Analysis of an IEEE 802.11 Network Activity during a Small Workshop

Zhe Zhou, Mark Claypool, and Robert Kinicki
{jeffz,claypool,rek}@cs.wpi.edu
Computer Science, Worcester Polytechnic Institute
100 Institute Road, Worcester, MA 01609, USA

***Abstract*—Understanding current performance and usage of IEEE 802.11 networks is valuable for effectively managing and deploying wireless networks in the future. This paper analyzes a wireless trace from NetGames'08 and compares the results with previous research. The analysis indicates increasing dominance of Web traffic, emerging p2p applications consuming significant uplink capacities, and differences in the overall volume and traffic patterns from prior wireless Internet traffic measurements. The workshop schedule greatly influences network usage with the number of people in the room significantly affecting the received signal strength.**

I. Introduction

Advances in WiFi technologies, especially IEEE 802.11-a/b/g, coupled with decreasing costs have made wireless Internet access more common than ever. Business and academic institutions often provide ubiquitous wireless access, while home users frequently deploy wireless to conveniently handle multiple networked devices. Currently, many professional gatherings, such as conferences and workshops, provide participants with free wireless Internet access. Understanding trends in usage and current 802.11 performance is valuable for effective management and planned deployment of future wireless local area networks (WLANs).

One of the earliest known studies of a public WLAN analyzed user behavior in the wireless network deployed at Stanford [7]. Similarly, other studies have analyzed wireless networks but without corresponding application level analysis [3]–[5]. Known works that have analyzed both wireless layer and application layer characteristics include an in-depth look at wireless traffic from the 2001 SIGCOMM conference [1] and multi-layer network analysis of the WLAN at the University of Dartmouth [2]. While these works have been valuable for better understanding WLAN usage, the former case is now 8 years old and both cases are from considerably larger forums that are not representative of the many small to medium conference gatherings.

This paper examines wireless, network and application level characteristics for a ACM workshop. The merits of this trace are: 1) *a, b and g*: to the best of our knowledge, this is the first analysis of a WLAN with simultaneous IEEE 802.11 a/b/g traffic; 2) *2008 data*: the recency of the data, from late 2008, has merit since IEEE 802.11 technologies and Internet applications are constantly evolving; and 3) *size*: the size of the forum, about 30-40 participants, represents a workshop or small conference size that is quite common, but not well-studied.

The 7th annual workshop on Network and Systems Support for Games (NetGames'08),[1] was a single-track, ACM-sponsored workshop that had about 35 participants, with about 20 of them bringing their laptops and using the provided wireless network. Wireless access to the Internet was provided throughout the workshop sessions, including outside the room during breaks. Over 2.7 GBytes of traffic traces were gathered by the wireless infrastructure and archived for post-processing and analysis when the workshop completed. The analysis of these traces seeks to answer the following questions:

1. *What is the classification, frequency and duration of wireless network use during the workshop?* Answer: both IEEE 802.11a and 802.11g are heavily used by two-thirds of the participants. The average aggregate bandwidth usage was only 700 Kbps, but the standard deviation was 1.36 Mbps with a maximum of 8.73 Mbps.

2. *What applications are used by workshop participants?* Answer: despite the plethora of applications available to users, traffic is increasingly dominated by the Web,[2] while only about one-third of the participants use email services.[3] The clear preference for Web applications may be attributed to their client-side simplicity and ability to pass through firewalls.

3. *What factors impact network use?* Answer: external organization factors, such as session breaks, have a profound impact on traffic patterns. Wireless signal strengths have a negative correlation with the number of users in the room.

The rest of this paper is organized as follows: Section II presents related work; Section III describes the methodology to gather and analyze the wireless traces; Section IV analyzes traffic on the physical, link, transport, application and client layers; and Section V summarizes our work.

II. Related Work

In one of the earliest known studies of public WLAN use, Tang and Baker used tcpdump and SNMP to trace 74 users in the Stanford Computer Science Department for 12 weeks [7]. They analyzed user behavior of the network, overall network traffic and load characteristics. Later work includes

[1] http://netgames2008.cs.wpi.edu/
[2] Ports 80 and 443.
[3] Ports 993 and 995.

trace analysis of the Dartmouth campus wireless network conducted by Henderson et al. [2] They characterized user behavior in terms of application use and found residential traffic dominated all other traffic especially in residences with newer students. They also showed that some wireless cards roamed excessively, being unable to settle down with one AP. They compared traces gathered in 2001 and 2003/4 and found trends including an increase in the user population and roaming.

Balachandran et al. studied the SIGCOMM'01 wireless network traffic [1]. Their work resembles our work in that their trace was collected over an academic conference with similar analysis, but differs in the scale of the conference with SIGCOMM being much larger than NetGames. A comparison for all these three works in relation to ours is presented in Section IV.

Saroiu and Gummadi characterized Web and p2p traffic seen at the border routers on the campus of the University of Washington [6]. They observed that outbound p2p traffic dominated. However, in the study at Dartmouth [2], the dominance of inbound traffic was observed. Our study also notices a dominance of inbound traffic, but a noticeable presence of outbound traffic.

Both [1] and [2] observed the dominance of Web traffic, while in the newer trace of [2] the dominance of Web was diminishing. Our studies reveals that Web traffic continues to dominate and has, in fact, expanded in its prevalence.

Other studies of wireless traces focused more on network performance than on user behavior. Jardosh et al. analyzed traces collected at the 62nd IETF meeting over two days in 2005 [3]. They found wireless rate adaptation rarely used the 2 Mbps and 5.5 Mbps data rates, and that these lower rates were detrimental to network performance during congestion. Rodrig et al. analyzed SIGCOMM'04 wireless traces and they concluded that 802.11 overhead was high and retransmissions are common [5]. Also, clients switched data rates more often than not, and most transmission times were spent sending at 1 Mbps.

All studies aforementioned either examined IEEE 802.11b networks or investigated 802.11g networks but only with 802.11b rates, while our paper looks at a WLAN that supports both 802.11b/g and 802.11a.

III. Methodology

A. Network Configuration

NetGames'08 was held in Salisbury Labs, a multi-story, brick and steel building at Worcester Polytechnic Institute in which an IEEE 802.11a/b/g network is deployed for campus use. Figure 1 illustrates the floor plan showing the locations of wireless access points (APs) and the conference room where the workshop was held.

Each individual AP simultaneously supports 802.11a and 802.11b/g. Clients negotiate with the AP to determine which mode (a, b or g) to use based on client support, signal strength, and the AP load. The APs dynamically load balance across all available 802.11a and 802.11b/g channels. The subnet of APs connects via a switch to the campus gigabit backbone and finally to the Internet.[4]

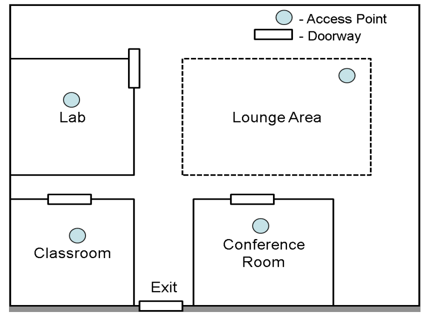

Fig. 1. Floor plan of NetGames'08 Forum

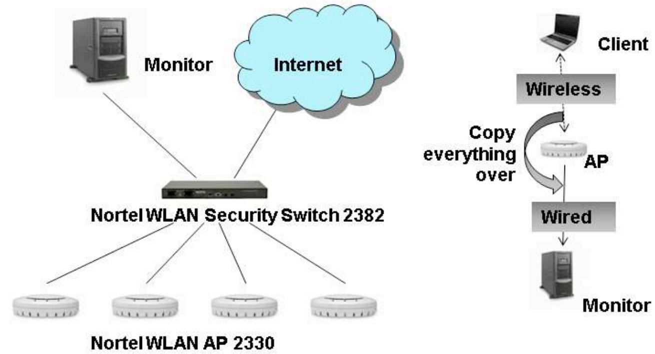

Fig. 2. Trace Collection Setup

B. Trace Collection

Figure 2 shows the trace collection setup. Transmitted and received wireless frames for the four APs in the subnet are forwarded via Ethernet to a monitor. Each frame collected by the AP contains link, network, transport, and application layer data encapsulated within the TaZman Sniffer Protocol (TZP) before being sent to the monitor in a UDP packet. The monitor stores `pcap` traffic it collects to local files for later offline analysis. Additionally, a laptop with a 802.11a/b/g compatible wireless adapter placed inside the conference room measured the received signal strength (RSSI) using a Windows Management Instrumentation script that queried the windows registry once every second.

The two-day workshop was held in October 2008 during a WPI term break when most of the students were away from campus. The four APs collected 20 million frames over the two day workshop. Since the first day had atypical network behavior in that wireless was not immediately available (owing to some difficulties encountered by the WPI network providers) and because users were not familiar with the environment (e.g., the location of power outlets), this study focuses on the

[4]For details of the WPI network infrastructure see http://www.wpi.edu/Academics/CCC/Netops/Network/

second day's traffic. Table I provides an overview of the trace statistics.

TABLE I
TRACE STATISTICS (2ND DAY)

Attribute	Values
Number of Channels	24
Total bytes transmitted	2.7 GBytes
Total hours of trace	7
Peak number of users	18
Peak throughput	8.4 Mbps

IV. ANALYSIS

Our analysis proceeds in a bottom-up fashion based on network layers: Section IV-A analyzes physical layer features, including channel information and RSSI variation with respect to the traffic (people and network) in the room; Sections IV-B and IV-C analyze link and transport layer traffic, respectively; Section IV-D shows the prominent applications used along with their inbound/outbound traffic rates during the day; and Section IV-E examines individual client activities and traffic.

A. Physical Layer

This section examines the physical layer aspect of the traffic traces including received signal strength indicator (RSSI) measurements obtained from the laptop in the conference room.

While control traffic, such as broadcast beacons, was sent on additional channels (2, 7, 48, 165, etc.), all observed traffic (shown in Table II) sent by the NetGames'08 participants appeared on channels 11 (IEEE 802.11b/g) and 60 (IEEE 802.11a). Generally, 802.11a provides data rates equivalent to 802.11g when wireless clients are close to an AP where the signal strength is good. Since the NetGames'08 participants were in a small conference room with the AP centrally located in the ceiling, the resultant strong signal strengths yielded more traffic on channel 60 than on channel 11. Our assumption is that when client laptops support 802.11a and have a good RSSI, 802.11a is employed to load balance the 802.11g channels.

TABLE II
CHANNEL TRAFFIC

Channel	Number of Frames	Volume (bytes)	Average Size (bytes)
60	6073329	1803797183	525
11	4712465	887727516	393

As a measure of link quality, the RSSI assists in AP association decisions and can be used in dynamic rate adaptation. RSSI values below a threshold are strong indicators of likely connection losses that interrupt or break applications relying on persistent connections. In this paper, the sniffing laptop was set to monitor channel 60 and recorded its RSSI value every second.

Figure 3 graphs RSSI over time in relation to background throughput on channel 60. RSSI varies considerably with little

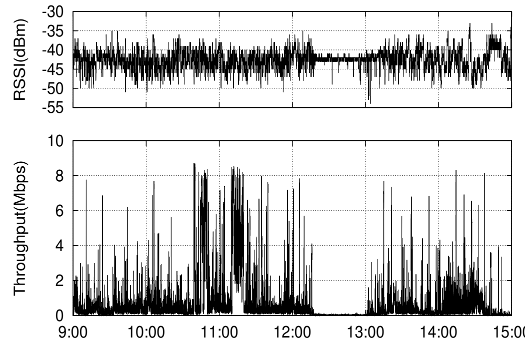

Fig. 3. RSSI Variation over Time

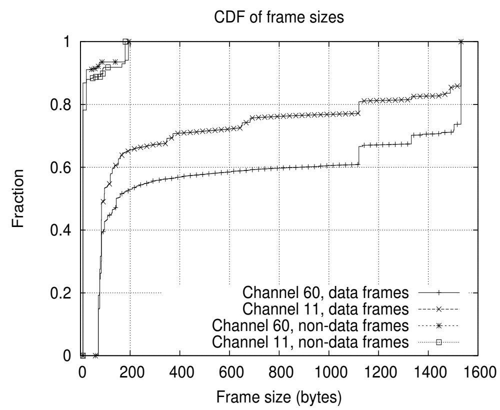

Fig. 4. CDF of Frame Sizes

visual correlation to background traffic spikes. When the room is occupied, the RSSI fluctuates in a range of -50 dBm to -35 dBm, even though the AP is in the ceiling directly overhead. However, when the room empties shortly after noon for a lunch break, there is a marked decrease in the RSSI variance. Note that -50 dBm to -35 dBm typically provides for good signal strength and participants in the room may see little variation in their application's performance. While the graph shows RSSI does not correlate directly with the traffic, it indicates that RSSI is affected by the people (and their laptops) in the room using the network.

B. Link Layer

Since channel 60 traffic resembles channel 11 traffic, for brevity, only channel 11 frames are analyzed in Table III. Most frames are data and acknowledgments. RTS is little used and most CTS frames are mixed 802.11b/g CTS-to-self frames.

Figure 4 provides the cumulative distribution function (CDF) for the data and non-data frames for both the heavily used channels, 11 and 60. While dominated by small ack and CTS frames, the non-data frame distribution includes the larger beacon frames. The figure shows 802.11a data frames to be, in general, larger than 802.11b/g data frames with approximately 10% more channel 60 data frames at the 1500-byte Ethernet

TABLE III
FRAME DISTRIBUTION (CHANNEL 11)

Type	Count	Frac.	Volume	Frac.	Size
Data	1809004	0.38	810593393	0.91	448
Ack	1831248	0.39	18312480	0.02	10
Beacon	204668	0.04	36840692	0.04	180
RTS	249	<0.01	3984	<0.01	16
CTS	690403	<0.01	6904030	0.15	10
Probe rqst	112062	0.02	8887496	0.01	79
Probe rspns	32227	<0.01	5381909	<0.01	167
Null data	31841	<0.01	764184	<0.01	24
Assoc rqst	143	<0.01	14433	<0.01	100
Assoc rspns	141	<0.01	6486	<0.01	46
Disassoc	13	<0.01	338	<0.01	26
Authentcte	311	<0.01	9330	<0.01	30
Deauthentcte	43	<0.01	1118	<0.01	26
Action	107	<0.01	7276	<0.01	68

MTU maximum. However, both channel frame distributions are dominated by the much smaller TCP ack packets.

C. Transport Layer

Transport layer (and some IP layer) data is presented in Table IV. Unlike in other conference measurement studies such as [1] where relatively low UDP usage is reported, 20% of the NetGames'08 traffic is UDP. Since smart p2p applications such as Skype will use UDP when participating p2p nodes are not behind NATed firewalls (such as at this workshop), the analysis in Section IV-D suggests this increased UDP component is most likely due to the one p2p file sharing application.

TABLE IV
TRANSPORT LAYER DATA DISTRIBUTION

Type	Count	Volume	Fraction
TCP	3215156	2002254869	0.79
UDP	1192505	510813319	0.2
GRE	28849	14670242	0.01
IGMP	3888	258310	<0.01
ICMP	3366	306671	<0.01
IPv6	125	28221	<0.01

D. Application Layer

Applications are classified according the IANA port numbers list, understanding that the list suggests, but does not force, port number assignments. Thus, user applications are determined mostly from well-known, server-side port numbers except for a single p2p file sharing application (uTorrent), which is identified by client-side port numbers. Table V provides the top 10 applications classified by port number with all remaining applications lumped into the 'other' bin. Clearly, this data confirms that Web traffic (`http` and `https`) dominate, contributing 73% of the traffic volume. While email (`pop3` and `imap4`), `nat-t` and `openvpn` traffic are not surprising members of the top 10 list, the p2p contribution sticks out as being different for NetGames'08 than for previous forums. Table VI compares the top five application types seen in NetGames'08 (where "session" is `ssh`, `vpn` and `rdp` combined) against those applications identified in three earlier studies: Stanford [7], Dartmouth [2] and SIGCOMM'01 [1]. "Average bytes" is the average number of bytes sent and received per user per hour.

TABLE V
APPLICATION CLASSIFICATION

App	Port	Count	Volume	Frac	Avg
http	80	2494223	1745068155	0.69	700
p2p	14900	721916	300657230	0.12	416
https	443	312781	109974470	0.04	352
nat-t	4500	235282	124695364	0.05	529
ssh	22	146601	28036250	0.01	191
imap4	993	102204	37451476	0.01	366
openvpn	1194	99302	63994068	0.03	644
netbios	137	60383	7628494	<0.01	126
rdp	3389	53970	20612017	0.01	382
pop3	995	43347	33558788	0.01	774
Other	n/a	105973	57408001	0.02	542

TABLE VI
APPLICATION COMPARISON ACROSS TRACES

Attrib	Stanford	SIGCOMM	Dartmouth	NetGames
Users	74	195	7134	20
APs	12	4	566	1
Length	12 weeks	52 hours	17 weeks	7 hours
Bytes	46 GBytes	4.6 GBytes	4.6 TBytes	2.7 GBytes
Avg. bytes	308 KBytes	440 KBytes	226 KBytes	19.3 MBytes
Top 5 apps	Web session FTP netbios X-term	Web (46%) session icp p2p email (6%)	Web (29%) file backup netbios p2p FTP	Web (73%) p2p session email (1%) netbios
Year	1999	2001	2003/4	2008

An important caveat here is that the applications in Table VI, defined by port number, are not likely to provide one-to-one mapping to actual user activities. For instance, port 80 is used to access email (via Web mail) as well as traditional Web browsing services. However, from another perspective, the proliferation of Web-based applications expands Web traffic to include such user activities as email, games, chat, video and live radio. For example, in our trace the increasing proportion of Web traffic is coupled with diminishing existence of email traffic.[5] While Internet use has diversified [2], the dominance of traffic labeled as "Web traffic" (`http`/`https`) has increased rather than decreased.

The presence of p2p traffic in the NetGames trace testifies to the popularity of such applications even within an academic conference setting. However, the fact that the p2p traffic is generated by a single client using `uTorrent` restricts our ability to draw general conclusions about current p2p impact. Yet, it is noteworthy that the lone p2p client in the NetGames trace contributes over 12% of the total traffic volume in an academic conference setting where the typical "leisure" applications supported by p2p are not expected to be a large part of the daytime traffic.

Figure 5 shows inbound and outbound traffic for the major applications. All the applications have more downstream traffic than upstream except for the p2p application where the total

[5]The Dartmouth trace fraction of email traffic is 6.4%.

upstream traffic exceeds the upstream traffic of all other applications combined. Measured application statistics are provided in Figure 6, showing the standard deviation of the mean and the coefficient of variation (CoV) in the overall traffic rate for each application type. To better understand application usage during the conference, Figure 7 aligns the throughput of the major applications over time.

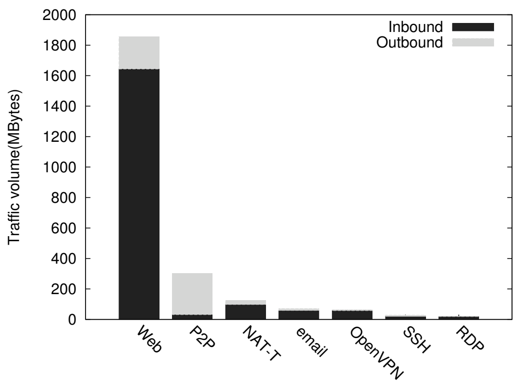

Fig. 5. Inbound & Outbound Traffic

Fig. 6. Application Statistics (in Mbps)

App	Min	Max	Median	Mean	Stddev	CoV
http	0.001	8.621	0.194	0.875	1.655	1.892
email	0.001	5.368	0.017	0.075	0.223	2.952
vpn	0.001	5.549	0.007	0.09	0.448	4.994
nat-t	0.001	4.35	0.045	0.114	0.218	1.917
ssh	0.001	4.116	0.003	0.021	0.118	5.608
rdp	0.001	0.098	0.058	0.056	0.14	2.491
p2p	0.001	0.01	0.262	0.172	0.118	0.686

Web traffic shows a wide range in load, with periods of low, medium and high throughput easily identifiable. Email throughput remains fairly low throughout the day, except for the throughput bursts in the early morning and late afternoon. The continuous, horizontal band at about 300 Kbps indicates periodic email synchronization. `Ssh` activities are more scattered and have high variability. While p2p throughput varies, it stays within the 0 to 400 Kbps region. `Vpn`, `nat-t` and `rdp` patterns are tied to the handful of individuals that used these applications.

The organization of the workshop added external factors that affect network use. The second day schedule for the workshop was as follows:

`9:00 - 10:30` Morning session (3 speakers)
`10:30 - 11:00` Break
`11:00 - 12:00` Panel session
`12:00 - 13:30` Lunch break
`13:30 - 15:00` Afternoon session (3 speakers)

The morning session has the most variation overall. In the first part of the morning there was high `nat-t` traffic, and then at about 9:20am there were a jump of network activities for Web, email and `nat-t`, showing a coordinated pattern. When the morning break started (indicated by the left-most vertical line), the overall throughput of Web and p2p dropped, indicating people left the room. After about 10 minutes, people

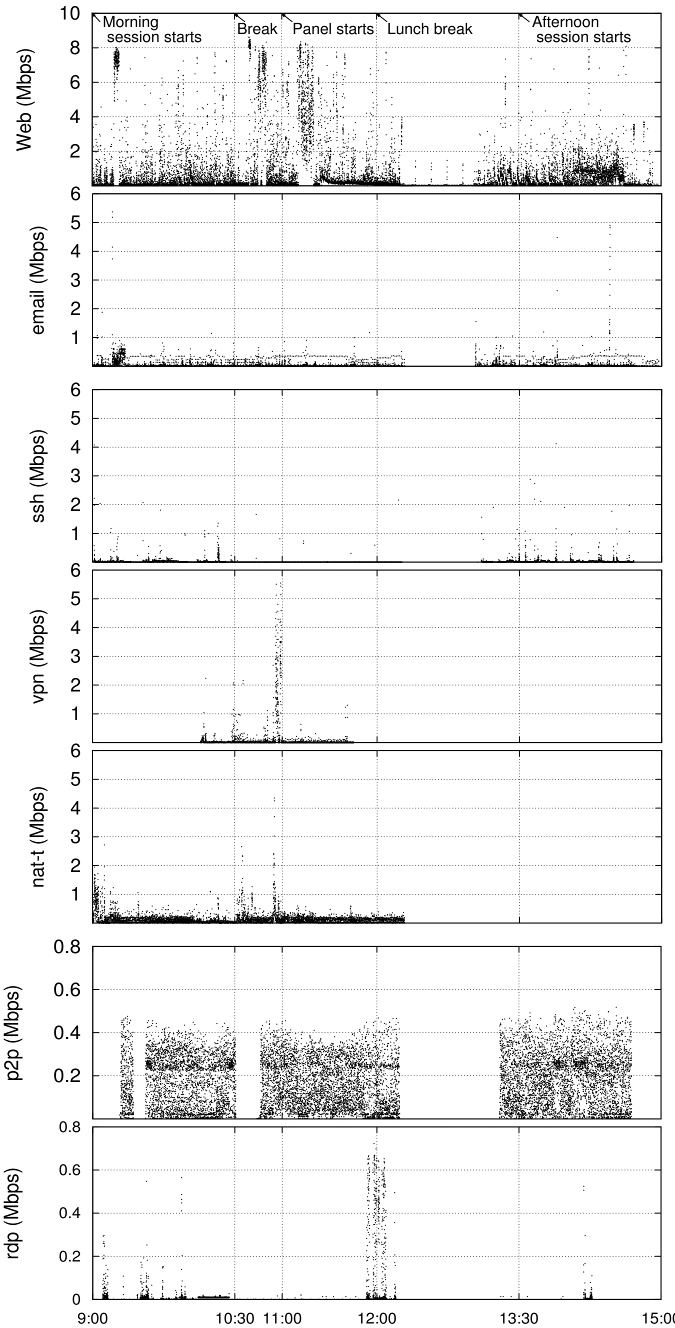

Fig. 7. Application Throughput over Time

started returning to the room and there was an increase in Web, `vpn`, `nat-t`, and p2p throughput. When the panel started at 11:00am, Web traffic diminished and then went up (but with fewer users, as shown in Figure 8), while `vpn` and `nat-t` went to their normal low. Most prominently, the lunch break at 12:00 provided a quiet period, with some traffic at the beginning and end. The small amount of periodic Web traffic seen during the lunch break was due to the authentication of clients to the WPI Web server, needed to retain guest access to the wireless network. In the afternoon session, `nat-t`, and `vpn` disappeared completely. Web traffic stayed comparably low to the morning except for a burst in the late afternoon. Email also had a burst in the late afternoon, possibly indicating some end-of-day checking and sending of email before the workshop ended and the participants dispersed. P2p's throughput covers large areas of the graph, except for during the breaks, but has a small Coefficient of Variation (CoV). `Ssh` has the largest CoV, followed by `vpn` and email, indicating more bursty traffic.

E. Clients

Client activeness provides another view into how participants used the wireless network. In particular, this section seeks an understanding of the number of clients that simultaneously used the wireless network during the workshop and how many clients used the popular applications of Web and email. A client is marked as active if it sends/receives a frame at least once a minute and active for Web/email if it sends/receives a packet to/from a Web/email port once a minute.

Figure 8 shows the number of active clients over time for Web, email, and all other activities. At any given time, there are fewer than 18 clients connected, about half of the 35 workshop participants. During lunch, the number of active clients drops to fewer than 4, and these clients generate periodic traffic that is likely an automatic refresh of Web pages or email. The number of active email clients is consistently from 1 to a maximum of 7.

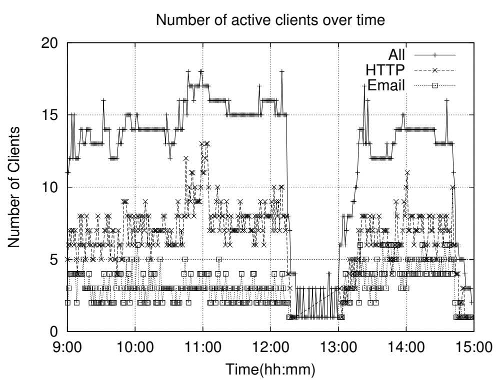

Fig. 8. Clients over Time

Figure 9 plots the inbound and outbound traffic volume for each client (identified by their MAC addresses), sorted from highest volume to lowest. Most clients used more downlink capacity, with the exception of client `e1:9c` that used more uplink capacity when running a p2p file sharing application.

Figure 10 shows the complementary CDF (CCDF) of the inbound and outbound traffic of all clients. The trendlines are not heavy-tailed.

V. Conclusion

Growth in wireless networks has led to the increasing availability of wireless Internet access in forums such as workshops and conferences. Analysis of wireless traces can help better understand network use. While previous work is either outdated or from large-scale wireless local area networks, this paper analyzes wireless Internet access from NetGames'08, a small-scale, single-track ACM workshop typical of many small-group gatherings. Our analysis examines the physical, link, transport, application and client layers. The findings show: Web remains and is perhaps increasing as the dominant application; p2p and `vpn` are emerging in use; wireless use is directly influenced by workshop organization; and people using wireless in the vicinity of the wireless access point significantly affect received signal strength. Researchers can use this analysis and these traces to build synthetic workloads for experiments or plan for improved deployment of wireless networks at future forums.

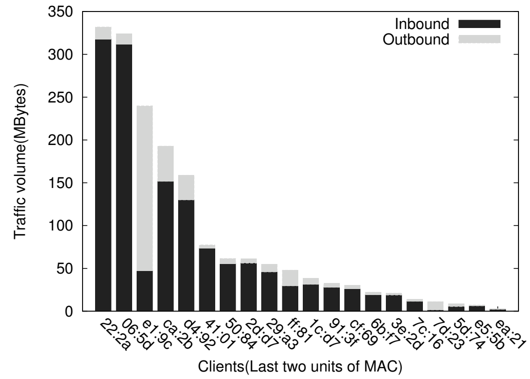

Fig. 9. Inbound & Outbound

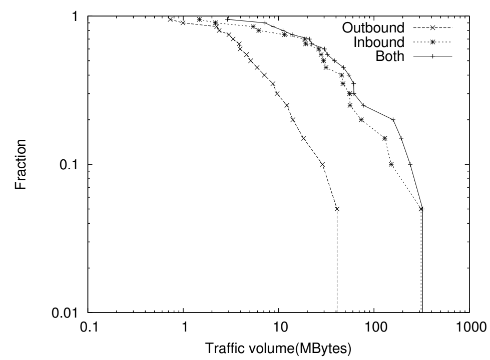

Fig. 10. CCDF of In & Out

References

[1] A. Balachandran, G. Voelker, P. Bahl, and V. Rangan. Characterizing User Behavior and Network Performance in a Public Wireless LAN. In *Proceedings of ACM SIGMETRICS*, Marina del Rey, CA, USA, 2002.

[2] T. Henderson, D. Kotz, and I. Abyzov. The Changing Usage of a Mature Campus-wide Wireless Network. In *Proceedings of MobiCom*, Philadelphia, PA, USA, 2004.

[3] Amit P. Jardosh, Krishna N. Ramachandran, Kevin C. Almeroth, and Elizabeth M. Belding-Royer. Understanding Congestion in IEEE 802.11b Wireless Networks. In *Proceedings of SIGCOMM IMC*, Berkeley, CA, USA, October 2005.

[4] R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan. Analyzing the MAC-Level Behavior of Wireless Networks in the Wild. In *Proceedings of ACM SIGCOMM*, Pisa, Italy, September 2004.

[5] M. Rodrig, C. Reis, R. Mahajan, D. Wetherall, and J. Zahorjan. Meas-based Characterization of 802.11 in a Hotspot Setting. In *Proceedings of ACM SIGCOMM E-WIND*, Phil., PA, USA, August 2005.

[6] S. Saroiu, P. Gummadi, and S. Gribble. Measuring and Analyzing the Characteristics of Napster and Gnutella Hosts. *Multimedia Systems Journal*, 9(2):170 – 184, August 2003.

[7] D. Tang and M. Baker. Analysis of a Local-Area Wireless Network. In *Proceedings of MobiCom*, Boston, MA, USA, August 2000.