}{v}\} =$$
$$Pr\{\log(t_{vho} + \frac{d_{exit}}{v}) \leq \log(\Psi) \leq \log(\kappa) + \frac{S_{tp} - S_{pod}}{10n}\}. \tag{5.13}$$

Equation 5.13 can be easily evaluated using the cumulative distribution function of the normally distributed random variable $\log(\Psi)$.

Next we derive the success probability in the $\overline{T_{PoD}}$ case, i.e., when the station does not trigger at the PoD, as described in Equation 5.7. In this case $t_{margin} = t_{ll} - t_{tp}$. From Equation 5.5 we find that,

$$S_{tp} - S_{ll} = -10n\log(d_{exit} + vt_{tp}) + 10n\log(d_{exit} + vt_{ll}). \tag{5.14}$$

In this case, $t_{margin} = t_{ll} - t_{tp}$, and by manipulating Equation 5.14, we can write

$$d_{exit} + vt_{tp} = 10^{\frac{P_{tx} - P_0 - S_{tp} - X}{10n}}, \tag{5.15}$$

$$t_{margin} = \frac{1}{v}(10^{\frac{S_{tp} - S_{ll}}{10n}} - 1)(d_{exit} + vt_{tp}), \tag{5.16}$$

$$t_{margin} = \frac{1}{v}(10^{\frac{S_{tp} - S_{ll}}{10n}} - 1)10^{\frac{P_{tx} - P_0 - S_{tp}}{10n}}10^{\frac{-X}{10n}}. \tag{5.17}$$

If we let $\varphi = \frac{1}{v}(10^{\frac{S_{tp} - S_{ll}}{10n}} - 1)10^{\frac{P_{tx} - P_0 - S_{tp}}{10n}}$ then Equation 5.17 can be rewritten as

$$t_{margin} = \varphi 10^{\frac{-X}{10n}}, \tag{5.18}$$

$$\log(t_{margin}) = \log(\varphi) - \frac{X}{10n}. \tag{5.19}$$

Now since $X \sim N(\mu, \sigma^2)$ then $\log(t_{margin}) \sim N(\log(\varphi) - \frac{\mu}{10n}, \frac{\sigma^2}{100n^2})$. Using Equation

5.7 we know that $\frac{X}{10n} < \frac{S_{pod}-S_{tp}}{10n}$ which can be used to obtain

$$t_{margin} > \varphi 10^{\frac{S_{tp}-S_{pod}}{10n}}. \tag{5.20}$$

Therefore, the second term in Equation 5.8 can be expressed as

$$Pr\{\log(t_{margin}) > (\log(\varphi) + \frac{S_{tp}-S_{pod}}{10n}) \cap \log(t_{margin}) > \log(t_{vho})\} = \\ Pr\{\log(t_{margin}) > \max(\log(\varphi) + \frac{S_{tp}-S_{pod}}{10n}, \log(t_{vho}))\}. \tag{5.21}$$

This can be evaluated using the cumulative distribution function of the randomly distributed variable $\log(t_{margin})$.

In Table 5.2 we also show the results of the building exit model computation. This was obtained using the model and its parameters derived in Section 5.3 from our measurements. The predictions using our experimental and model-based values are very similar. However, as the value of S_{ll} decreases, the experimental value of the failure probability is higher than that evaluated using the analytic model. This slight discrepancy between the two values is attributed to our analytic model not accounting for the missed beacons which cause more failed handovers when the S_{ll} is close to the noise floor. However, the results from the model give surprisingly good correspondence over a wide range of parameter values.

5.4 Conclusions

The limited data rate use algorithm proposed in this chapter allows mobile stations performing VHO to maintain the WLAN link at low RSSI values without adversely affecting the WLAN capacity. This behavior is ordinarily prevented in APs trying to maintain the WLAN capacity by restricting their basic service set rates advertised. Results obtained from analytic models and simulations show that a WLAN AP which implements our LDU algorithm can reduce the VHO dropping probability by roughly an order of magnitude. These improvements come in lieu of doubling the new call blocking probability which is more than acceptable given the absolute values observed of this new call blocking probability. Using the collected data, we have also extracted some simple model parameters which can be used to accurately compute vertical handover behavior in certain common situations.

(a) $S_{PoD} = -50\text{dBm}$ $S_{tp} = -55\text{dBm}$

LDU

$S_{disassociate}$	−80dBm	−85dBm	−90dBm	−93dBm
Link Loss (Model)	0.1206	0.0119	0.0004	0.0000
Link Loss (Experimental)	0.1569	0.0196	0.0196	0.0098
New Call Blocking Loss	0.0110	0.0105	0.0103	0.0107

No LDU

(b) $S_{PoD} = -55\text{dBm}$ $S_{tp} = -60\text{dBm}$

LDU

$S_{disassociate}$	−80dBm	−85dBm	−90dBm	−93dBm
Link Loss (Model)	0.2094	0.0290	0.0014	0.0001
Link Loss (Experimental)	0.1636	0.0273	0.0273	0.0000
New Call Blocking Loss	0.0104	0.0111	0.0179	0.0099

No LDU

(c) $S_{PoD} = -60\text{dBm}$ $S_{tp} = -65\text{dBm}$

LDU

$S_{disassociate}$	−80dBm	−85dBm	−90dBm	−93dBm
Link Loss (Model)	0.4694	0.0711	0.0053	0.0007
Link Loss (Experimental)	0.4590	0.1000	0.0250	0.0250
New Call Blocking Loss	0.0105	0.0135	0.0211	0.0220

No LDU

(d) $S_{PoD} = -65\text{dBm}$ $S_{tp} = -70\text{dBm}$

LDU

$S_{disassociate}$	−80dBm	−85dBm	−90dBm	−93dBm
Link Loss (Model)	0.8770	0.1655	0.0198	0.0034
Link Loss (Experimental)	0.7213	0.2049	0.0328	0.0246
New Call Blocking Loss	0.0106	0.0143	0.0216	0.0221

No LDU

Table 5.2: New Call and Link Loss Probability

Chapter 6

Conclusions

In this thesis we considered the soft vertical handover of dual-mode cellular/WLAN handsets in loosely-coupled, enterprise-centric architectures, where newly established calls are anchored at the enterprise using a PSTN gateway/PBX. In this architecture a WLAN-to-WWAN handover poses a strict timing constraint due to the abruptness of WLAN signal strength degradation when crossing a WLAN hotzone boundary. We presented a measurement-based study describing the behavior of WLAN signal behavior in WLAN-to-WWAN transition regions in the context of vertical handover. Using WLAN signal strength traces collected as mobile stations traversed exit regions we quantified the vertical handover success probability of classical threshold-based VHO algorithms. It was found that it is very difficult to successfully complete the handover without carefully engineering the WLAN deployment, highlighting the fundamental difference between VHO and handover in homogeneous networks.

We also showed that in most cases the classical WLAN-to-WWAN VHO algorithms come short of achieving success probabilities similar to those found in homogeneous WWAN systems. The exception to these cases (i.e., where $P_{success}$ was comparable to those of the homogeneous WWAN systems) restrict the designer to a small set of S_{tp} and S_{ll} values and is achieved only by the mobile user triggering as early as possible and continuing to drag the WLAN link to values very close the noise floor, adversely affecting the system capacity. To preserve system capacity while allowing the MS to maintain the WLAN at very low RSSI values during the VHO procedure, we proposed a "limited data rate use" algorithm which greatly improved the probability of seamless handover. Through analytic modelling and simulations we showed that VHO success probability could experience many orders of magnitude of improvement without impacting the original system capacity. Additionally, we summarized our observations of the RF signal behavior in the transition region by proposing a propagation model which was used to accurately compute vertical handover behavior in certain common situations.

An architecture was proposed to allow for bi-directional handovers in loosely-coupled systems using conference bridging available at enterprise PSTN gateways. A dual-mode handset testbed was used to show that GSM-to-WiFi and WiFi-to-GSM vertical handovers could be performed seamlessly via the use of a hairpin device that performs the required SIP signaling. In both cases, we have obtained 20 msec controllable overlap between the two legs from the DM to the conference room before the handover was completed. Conventional conferenced handover requires that all intra-enterprise calls be routed through the gateway at call establishment. We proposed a more scalable mechanism (i.e., DACH) whereby active intra-enterprise calls were dynamically anchored at the the conference room just prior to the initiation of the vertical handover. An analytic characterization of the scalability of DACH was given. It was shown that orders of magnitude reduction in new call blocking and handover dropping probabilities are possible. Furthermore, we have shown that DACH effectively decouples the handover rate from the call blocking and dropping probabilities, and maintains these probabilities at least two orders of magnitude less than those of the original conferencing mechanism.

Finally, this work has spurred several research directions including the capacity deficit problem resulting from the mobile user consuming more capacity when dragging the WLAN link during the VHO. We are also considering VHO-friendly WLAN deployment algorithms such that the WLAN propagation in the PoD regions allows for acceptable success probabilities when using threshold-based triggering.

Bibliography

[BCI04] M. Bernaschi, F. Cacace, and G. Iannello. Vertical Handoff Performance in Heterogeneous Networks. In *Proc. of ICPPW'04*, pages 100–107, Aug. 2004.

[BD01] B. Biggs and R. Dean. SIP Call Control: Call Handoff (Internet Draft), 2001.

[BDA05] N. Banerjee, S. K. Das, and A. Acharya. SIP-Based Mobility Architecture for Next Generation Wireless Networks. In *Proc. of IEEE PerCom'05*, pages 181–190, Mar. 2005.

[BHJ03] H. Bing, C. He, and L. Jiang. Performance Analysis of Vertical Handover in a UMTS-WLAN Integrated Network. In *Proc. of IEEE PIMRC'03*, volume 1, pages 187–191, Sept. 2003.

[BSL05] D. Bultmann, M. Siebert, and M. Lott. Performance Evaluation of Accurate Trigger Generation for Vertical Handover. In *Proc. of IEEE PIMRC'05*, volume 3, pages 2034–2039, Sept. 2005.

[Cam01] G. Camarillo. *SIP Demystified.* McGraw-Hill Professional, 2001.

[CDM04] A. Calvagna and G. Di Modica. A User-Centric Analysis of Vertical Handovers. In *Proc. of ACM WMASH'04*, pages 137–146, New York, NY, USA, 2004. ACM Press.

[CSC+04] L. Chen, T. Sun, B. Chen, V. Rajendran, and M. Gerla. A Smart Decision Model for Vertical Handoff. In *Proc. of ANWIRE'04*, May 2004.

[CVS+04] R. Chakravorty, P. Vidales, K. Subramanian, I. Pratt, and J. Crowcroft. Performance Issues with Vertical Handovers - Experiences from GPRS Cellular and WLAN Hot-spots Integration. In *Proc. of IEEE PerCom'04*, pages 155–164, Washington, DC, USA, 2004. IEEE Computer Society.

[Dig] Digium. Asterisk - Open Source PBX. http://www.asterisk.org/.

[ETS01] Requirements and Architectures for Interworking between HIPERLAN/3 and 3rd Generation Cellular Systems. Technical Report TR 101 957, ETSI, Aug. 2001.

[Eve94] D. E. Everitt. Traffic Engineering of The Radio Interface for Cellular Mobile Networks. *Proceedings of the IEEE*, 82:1371–1382, 1994.

[FCL98] Y. Fang, I. Chlamtac, and Yi-Bing Lin. Channel Occupancy Times and Handoff Rate for Mobile Computing and PCS Networks. *IEEE Transactions on Computers*, 47:679–692, 1998.

[Gas05] M. Gast. *802.11 Wireless Networks: The Definitive Guide*. Second edition, 2005.

[GPVD99] E. Guttman, C. Perkins, J. Veizades, and M. Day. Service Location Protocol, Version 2 (RFC - 2608), 1999.

[GS06] R. Guha and S. Sarkar. Characterizing Temporal SNR Variation in 802.11 Networks. In *Proc. of IEEE WCNC'06*, volume 3, pages 1408–1413, Apr. 2006.

[Gu87] R. Gurin. Channel Occupancy Time Distribution in a Cellular Radio System. *IEEE Transactions on Vehicular Technology*, 36:89–99, Aug. 1987.

[GV96] A. Gulbrandsen and P. Vixie. A DNS RR for Specifying the Location of Services (DNS SRV), (RFC - 2052), 1996.

[HHSDH04] C. Hyun-Ho, O. Song, and C. Dong-Ho. A Seamless Handoff Scheme for UMTS-WLAN Interworking. In *Proc. of IEEE GLOBECOM'04*, volume 3, pages 1559–1564 vol. 3, Nov. 2004.

[Hil01] A. Hills. Large-Scale Wireless LAN Design. *IEEE Communications Magazine*, 39:98–107, 2001.

[HJ98] M. Handley and V. Jacobson. SDP: Session Description Protocol (RFC - 2327), 1998.

[HR86] D. Hong and S. S. Rappaport. Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Nonprioritized Handoff Procedures. *IEEE Transactions on Vehicular Technology*, 35:77–92, Aug. 1986.

[HT04] D. P. Hole and F. A. Tobagi. Capacity of an IEEE 802.11b Wireless LAN Supporting VoIP. In *Proc. of IEEE ICC'04*, volume 1, pages 196–201, Jun. 2004.

[IEE97] IEEE. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE Standard 802.11-1997, IEEE, Nov. 1997.

[jai08] JSR-000032 JAIN(TM) SIP Specification. http://jcp.org/aboutJava/communityprocess/final/jsr032/, Feb. 2008.

[JL04] A. Johnston and O. Levin. Session Initiation Protocol Call Control - Conferencing for User Agents, Jul. 2004.

[KTZ04] P. Khadivi, T. D. Todd, and D. Zhao. Handoff Trigger Nodes for Hybrid IEEE 802.11 WLAN/Cellular Networks. In *Proc. of QSHINE'04*, pages 164–170, Oct. 2004.

[Kuc91] A. D. Kucar. Mobile Radio: an Overview. *IEEE Communications Magazine*, 29:72–85, 1991.

[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.

[LSR01] J. Lennox, H. Schulzrinne, and J. Rosenberg. Common Gateway Interface for SIP (RFC - 3050), 2001.

[LWS04] J. Lennox, X. Wu, and H. Schulzrinne. Call Processing Language (CPL): A Language for User Control of Internet Telephony Services (RFC - 3880), Oct. 2004.

[Mah03] R. Mahy. A Call Control and Multi-Party Usage Framework for The Session Initiation Protocol (SIP) (Internet Draft), Oct. 2003.

[Mal07] K. El Malki. Low Latency Handoff in Mobile IPv4 (RFC - 4881), Jun. 2007.

[MBD04] R. Mahy, B. Biggs, and R. Dean. The Session Initiation Protocol (SIP) "Replaces" Header (RFC - 3891), Sept. 2004.

[MSA04] A. Mishra, M. Shin, and W. A. Arbaush. Context Caching Using Neighbor Graphs for Fast Handoffs in a Wireless Network. In *Proc. of INFOCOM'04*, volume 1, page 361, Mar. 2004.

[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 Journal on Personal Communications*, 7:34–47, 2000.

[Pra00] N. R. Prasad. IEEE 802.11 System Design. In *Proc. of IEEE ICPWC'00*, pages 490–494, Dec. 2000.

[Rap95] T. S. Rappaport. *Wireless Communications: Principles and Practice.* Prentice-Hall, 1995.

[Rap96] T.S. Rappaport. *Wireless Communications : Principles and Practice.* Prentice Hall, 1996.

[RLS99] J. Rosenberg, J. Lennox, and H. Schulzrinne. Programming Internet Telephony Services. *IEEE Network Magazine*, 13:42–49, 1999.

[Ros06] J. Rosenberg. Framework for Conferencing with the Session Initiation Protocol (RFC - 4353), Feb. 2006.

[RPSC04] J. Rosenberg, J. Peterson, H. Schulzrinne, and G. Camarillo. Best Current Practices for Third Party Call Control (3PCC) in the Session Initiation Protocol (SIP) (RFC - 3725), Apr. 2004.

[RSC+02] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. SIP: Session Initiation Protocol (RFC - 3261), Jun. 2002.

[RSL06] J. Rosenberg, H. Schulzrinne, and O. Levin. A Session Initiation Protocol (SIP) Event Package for Conference State (RFC - 4575), Aug. 2006.

[SAT06] M. N. Smadi, V. Azhari, and T. D. Todd. A Measurement-Based Study of WLAN to Cellular Handover. In *Proc. of IEEE MASS'06*, Oct. 2006.

[SCFJ03] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: A Transport Protocol for Real-Time Applications (RFC - 3550), Jul. 2003.

[ser08] SIP Express Router. http://www.iptel.org/ser/, Feb. 2008.

[SJ01] H. Sinnreich and A. B. Johnston. *Internet Communications Using SIP: Delivering VoIP and Multimedia Services with Session Initiation Protocol.* John Wiley & Sons, 2001.

[SJP07] R. Sparks, A. Johnston, and D. Petrie. Session Initiation Protocol Call Control - Transfer (Internet Draft), Nov. 2007. Expires Apr. 2005.

[SK98] M. Stemm and R. Katz. Vertical Handoffs in Wireless Overlay Networks. *Mobile Networks and Applications 3*, pages 335–350, 1998.

[SKBMM06] M. Souryal, L. Klein-Berndt, L. Miller, and N. Moayeri. Link Assessment in an Indoor 802.11 Network. In *Proc. of IEEE WCNC'06*, volume 3, pages 1402–1407, Apr. 2006.

[SKK01] R. Shirdokar, J. Kabara, and P. Krishnamurthy. A QoS-Based Indoor Wireless Data Network Design for VoIP Applications. In *Proc. of IEEE VTC'01*, volume 4, pages 2594–2598, Oct. 2001.

[SMA04] M Shin, A. Mishra, and W. Arbaugh. Improving the Latency of 802.11 Hand-Offs Using Neighbor Graphs. In *Proc. of ACM MobiSys '04*, pages 70–83, New York, NY, USA, 2004. ACM Press.

[Spa03] R. Sparks. The Session Initiation Protocol (SIP) Refer Method (RFC - 3515), 2003.

[Spa04] R. Sparks. The Session Initiation Protocol (SIP) Referred-By Mechanism (RFC - 3892), Sept. 2004.

[SR99] H. Schulzrinne and J. Rosenberg. The IETF Internet Telephony Architecture and Protocols. *IEEE Network Magazine*, 13:18–23, 1999.

[SR00] H. Schulzrinne and J. Rosenberg. The Session Initiation Protocol: Internet-Centric Signaling. *IEEE Communications Magazine*, 38:134–141, 2000.

[STK+] M. N. Smadi, T. D. Todd, V. Keys, S. V. Azhari, and D. Zhao. Dynamically Anchored Conferencing Handoff for Dual-Mode Cellular/WLAN Handsets. *To appear in ACM Transactions on Wireless Networks.*

[STZK06] M. Smadi, T. D. Todd, Dongmei Zhao, and Vytas Kezys. Dynamically Anchored Conferencing Handoff for Dual-Mode Cellular/WLAN Handsets. In *Proc. of IEEE ICC'06*, volume 5, pages 2028–2033, 2006.

[SW00] H. Schulzrinne and E. Wedlund. Application-Layer Mobility Using SIP. *ACM SIGMOBILE Mobile Computing and Communications Review*, 4(3):47–57, 2000.

[Sys05] Cisco Systems. Cisco 7920 Wireless IP Phone Design and Deployment Guide. Online, Oct. 2005.

[UMA08] Unlicensed Mobile Access. http://www.umatechnology.org/, Feb. 2008.

[Uni] International Telecommunication Union. Recommendation G.114, 1996. One-Way Transmission Time.

[voi08] VoiceXML. http://voicexml.org, Feb. 2008.

[YY95] T. P. Yum and K. L. Yeung. Blocking and Handoff Performance Analysis of Directed Retry in Cellular Mobile Systems. *IEEE Transactions on Vehicular Technology*, 44:645–650, 1995.

[Zha02] Yuan Zhang. SIP-based VoIP Network and its Interworking with the PSTN. *Electronics and Communication Engineering Journal*, 14:273–282, 2002.

[ZM04] F. Zhu and J. McNair. Optimizations for Vertical Handoff Decision Algorithms. In *Proc. of IEEE WCNC'04*, volume 2, pages 867–872 Vol.2, Mar. 2004.

[ZS05] J. Zhang and I. Stojmenovic. *Handbook on Security*. John Wiley & Sons, Dec. 2005.

[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 *Proc. of ICAP'03*, volume 1, pages 63–66 vol.1, Mar. 2003.