Scanning Delays in 802.11 Networks

David Murray
Murdoch University
D.Murray@murdoch.edu.au

Michael Dixon
Murdoch University
M.Dixon@murdoch.edu.au

Terry Koziniec
Murdoch University
T.Koziniec@murdoch.edu.au

Abstract

The proliferation of Wireless LANs and the increasing integration of voice into data networks has created the potential for VoWLANs (Voice over Wireless Local Area Networks). This technology has immense cost saving potential and the ability to provide better service and functionality. However, before the integration of VoWLAN is possible, handoff delays must be reduced. Currently, the connectivity transition that occurs from moving between APs (Access Points) is too long, causing poor voice quality and call dropouts. An experimental approach is used to investigate a particular handoff delay known as the scanning delay. The study concludes that the primary source of delay in the scanning process is caused by overlapping channels in the 2.4GHz band.

1. Introduction

In recent times, wireless LAN and VoIP (Voice over IP) technologies have experienced rapid growth. Wireless LANs have matured to the stage where they are being utilized in large scale, city-wide deployments. Equally, VoIP (Voice over IP) is being embraced in a number of markets. Large organizations have begun using VoIP for corporate communications and individuals are using products such as Skype for free global telephony.

VoWLAN (Voice over WLANs) is the combination of VoIP and WLANs. It aims to provide cell phone like service at the cost of a VoIP call. This technology has generated considerable interest with cell phone manufacturers now incorporating WiFi chipsets and SIP software into phones. However, before this technology can become widespread, several technical challenges must be solved. One of these challenges is handoff.

When a wireless client moves between two APs (Access Points), it must handoff to maintain network connectivity. The cause of handoff is shown in Fig 1.

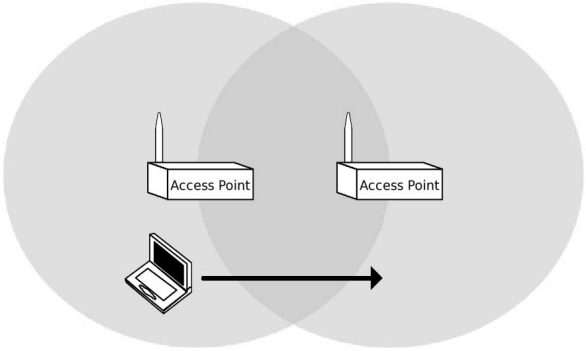

Fig 1. Handoff in 802.11 networks

During the handoff process, wireless clients are unable to send or receive data transmissions resulting in call dropouts. The ITU (International Telecommunication Union) specify that this delay should be less than 50ms [1]. However, prior studies have shown that the handoff delay far exceeds this target [2-4]. The excessive length of handoff is exacerbated by the frequency with which these handovers are performed. Due to the limited cell size of 802.11 APs, highly mobile users will be required to handoff many times over the duration of a call.

This paper is organized as follows. In section 2, layer 2 and layer 3 handoffs are differentiated, prior work is reviewed and the research scope is refined. Section 3 provides an in-depth discussion of scanning in wireless networks. Section 4 calculates the theoretical length of the scanning delay which is compared with prior experimental studies. Section 5 explains our experimental design and section 6 shows the results of our study. Our findings and the implications of our study are discussed in section 7 and the paper is concluded in section 8.

2. Handoff Research

Many different types of handoff exist. This section describes the difference between layer 2 and layer 3 handoff and reviews the proposed improvements. Layer 2 handoff is the most common type of handoff.

The 2007 International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST 2007)
0-7695-2878-3/07 $25.00 © 2007 IEEE

It occurs when a wireless client moves between APs within the same IP subnet. A Layer 3 handoff occurs when a wireless client moves between two APs in different subnets introducing IP addressing issues.

2.1. Layer 2 handoff

Layer 2 handoffs require clients to perform three phases: scanning, authentication and association. In the scanning phase clients search for and select a new AP. In the authentication phase clients identify themselves to the new AP and in the association phase the clients concurrent sessions are transferred from the old to the new AP.

The scanning process has been measured in a number of studies [2-4]. These results suggest that scanning delays vary between 70ms and 600ms. The authentication phase can be equally time consuming varying between a few milliseconds [2] and over one second [6] depending on the type of authentication. The association phase is a vendor specific operation. Although it is relatively short, approximately 15ms [2], delays are dependent on the number of sessions transferred between APs.

A number of different groups are working on mechanisms to reduce these delays. The 802.11 TGr is working towards a fast roaming amendment, 802.11r. This group is attempting to reduce authentication delays by proactively sending client information to adjacent APs prior to handoff. This enables wireless clients to pre-authenticate. Some research [6] claims that these mechanisms can reduce authentication delays from 1100ms to 50ms. Similar solutions to the association delay were proposed in the now expired 802.11F recommended practice specification [7]. Work has shown that proactive association can reduce delays from 15ms to 1.5ms [8].

2.2. Layer 3 handoff

Another major type of handoff is layer 3 handoff. This is performed when wireless clients move between APs in different subnetworks requiring clients to change network or IP address. The IETF's Mobile IP specification provides transparency to these IP address changes. A number of extensions [9-10] provide mechanisms to reduce the delays introduced by Mobile IP and Mobile IPv6.

3. Scanning

Despite the lengthy delays imposed by scanning, no mutually agreed solution exists. This section describes scanning in detail. Scanning is the process of searching for the best AP. Typically, SNR (Signal to Noise Ratio) is used to trigger the need to scan and to determine the best AP. However, this research investigates delays as opposed to heuristics.

In large wireless networks where many APs are needed to cover an area, APs are placed on different channels to avoid interference. When a wireless client moves away from its current AP, the SNR drops and the client is forced to scan for other suitable APs. During this period, the wireless client is unable to send or receive data. To find new APs, wireless clients must switch between and probe each channel individually. The number of channels or frequencies that must be scanned will depend on the country and the physical layer mode.

The 2.4GHz spectrum, used by 802.11b and 802.11g, has 11 or 14 channels depending on the country. The 5GHz spectrum, used by 802.11a, originally had 8 channels but recent amendments [11] have opened the spectrum to allow up to 24 channels. A crucial difference between channel allocation in the 2.4GHz and 5GHz spectrum is the amount of channel spacing. Every channel in the 5GHz spectrum is an independent, non-overlapping channel. Conversely, in the 2.4GHz spectrum only three channels namely 1, 6 and 11 are non-overlapping.

Scanning these channels can be done passively or actively. Passive scanning requires clients to wait for periodically broadcasted beacons. By default APs transmit beacon every 100ms. Subsequently, passively scanning clients must wait at least 100ms on every channel to ensure that all beacons are collected.

Channels can also be scanned actively which requires clients to proactively probe for APs. Active scanning is the fastest and most commonly implemented mechanism. The active scanning process begins with the client switching to a new channel and transmitting a probe. Following the probe, the client starts a timer. If no transmissions are heard by a time called the minimum channel time, the channel is declared empty and the wireless client restarts the process on a new channel.

However, if 802.11 traffic is heard, the client concludes that one or more APs may exist on the current channel. The client will subsequently remain on the channel until the expiry of the maximum channel time.

4. Real and Theoretical Active Scanning Delays

This section calculates the theoretical duration of scanning delays. The results are compared with the delays measured in prior experimental work [2-4]. The

theoretical network which we use to calculate delays contains two APs on different channels in the 2.4GHz spectrum. Our theoretical 802.11b/g clients must scan all 11 channels in the 2.4GHz spectrum. This estimation is divided into 3 components. These are: the amount of time required to switch channel and transmit a probe, the minimum channel time and the maximum channel time.

The wireless client in our theoretical network must switch between and transmit a probe on all 11 channels. Estimating channel switch times is difficult. Hardware channel switches in 802.11 equipment is between 40µs and 200µs [12]. Actual delays may be higher as a result of driver code paths. Our experimentation on Orinoco, Atheros and Intersil wireless chipsets found switching delays of 3ms, 5ms and 20ms respectively. Although these results were synonymous with other work [13], it is likely that the results were affected by our testing environment. A conservative estimate is that channel switching takes between 1ms and 5ms. The transmission of a probe takes 2ms. This is calculated from the MAC contention time and the time required to transmit a probe at 1 Mb/s. Subsequently, the total time required to switch to a new channel and transmit a probe will take between 3ms (1ms + 2ms) and 7 ms (5ms + 2ms) as shown in Fig 2.

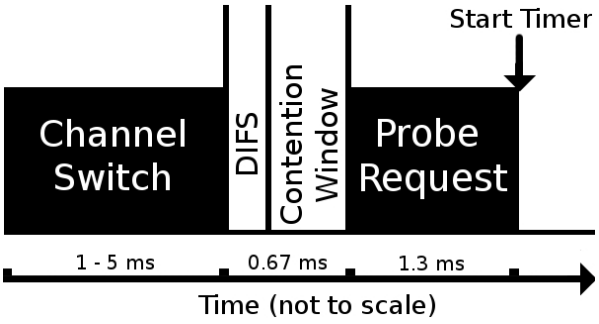

Fig 2. Channel switch and probe transmission

Following the transmission of a probe, the wireless client will start a timer and wait for a response. As stated in section 3, if no response is heard after the minimum channel time, the wireless client will switch to the next channel. However, if a response is heard, the client must wait on the channel for a longer period of time called the maximum channel time to ensure that replies from multiple APs can be received. In our theoretical network two APs operate on different channels. Subsequently, on 9 of the 11 channels, traffic will not be detected and the wireless client will switch channel after the minimum channel time.

The minimum channel time can be set relatively short [3]. The time required to assess whether a channel is occupied is the same as the contention time, 0.67ms. If the clients probe was heard, the AP will immediately transmit a response. However, if the AP or another station has priority over the medium and begins transmitting, the wireless traffic indicates that an AP is present but may be busy. In either case, the wireless client can quickly assess whether a channel is occupied. Subsequently, the minimum channel time can be set between 1ms and 2ms [3].

The maximum channel time must be considerably longer than the minimum channel time. As access to the wireless medium is randomly shared, the APs probe response time is dependent on AP load and the number stations competing for the medium. Subsequently, maximum channel timers must be set high enough to ensure that even heavily loaded APs can respond. Simulations have revealed that maximum channel times in 802.11b networks can be set between 10ms [3] and 27ms [14].

To finalize our estimates, empty channels will incur the channel switch and probe delay followed by the minimum channel time. This should take at least 4ms (3ms + 1ms) and at most 9ms (7ms + 2ms). Scanning busy channels requires the same process but clients must wait for the expiry of the maximum channel time. Channels with APs will take at least 13ms (3ms + 10ms) and at most 34ms (7ms + 27ms) to scan. With 9 empty channels and 2 busy channels, scanning delay should be at least 62ms ((4ms × 9chans) + (13ms × 2chans)) and at most 149ms ((9ms × 9chans) + (34ms × 2chans)) depending on timers and switching delays.

Considering that prior experimental work has measured scanning delays between 70ms and 600ms [2-4], these estimates lead to uncertainty over the source of delays. Are high scanning delays a product of relaxed timers? Some studies [2], [4] observed a high degree of variation in delays using the same test-bed. How can a function based on timers show large variations in delay? This study uses an experimental approach to investigate scanning to account for delays.

5. Experimental Design

To analyze delays, an experiment was designed to capture the scanning process to enable empirical measurement and a frame-by-frame analysis. Prior studies [2-4] have investigated scanning in a similar manner however this experiment is unique as it also examines the 802.11a standard.

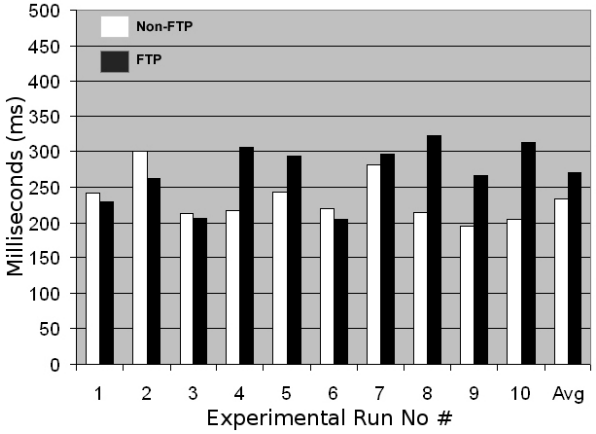

Fig 3. Client: Cisco .11b -- AP: Linksys

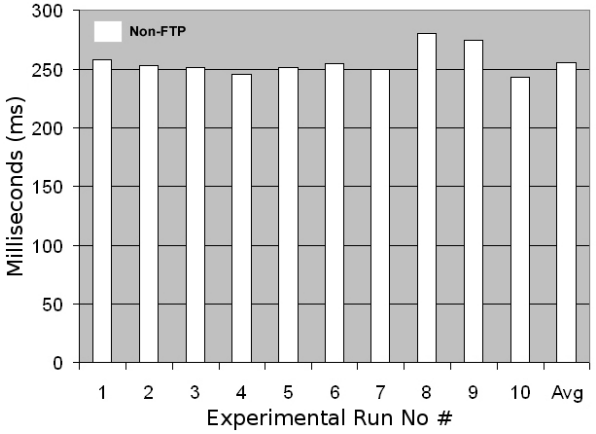

Fig 4. Client: Enterasys .11b – AP: Linksys

Ethereal network analyzer software was used to collect, store and timestamp packets. However, the 802.11 scanning process is significantly more difficult to capture than traditional wired Ethernet. Firstly, scanning frames are a management frame and are only able to be captured in a special promiscuous wireless mode known as RF (Radio Frequency) monitor mode. Wireless interfaces in RF monitor mode are unable to transmit frames and cannot operate as a client. Subsequently, unlike wired Ethernet, a separate interface is required to capture and store management frames. Secondly, as the scanning process occurs over multiple channels, multiple interfaces are required to capture the entire process. This study used four interfaces in two PCs to capture traffic. The clocks of these PCs were syncronised using NTP (Network Time Protocol). The packet capturing interfaces were Atheros 802.11a/b/g cards running the Linux MADWiFi driver.

A number of experimental variables were used.

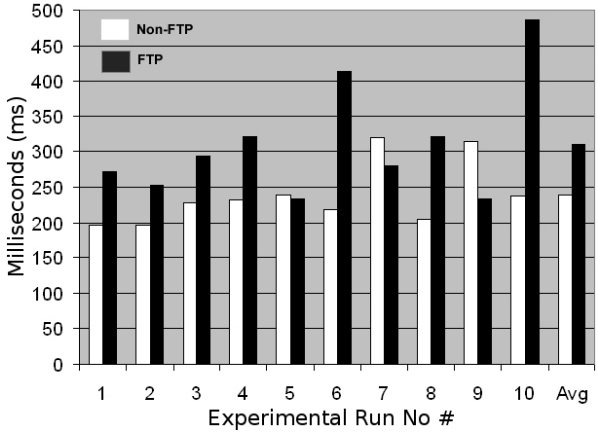

Fig 5. Client: Aironet .11b – AP: Cisco

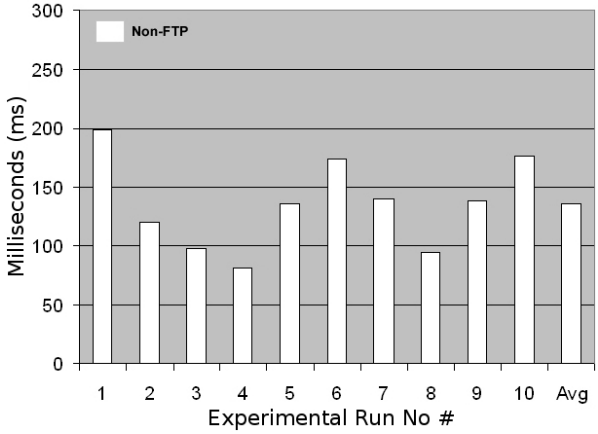

Fig 6. Client: Enterasys .11b – AP: Cisco

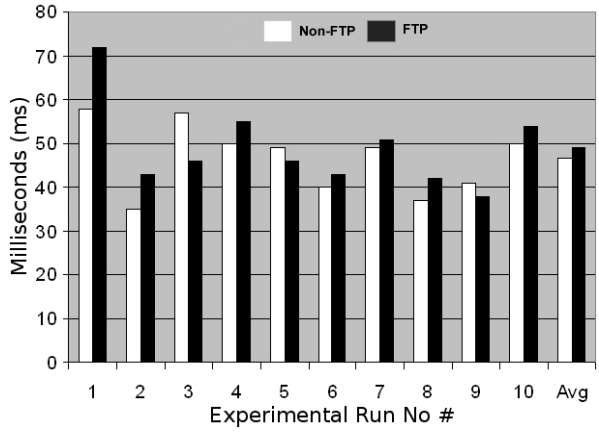

Fig 7. Client: Cisco .11a – AP: Cisco

Three Different client cards were tested including: Cisco Aironet 802.11b, Enterasys RoamAbout 802.11b and Cisco Aironet 802.11a wireless cards. Also, two different APs were tested: Cisco 1200s with 802.11a/b radios and Linksys WRTs running OpenWRT with 802.11b/g radios.

In addition to varying the client card and AP, we also investigated the impact of background or FTP traffic. By adding another client into the wireless network and starting a large FTP file transfer, the effect of heavy traffic loads on scanning delays could also be examined.

Handoff was induced by physically moving a laptop containing one of the five wireless cards between the APs. Each wireless card was tested 10 times and averaged. The purpose of this experiment was not to review vendors scanning algorithms, but to reveal why actual delays are higher than our theoretical estimates.

6. Results

The results of the tests are shown in Fig 3, 4, 5, 6 and 7. The Cisco 802.11b wireless card had an average scanning delay of ~250ms when scanning the Linksys APs (Fig 3) and ~275ms when scanning the Cisco AP (Fig 5). The black and white bars indicate scanning with and without background FTP traffic. Delays are slightly higher when scanning a channel with FTP traffic.

The Enterasys wireless card displayed highly variable scanning times between ~75ms and ~275ms. Fig 4 and Fig 6 confirm observations in prior work [3] showing that scan times vary with the AP. The Enterasys card was not capable of roaming to a channel saturated with FTP traffic possibly due to an additional roam criterion or heuristic. Subsequently, results have not been recorded

The results of the Cisco 802.11a wireless card scanning the Cisco APs are shown in Fig 7. As the Linksys AP does not support 802.11a no results are recorded. Few prior studies have investigated scanning in 5GHz networks. One study [4] measured scan times of an 802.11a card but did not specify their method for capturing 802.11a packets or the wireless client they used. They found that scanning delays were between 900ms and 1000ms. Our results suggest the opposite. Fig 7 shows the Cisco 802.11a card scanning the 8 5GHz channels in an average of 46ms. This is approximately five times faster than the Cisco 802.11b wireless card.

7. Discussion

A pivotal question which stems from the results is: why are scan times approximately five times lower using the Cisco 802.11a card than the Cisco 802.11b card? A number of factors may contribute to lower scan times in the 802.11a wireless card. Firstly, the Cisco 802.11a wireless card only scanned eight channels whereas the 802.11b wireless cards scanned 11 channels. Secondly, 802.11b/g wireless cards transmit management frames at 1 Mb/s and 802.11a wireless cards transmit management frames at 6 Mb/s. Furthermore, the 802.11a standard has lower contention times. High data rates and low contention times allow frames to be serialized onto the medium faster. As less time is required to transmit and receive frames, it is permissible that maximum channel times could be set lower in 802.11a cards reducing scanning delays.

While these factors may play a role, frame-by-frame analysis of the packet captures reveals that channel overlap is the primary reason for low scanning delays in 802.11a wireless cards.

The division of channels in the 5GHz spectrum is non-overlapping. The 2.4GHz spectrum consists of 11 channels, of which only 3 are non-overlapping. 802.11b packet captures show 802.11b APs responding to many probe requests for each clients scan. Comparatively, the 802.11a APs only respond to one probe request per probe request.

The transmission of superfluous probe responses in 802.11b APs is a result of responses to probes on overlapping channels. This has a significant effect on total scan times. As discussed in our theoretical estimates, channels whereby traffic is detected require wireless clients to wait for their maximum channel time to expire. The reason that 802.11b scan times are higher than our theoretical estimates is because clients are waiting the duration of their maximum channel time on channels that overlap with the APs designated channel.

This phenomenon is difficult to prove as management frames do not specify the channel on which they were transmitted. This makes it difficult to confirm that, for example, a probe transmitted on channel 2 is being replied to by an AP on channel 1. However, channel information is provided by the packet capturing wireless interface in RF monitor mode. This information specifies the channel that the packet was captured on. Through experimentation, we noticed that a packet capturer on channel 3 could capture AP transmissions on channel 1.This is demonstrative that packets sent on one wireless channel, can be processed and interpreted by interfaces on different channels.

The overlapping channels theory also explains how a process based on timers can display large variations seen in this and previous studies [2], [4]. As a result of physically where and when a client scans, an AP may or may not receive probe requests on any number of overlapping channels. Consequently, a wireless card, may, or may not, wait for the duration of the maximum channel timer on a given channel. High transmit

powers and close proximity to APs can all exacerbate the degree of channel overlap.

It also explains how different APs can affect the scanning delay; an operation driven by client timers. The varying physical layer IF (Intermediate Frequency) filtering characteristics and sensitivities of APs will result in different abilities to detect and respond to probe requests on overlapping channels

This novel concept is unique to the scanning phase as connecting to an AP mitigates the problem. Once connected, frames are filtered using the BSSID (Basic Service Set IDentifier) or the MAC address of the AP.

Some studies [3], [15] have suggested that scanning delays can be reduced by sending lists to clients with the channels upon which nearby APs reside. Pruning unused channels could dramatically reduce scan times and mitigate the overlapping channel problem. However, such mechanisms require each AP to individually learn of its physically adjacent neighbors.

A simple way to reduce scanning delays is to send channel information in the header of a probe request frame. APs could therefore ignore frames transmitted on different channels. This may not entirely mitigate the problem as clients may still detect traffic when scanning overlapping channels and wait the duration of their maximum channel time, it could alleviate the problem in many circumstances.

8. Conclusion

This study measured scanning delays with different wireless cards, APs, and levels of background traffic. It was found that 802.11b scanning delays were significantly higher than theoretical calculations. Further investigation revealed that the primary source of delay is caused by an inability for wireless clients and APs to distinguish between management frames transmitted on overlapping channels. Currently, no layer two mechanisms prevent probe requests from being received by APs on overlapping channels. We propose sending channel information in probe requests as a simple yet effective mechanism to reduce scanning delays.

9. References

[1] CCITT, "General Characteristics of International Telephone Connections and International Telephone Circuits", ITU (International Telephone Union), 1988.

[2] A. Mishra, M. Shin, and W. Arbaugh, "An Empirical Analysis of the IEEE 802.11 MAC Layer Handover Process", *ACM SIGCOMM Computer Communications Review*, April 2003, pp. 93-102.

[3] H. Velayos, and G. Karlsson, "Techniques to reduce IEEE 802.11b MAC Layer Handover Time", *ICC IEEE Conference on Communications*, June 2004, vol 7, pp. 3844-3848.

[4] S. Shin, A.G. Forte, A.S. Rawat, and H. Schulzrinne, "Reducing MAC Layer Handoff Latency in IEEE 802.11 Wireless LANs", *In Proceedings of the 2nd International Workshop on Mobility Management & Wireless Access Protocols*, New York, 2004, pp. 19-26.

[5] J-O. Vatn, "An Experimental Study of IEEE 802.11b Handover Performance and its Effect on Voice Traffic", *Technical Report*, KTH Royal Institute of Technology, July 2003.

[6] A. Mishra, M. Shin, and W. Arbaugh, "Proactive Key Distribution using Neighbor Graphs", *IEEE Wireless Communications*, February 2004 Vol 11:1, pp. 26-36.

[7] IEEE, "802.11F - Recommended Practice for Multi-Vendor Access Point Interoperability via an Inter-Access Point Protocol Across Distribution Systems Supporting IEEE 802.11 Operation", 2003

[8] A. Mishra, M. Shin, and W. Arbaugh, "Context Caching using Neighbor Graphs for Fast Handoff in a Wireless Network", *IEEE INFOCOM Conference on Computer Communications*, March 2004, Vol 1, pp. 351-361.

[9] IETF, "Low Latency Handoffs in Mobile IPv4", *IETF Draft*, 2005.

[10] IETF, "Mobility Support in IPv6", *RFC 3775*, 2006.

[11] IEEE, "802.11h - Spectrum and Transmit Power Management Extensions in the 5GHz band in Europe", 2003

[12] P. Bahl, R. Chandra, and J. Dunagan, "SSCH: Slotted Seeded Channel Hopping for Capacity Improvement in IEEE 802.11Ad-hoc Wireless Networks", *Proceedings of the 10th Annual International Conference on Mobile Computing and Networking*, New York, September 2004, pp 216-230.

[13] I. Ramani, and S. Savage, "Syncscan: Practical Fast Handoff for 802.11 Infrastructure Networks", *IEEE INFOCOM The Conference on Computer Communications*, March 2005, Vol 1, pp 675-684.

[14] R. Pries, and K. Heck, "Simulative Study of the WLAN Handover Performance", *OPNETWORK 2005*, Washington D.C., August 2005.

[15] M. Shin, A. Mishra, and W. Arbaugh , "Improving the Latency of 802.11 Hand-offs using Neighbor Graphs", *In Proceedings of the 2nd International Conference on Mobile Systems, Applications, and Services*, Boston, June 2004.