Malware in IEEE 802.11 Wireless Networks

Brett Stone-Gross[1], Christo Wilson[1], Kevin Almeroth[1], Elizabeth Belding[1], Heather Zheng[1], and Konstantina Papagiannaki[2]

[1] Department of Computer Science,
University of California, Santa Barbara
{bstone,bowlin,almeroth,ebelding,htzheng}@cs.ucsb.edu
[2] Intel Research
Pittsburgh, PA
dina.papagiannaki@intel.com

Abstract. Malicious software (malware) is one of the largest threats facing the Internet today. In recent years, malware has proliferated into wireless LANs as these networks have grown in popularity and prevalence. Yet the actual effects of malware-related network traffic in open wireless networks has never been examined. In this paper, we provide the first study to quantify the characteristics of malware on wireless LANs. We use data collected from the large wireless LAN deployment at the 67^{th} IETF meeting in San Diego, California as a case study. The measurements in this paper demonstrate that even a single infected host can have a dramatic impact on the performance of a wireless network.

1 Introduction

There has been ample research on the separate topics of malware and wireless networks. A majority of malware research has focused on propagation modeling, detection, and application characterization [3][5][8]. The impact of malware induced traffic on the performance of wired networks has been largely ignored, because the effects of additional ingress and egress flows are mitigated by faster access technologies and more bandwidth. However, limited resources in wireless networks and the inherently broadcast nature of the medium creates valid concerns when considering network performance. This work analyzes these effects which include MAC layer retransmissions, management frame collisions, and an overall performance degradation due to increased congestion.

Wireless networks have been examined through experimental measurements and simulations. Many studies have assessed wireless performance on deployed networks [1][9][10][13]. Rodrig *et al.* captured wireless traffic and analyzed the efficiency of the 802.11 protocol [12]. They present how the efficiency significantly degrades during periods of high contention with the majority of packets requiring link layer retransmissions due to packet loss and transmission errors. These results are consistent with our own findings. Jardosh *et al.* examined methods for detecting congestion in large-scale wireless networks [7]. They propose that monitoring the channel busy time is a good measure of channel utilization. In

M. Claypool and S. Uhlig (Eds.): PAM 2008, LNCS 4979, pp. 222–231, 2008.

addition, network throughput and goodput can be used as metrics to identify congestion. Heusse *et al.* [6] found that anomalies in current multi-rate adaption algorithms of 802.11 cause an overall reduction in network performance, especially during periods of congestion. We also observed this behavior during several malware attacks. What all of these studies lack is an accounting of the extraneous packets that are injected into the network by malicious software.

We are the first to quantify, characterize, and correlate the effects of malicious network traffic on wireless performance. We believe that analyzing the effects malware can have on wireless networks is important. The applications of our research can lead to more realistic traffic models, justify the need for network protection, and improve the quality of service in wireless networks. In addition, recognizing these effects are beneficial in wireless network diagnostics [2][4].

The remainder of this paper is organized as follows. Section 2 describes our data collection and filtering process. In Section 3, the data sets are summarized. The effects that malware produced in the wireless network are examined in Section 4. Section 5 concludes with an overall summary of our findings.

2 Data Collection and Filtering

The wireless network deployed at the 67^{th} IETF meeting was unusual due to both its large size and heavy utilization. The network provided an excellent opportunity to analyze the characteristics and prevalence of malware. With more than 1,700 unique users on the network, the resulting trace provided the equivalent of a small Internet Service Provider's (ISPs) perspective of malware attacks. Details of our data collection process at the IETF meeting and our subsequent malware identification process are discussed in this section.

2.1 Experimental Setup

The on-site network at the IETF meeting consisted of 30 802.11 a/b/g access points routed to a 44.7Mbps T3 backhaul link to the Internet. Participants utilized the Dynamic Host Configuration Protocol (DHCP) to obtain a publicly routable IP address in the 130.129/16 address range. No MAC layer encryption, Network Address Translation (NAT) devices, or firewalls were present in between the access points and the backhaul connection.

We collected data from two vantage points:

1. *Trunk Data Set*: Full data traces were recorded from a trunk mirror port on the router which managed the backhaul Internet link.
2. *Wireless Data Set*: Wireless sniffers were strategically positioned around the meeting near popular access points to record wireless traffic, as shown in Figure 1. Each wireless sniffer consisted of an IBM or Toshiba laptop with an Atheros chipset. Each sniffer was configured in RFMon mode to capture all management and data frames. Based on previous measurements [7], we estimate that each sniffer recorded more than 90% of frame transmissions.

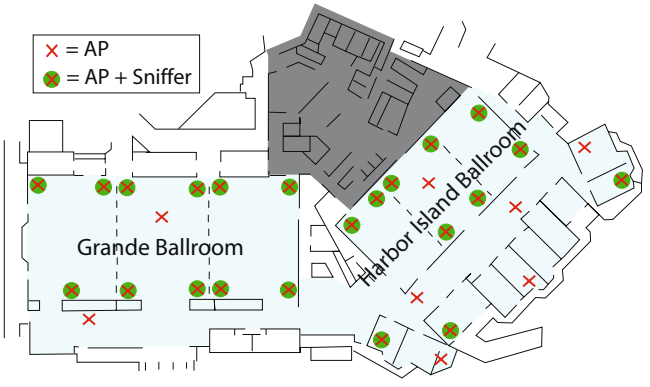


Fig. 1. Locations of wireless APs and data collection sniffers at the IETF meeting

Over 511 gigabytes of uncompressed data were collected at the trunk port along with another 131 gigabytes of uncompressed data recorded by the wireless sniffers. The data collected from the trunk port included some packets destined for a small on-site terminal room. This location was the only place in which attendees could access a wired Ethernet connection. We were able to identify traffic from the terminal room from the fixed set of IP addresses assigned by DHCP, by comparing IP addresses in both traces, and confirmed that less than 10% of the traffic observed in the trunk data set came from the terminal room.

2.2 Filtering Heuristics

In order to isolate malicious traffic from the normal flows present in the data set, we created a set of heuristic-based filters to detect abnormal behavior. We designed the filters around a set of assumptions about known malware behavior patterns, and then constructed an identification and measurement system. We observed that malware's traffic exhibits two primary types of traffic patterns:

- *Scanning behavior*: Worms and Trojans are typically spread by scanning large sequences of IP addresses on known ports. The scans search for vulnerable or weakly protected services (*e.g.*, default, weak or non-existent passwords) that can be exploited.
- *Flooding behavior*: Malware is often directed to attack other computers by flooding them with connection attempts (*e.g.*, a SYN flood).

One of the key characteristics of scanning behavior is that the machine in question will contact an abnormally large number of different IP addresses. This behavior will occur repeatedly on known vulnerable ports. Flooding behavior is best characterized as one machine initiating an unusually large number of connection attempts to one particular IP address.

For both behavior patterns, malicious traffic flows are often unidirectional and almost always short-lived. In the former pattern, scan attempts are often directed at unused IP addresses, or towards machines with firewalls which results in unidirectional traffic. SYN floods are by definition, unidirectional. If a scanner does manage to find a live target, it will attempt to either infect the host or guess the host's password, both of which are relatively brief affairs. Attempts may be repeated, but the connection is broken and reset each time, leading to bursty traffic flow characteristics. Another important consideration is that certain forms of malware including adware, keyloggers, and open relay proxies generate smaller amounts of network traffic and are consequently harder to identify. Therefore, the rest of our results should be considered as a lower bound of malware present.

3 Wireless and Trunk Data Analysis

Before we examined our wireless data set, we first developed a more general characterization of the network activity at the IETF using the trunk data set. Besides deriving network statistics, we used the trunk data set as the basis to identify malicious flows, which we later correlated with the more restricted data set obtained from the wireless sniffers.

3.1 Malicious Traffic Analysis

We begin by analyzing the malicious traffic present in the trunk data set. There were 109,740 unique external IP addresses in the trace, and 3,941 were implicated in malicious behavior, or about 3.6%. We identified 1,786 internal IP addresses, and out of this set 14 (0.8%) showed indications of malicious activity.

Overall, 272,480,816 egress TCP packets were sent over the course of the meeting, of which 4,076,412 (1.5%) were involved in malicious flows. 284,565,595 ingress TCP packets were received, of which 2,765,683 (1.0%) were malicious. In general these results appear consistent with a study by Kotz and Essien [9]. They recorded observing 0.9% of TCP traffic being sent to Microsoft RPC port 445, which they correlate with denial-of-service attacks against Windows 2000 machines. In our case, since we quantify scanning as well as flooding attacks across multiple services, our results represent a more complete view of overall malicious traffic percentages.

Although malicious TCP traffic accounted for an average of 1% of the total traffic at the IETF meeting, it accounts for a much larger percentage of TCP control traffic, defined as SYN and SYN-ACK packets. Thus, when data packets are not considered, the magnitude of malicious traffic becomes much more pronounced (as displayed in Figures 2 and 3). From this data, malicious flows are shown to account for a substantial portion of total TCP connection requests, occasionally rising above 50%. During a massive SSH password cracking attempt on Friday morning, nearly 100% of all TCP control traffic was part of the attack, and is clearly evident in Figures 2 and 3. In addition to conducting an analysis of malware behavior within the IETF network, we also attempted to isolate what

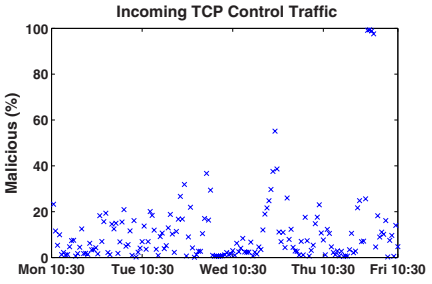


Fig. 2. Instantaneous percentage of incoming malicious TCP traffic

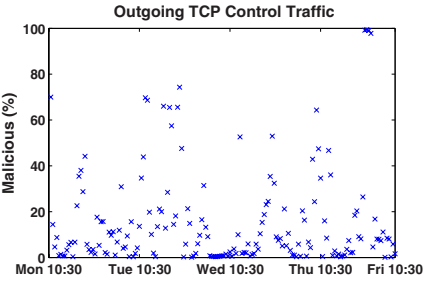


Fig. 3. Instantaneous percentage of outgoing malicious TCP traffic

effects such traffic had on the wireless medium itself. Although we were able to identify many attacks in the trunk data set, pinpointing these same attacks in the wireless data set proved to be difficult since our sniffers did not observe all wireless LAN traffic across all access points. From the set of malicious flows that were detectable in the wireless data sets, many proved unsuitable for analysis. The reasons include the following:

1. Ingress attacks that involved only a few total packets.
2. Egress scanning attacks which, though long lived, only generated a few packets per second.
3. Ingress port scans that were distributed over hosts on all 30 access points.
4. Backscatter from DoS attacks throughout the Internet that produced unsolicited TCP SYN ACKs, resets, and ICMP replies [11].

Although the preceding cases were not ideal for analyzing MAC characteristics, these attacks still had an overall effect as more than 1% of all packets were malicious and present in the wired and wireless data sets. The most substantial effects on wireless performance were produced by malicious flows that originated within the network. Therefore, we examined several of these egress flows under light and heavy channel utilization.

4 Quantifying the Impact of Malware

As previously discussed in Section 3.1, malicious egress flows were well suited for our analysis since these flows consumed more bandwidth, and caused more collisions than malicious ingress flows. In order to understand the impact of these malicious flows on the MAC layer, we aggregated statistics for channel utilization, throughput, probe requests/responses, data packets/retries/acknowledgments, and transmission rates. At the transport layer we computed the TCP Round-Trip-Times (RTT) to determine the end-to-end delay.

Table 1. The effects on TCP RTT of an ICMP flood and NetBIOS attack

	Non-Attack Interval	During Attack	Percent Increase
Avg (Egress)	64.7 ms	99.2 ms	53.23%
Avg (Ingress)	23.4 ms	36.1 ms	54.36%
Median (Egress)	41.6 ms	85.0 ms	104.33%
Median (Ingress)	3.2 ms	6.8 ms	112.50%

4.1 Malware Attacks in Wireless Networks

We performed a detailed analysis of two of the largest attacks occurring in the wireless data sets during the meeting based on packets per second and bandwidth. These types of attacks were also the most common that we observed. They included an ICMP ping flood combined with a NetBIOS exploit and a TCP SYN Flood.

ICMP Flood and NetBIOS Exploit. One of the largest network attacks observed during the entire meeting was an ICMP ping sweep across a range of IP addresses. The attack was used to probe for machines and prepare for a subsequent NetBIOS worm exploit. The malicious flow persisted for approximately 18 minutes and 7 seconds occurring late Thursday afternoon during the plenary session between 17:02:38 and 17:20:45. The attack created 79,289 packets at an average rate of 117 packets per second with a maximum burst of 235 packets per second. The impact of the flow drove the channel utilization to nearly 100%, and caused both a rise in the number of link layer data retries (retransmissions) and a reduction in the transmission rates (shown in Figure 4). The metric in Figure 4(b) shows the two primary ranges of transmission rates of 11-18Mbps and 48-54Mbps that were used by wireless clients. The rectangular regions in Figure 4 and 5 indicate the periods of malicious traffic flow.

As part of our analysis, we also discovered a brief period in the middle of the ping flood just after 17:09:00 when the attack halted. This temporary pause resulted in a reduction in utilization, an increase in data transmission rates, and fewer data retries. Unfortunately we were not able to determine why the attack was suspended during this two minute interval, but we conjecture that the infected machine may have become unresponsive and was rebooted.

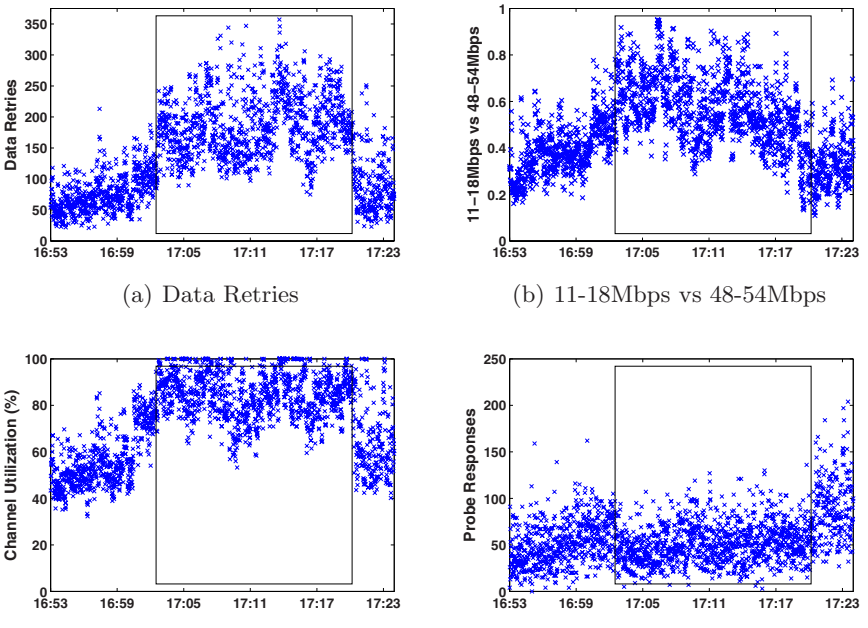
An additional result that we observed in our analysis was that overall, the combined throughput on the channel remained relatively constant at 4,412 KB/s over the course of the attack. However, the average and median RTT increased by more than 50% and 100% respectively for all TCP flows. Table 1 displays the average and median RTTs for a 10 minute interval before and after the attack with respect to the RTT during the attack.

There are several conclusions that can be drawn based on these results. First, the attacker was not only able to adversely affect other clients' performance, but also obstruct the access point's probe responses to clients who were searching for access points via probe requests. This is evident in Figure 4(d), which illustrates the spike in probe responses immediately after the attack occurred.

Consequently, the attack exacerbated a problem in the wireless network in that probe requests and responses were essentially jammed during heavy utilization. Access point control packets such as beacons, probes, and other management frames were also lost or delayed, and therefore served no productive purpose and only contributed to the overall network congestion.

A reduction in client transmission rates occurred due to the Auto Rate Fallback (ARF) mechanism, as illustrated in Figure 4(b), due to increased packet loss. As a result, packet transmission times increased, which further increased the channel busy time. The purpose of ARF is to combat lossy channel conditions by sending data at lower rates (*i.e.*, provide more robust modulation and coding schemes), and thus decrease the likelihood that data is lost because of radio noise. However, using the ARF strategy is a poor choice in this case since dropped packets are due to packet collisions and not noise interference. During these congested periods, this behavior created a negative feedback loop as client queues filled, but were unable to effectively drain due to contention compounded by slower transfer rates. Therefore, the delay for each host increased as they continuously waited for the channel to become idle.

The dramatic increase in TCP delay, as shown in Table 1, can be attributed to the additional strain that this attack placed on the link layer. Accordingly, the attack produced a large amount of data retransmissions. During the attack nearly 25% of all MAC layer frames were retransmissions, and at the peak of the attack



(a) Data Retries

(b) 11-18Mbps vs 48-54Mbps

(c) Channel Utilization

(d) Probe Responses

Fig. 4. ICMP Flood and NetBIOS exploit effects on the wireless medium

almost 50% of all packets were retransmissions. In addition, as clients following the ARF procedure reduced their transmission rates, the channel became even more congested as transmissions took longer to complete. These characteristics had a significant impact on TCP delay due to the fact that these MAC layer delays and losses were assumed to be caused by end-to-end congestion. Hence TCP transmission timeouts occurred, which reduced the congestion window.

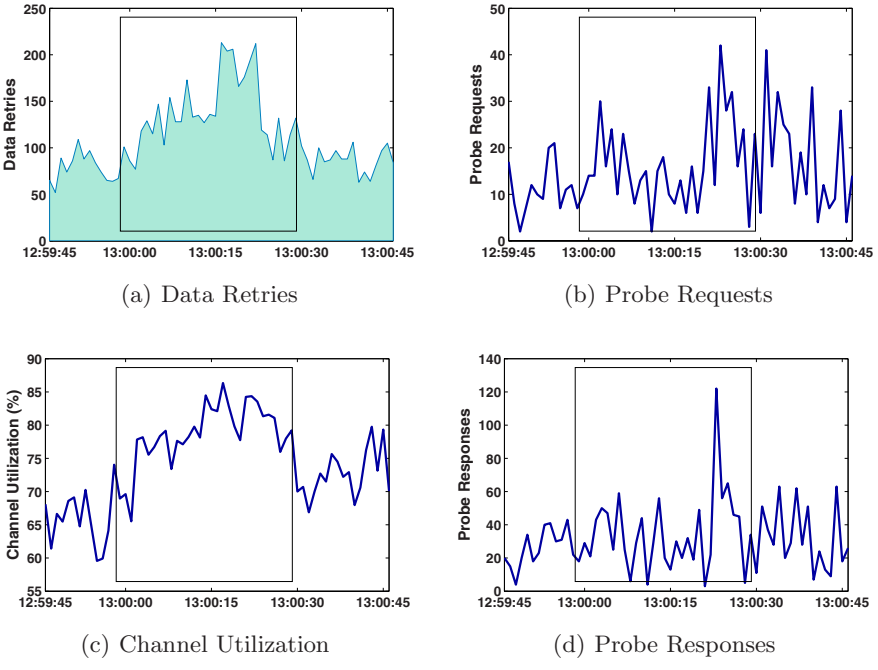
TCP SYN Flood. Another one of the more obvious attacks that we observed was a TCP SYN flood directed at an external server on Port 80 involving over 6,000 connection requests. The attacker in question emitted three bursts of attack traffic that began Thursday afternoon at 12:59:57 and numbered up to 109 packets per second for 30 seconds.

Figure 5 combines several of these measurement metrics during the initial attack, which lasted for only 30 seconds. The peaks in the numbers of data packets correspond to periods of attack. As shown in Figure 5(c), the aggregate channel utilization for this particular access point, while elevated, was not near bottleneck limits. What was most impacted by the SYN flood was the data retry rate, which peaked in the midst of the attack. This result indicated a higher rate of contention and collisions at the MAC layer due to the attacker's rapid transmission of single SYN packets. The result was an increase in the overall end-to-end latency as the MAC layer struggled to reliably deliver packets. During this attack, the average RTT increased by more than 33% with 16% of all frames consisting of MAC layer retransmissions. At the peak of the attack, more than 30% of all frames were data retransmissions.

Additionally, the aggregate number of probe requests and probe responses to and from all access points increased during the initial attack as illustrated in Figures 5(b) and 5(d). This result indicates that the attacker may have aggravated existing hidden terminal problems, thereby causing collisions and data retries. This behavior then triggered nearby clients that were connected to the same access point to begin probing for other access points offering better connectivity. While these effects do not appear catastrophic, it is evident that the probe responses and data retries increased by more than twice their averages over regular traffic intervals. Analogous to the ICMP ping flood, the number of probe responses more than doubled immediately after the attack. This behavior occurred in response to the outstanding probe requests that were partially blocked during the attack interval.

4.2 Effects of Malicious Flows on Wireless Performance

Our findings show that the presence of active malware in a congested wireless network harms performance by reducing client transmission rates and increasing data retries. The results also demonstrate that the end-to-end delay for TCP connections rise commensurately with slower data rates and greater numbers of packet collisions. These effects would likely have a significant impact on real-time applications. Under heavy utilization, access point management frames can be obstructed and increase the delay in client handoffs, authentications, and



(a) Data Retries (b) Probe Requests

(c) Channel Utilization (d) Probe Responses

Fig. 5. TCP SYN flood effects on the wireless medium

associations, further degrading performance. By comparing the effects of the NetBIOS attack with the TCP SYN flood, we can determine that faster sending rates and larger packets have a more significant effect on the wireless medium since the channel is busy for longer periods of time. In addition, the 802.11 CSMA protocol worked well in preventing small TCP SYN packets from dominating the channel during malicious traffic flows.

5 Conclusion

The study of malware on wireless systems is becoming increasingly important as more devices communicate openly over-the-air. In this paper, we analyzed the effects that malware-driven attacks can have on 802.11 performance. The most severe consequence is an increase in RTTs, which can hinder real-time communication. Wireless quality of service is also virtually impossible without developing mechanisms to reduce unwanted link layer contention.

The results that we present are from single attackers' outgoing malware attacks. Left unabated, the prevalence of malware will lead to a higher concentration of attackers and potentially deny service to legitimate users. This makes the protection of connected machines an especially pertinent objective for wireless network operators. In addition, as worms and botnets become more sophisticated, we believe that the exploitation of wireless networks by mining sensitive

information from unencrypted transmissions will become routine. Malware will also adapt to preserve its own anonymity by spoofing the source of attacks. Consequently, the effects of multiple compromised machines on a single wireless access point will become more significant as malware evolves to specifically exploit the wireless medium. Therefore, a lightweight solution will be essential to ensure optimal network performance and protect users' sensitive data.

References

1. Balachandran, A., Voelker, G.M., Bahl, P., Rangan, P.V.: Characterizing User Behavior and Network Performance in a Public Wireless LAN. In: Proc. of ACM SIGMETRICS, Marina Del Rey, CA, June 2002, pp. 195–205 (2002)
2. Chandra, R., Padmanabhan, V., Zhang, M.: WiFiProfiler: Cooperative Diagnosis in Wireless LANs. In: Proc. of MobiSys, Uppsala, Sweden (June 2006)
3. Chen, Z., Gao, L., Kwiat, K.: Modeling the Spread of Active Worms. In: Proc. of IEEE INFOCOM, San Francisco, CA (April 2003)
4. Cheng, Y., Afanasyev, M., Verkaik, P., Benko, P., Chiang, J., Snoeren, A., Savage, S., Voelker, G., Kwiat, K.: Automating Cross-Layer Diagnosis of Enterprise Wireless Networks. In: Proc. of ACM SIGCOMM, Kyoto, Japan (August 2007)
5. Gu, G., Porras, P., Yegneswaran, V., Fong, M., Lee, W.: BotHunter: Detecting Malware Infection Through IDS-Driven Dialog Correlation. In: Proc. of Usenix Security Symposium, Boston, MA (August 2007)
6. Heusse, M., Rousseau, F., Berger-Sabbatel, G., Duda, A.: Performance Anomaly of 802.11b. In: Proc. of IEEE INFOCOM, San Francisco, CA (March 2003)
7. Jardosh, A., Ramachandran, K., Almeroth, K., Belding-Royer, E.: Understanding Congestion in IEEE 802.11b wireless networks. In: Proc. of Internet Measurement Conference, Berkeley, CA (October 2005)
8. Kirda, E., Kruegel, C., Banks, G., Vigna, G., Kemmerer, R.: Behavior-based Spyware Detection. In: Proc. of Usenix Security Symposium, Vancouver, Canada (August 2006)
9. Kotz, D., Essien, K.: Analysis of a Campus-wide Wireless Network. In: Proc. of ACM MOBICOM, Atlanta, GA (September 2002)
10. Meng, X., Wong, S., Yuan, Y., Lu, S.: Characterizing Flows in Large Wireless Data Networks. In: Proc. of ACM MOBICOM, Philadelphia, PA (September 2004)
11. Moore, D., Voelker, G.M., Savage, S.: Inferring Internet Denial-of-Service Activity. In: Proc. of Usenix Security Symposium, Washington D.C (August 2001)
12. Rodrig, M., Reis, C., Mahajan, R., Wetherall, D., Zahorjan, J.: Measurement-based Characterization of 802.11 in a Hotspot Setting. In: Proc. of ACM SIGCOMM Workshop on Experimental Approaches to Wireless Network Design and Analysis (E-WIND), Philadelphia, PA (August 2005)
13. Schwab, D., Bunt, R.: Characterizing the Use of a Campus Wireless Network. In: Proc. of IEEE INFOCOM, Hong Kong, China (March 2004)