A Dense Wireless LAN Case Study

***Abstract*—**
While usage and trends for campus wireless LANs have been extensively studied through various deployment instances, very few studies have been done to study short term deployments of wireless LANs, such as those provisioned at the conference venues. We conduct a detailed study of a dense wireless LAN consisting of 82 access points (APs) deployed to serve over 5000 SuperComputing 2004 users. By studying traffic generated by all layers of the TCP/IP protocol stack, we make several observations that differ from those made by the previous wireless LAN studies. In particular, we find that: applications lacking congestion control were more popular than previously reported, channel error rates were higher, a very high percentage of broadcast traffic was present, and that the client mobility was higher than what was previously observed. The two novel features of our custom traffic analyzer that made these observations possible are: 1) it looks at traffic both from the perspective of the total number of bytes, as well as the total number of packets and 2) it conducts a bi-directional traffic analysis from wireless clients to APs and from APs to clients.

I. Introduction

Wireless LANs (local area networks) are now an expectation at airports and conferences, on campuses, and in hotels. They offer mobility to their users and are an ideal choice where temporary network connectivity is desired due to the ease of deployment.

While many studies have been conducted [1], [2], [3], [4], [5], [6], [7] to understand the usage and trends of the wireless LANs deployed on campuses, very few [8] have been done to characterize the temporary deployments of wireless LANs, like those typically provisioned at the conference venues.

This paper reports on the study of a dense wireless LAN consisting of 82 access points (APs) that was provisioned for the 5000 SuperComputing 2004 (SC2004) attendees in a $236,900$ square feet exhibit floor. The data for this study was collected by the Porcupine [9] wireless data analysis platform, which sniffed 802.11b traffic passively from the SC2004 exhibit floor. We used a custom traffic analysis tool[1] which considered two traffic perspectives. The first, *bytes* perspective, looked at the aggregate bytes generated by various layers of the TCP/IP protocol stack. The second, *packets* perspective, considered the total number of packets generated by each protocol. The packets perspective is a novel and more appropriate way of looking at traffic because of the explicit contention resolution protocol required to secure a 802.11 wireless channel for each packet transmission. Another novelty of our custom traffic analyzer is that in addition to allowing aggregate traffic analysis, it allows traffic analysis to be performed individually for each direction of traffic: from client to APs and from APs to clients. Prominent observations from our analysis include:

- Transport protocols UDP and IGMP caused more packets to be scheduled than TCP. Given that they do not perform any type of congestion control, this could degrade the performance for TCP-based applications.
- A custom private encapsulation protocol, protocol99 (we name it so due to the protocol number contained in the protocol field of the IPv4 packets), consumed a significant amount of bandwidth, both in total number of bytes and packets.
- Though the overall channel error rates were $5-8\%$, channel error rates for frames from the clients to their APs were about 25%, implying that one in every three frames from the clients were lost!
- Broadcasts due to protocols like IGMP and protocol99 comprised about 44% of the total packets sent by the APs to their clients. This appears to be an unusually high percentage and could have caused client performance to degrade.
- The median available bandwidth to long-lived bandwidth intensive TCP flows was a mere 130kbps, a very small fraction of the 11Mbps allowed by 802.11b.
- The distribution of the number of APs that a client associated with during each day was exponential, implying that some clients were highly mobile. One of the effects of this was the significant percentage of DHCP traffic observed in our workload.

The rest of the this paper is organized as follows. Section II describes the related work. Data collection is described in section III and section IV describes the results of the analysis performed on the SC2004 workload. Section V concludes with a discussion of the

[1]We found that open source traffic analysis tools such as fprobe [10] and ARGUS [11] do not provide the necessary granularity for detailed traffic classification at the link layer. As a result, we wrote a custom tool for the analysis presented in this paper.

issues raised by this research.

II. Related Work

Tang and Baker [1] conducted among the very first studies of wireless LANs. They used tcpdump and SNMP data to analyze the behavior of 74 users served by 12 APs in the Stanford Computer Science Department over a 12-week period in 2000. Their focus was on characterizing user mobility, wireless LAN usage pattern, application mix, and peak throughput rates achieved. They found that users tended to not move much within the building and 80% of the time the peak wireless LAN usage was caused a single user and application. Further, the maximum simultaneous users in their trace was 17, allowing peak throughputs to reach as high as 5.6Mbps. They introduced the notion of using number of packets to measure traffic (in addition to total bytes) and found *http* to be the most popular application both in terms of the number of packets and bytes.

Hutchins and Zegura [2] analyzed Georgia Tech campus trace (not including the residential halls) for 444 wireless users served by 109 APs over a two month period. Their main focus was on mobility and user session information using authentication logs. They observed diurnal usage patterns and that half the users tended to move across several of the 18 buildings while the other half did not move at all.

Kotz et al have conducted two studies of Dartmouth College campus-wide wireless LAN. The trace used in their first study [3] contained information about 1706 users across 476 APs over a 12 week period in Fall 2001. The trace for their second study [4] was even more extensive and contains information about 7000 unique users using 550 APs over the course of a 17-week period in 2003/4. Both their studies focussed on macro-characteristics of the campus wireless LAN and track user mobility, session duration, application mix, and overall traffic. In particular, their second study observed an increase in the user population, roaming, and overall traffic. The applications used on campus also changed significantly in 2003/4 as compared to 2001, with *http* consuming only 28.6% of traffic bytes compared to the 62.9% in the earlier study. Also, the bandwidth used by peer-to-peer (P2P) applications increased from 5.2% in 2001 to 19.3% in 2003/4.

Balazinska and Castro [5] analyzed a trace containing information about 1366 corporate users on 117 APs over 4 weeks. They focused on population characteristics, load distribution across APs, user activity, and mobility. They found the load on a AP to be dependent on its location, and that the aggregate data transfer rate seen by an AP did not depend on the number of users associated with it, but rather on which users are present. Just like the other studies, their users tended to not move around much, and when they moved their network usage remained unchanged. While this study complemented the existing studies in terms of many of their findings, its main contribution was in characterizing user mobility. They found that the probability distributions of *persistence* (a measure of how long users stay continuously associated with the same AP) and *prevalence* (reflects how frequently users visit various locations) followed power laws.

More recently, Schwab and Bunt have analyzed the wireless LAN at University of Saskatchewan [6]. Their trace covers 136 users on 18 APs over 7 days and contains information about each packet sent on the LAN and the authentication records of all users. They found that fluctuating signal strength at the edge of the APs range caused exceptionally high number of authentications per user. Also, *http* comprised 28% of packets. They were only able to identify 1.5% of P2P traffic but it is likely that much of the 35% of the unidentified TCP packets contributed to P2P traffic as well.

Chinchilla et al [7] tracked logs from 222 APs and 7694 *Web* users over a 11 week period on the University of North Carolina campus wireless LAN. Their goal was to explore characteristics of the wireless environment in accomplishing caching, prefetching, coverage planning, and resource reservation. They found that 13% of the unique URLs contributed to 70% of the accesses by the wireless *Web* clients and that user caches and caches attached to APs would be beneficial to deploy.

While all the above studies analyze macro-characteristics of wireless LANs like overall usage patterns, user mobility, traffic characteristics, and application mix, Balachandran et al [8] focussed on micro-characteristics like the bandwidth requirements of individual users and the load on individual APs. They analyzed user and network performance of 195 attendees of SIGCOMM2001 that were served by 4 APs. Their wireless LAN was much smaller than ours (4 APs and 195 users) but similar in nature because it is set up to serve users for a short duration and no prior information about the usage is known. We find our application mix and traffic characteristics to be different from theirs. This could be due to a variety of reasons: 1) our user base of 1699 users is much bigger and more *diverse* than theirs and 2) our trace is 3 years newer and Internet applications have evolved since then.

III. Data Collection

The data for this study was collected on November 8 and 11 during the SuperComputing 2004 (SC2004) conference. SC2004 was held in the David L. Lawrence Convention Center in Pittsburgh, PA on November 6-12, 2004 and brought together researchers from industry, academia, federal government, and not-for-profit sectors on issues of high performance computing, networking, and storage. The conference program featured exhibits of products and research , and a technical program comprised of peer-reviewed papers. Network connectivity for the $5000+$ SC2004 participants was provided by both wired and wireless (802.11a/b/g) networks in the $236,900$ square feet exhibit floor, select common areas of the convention center, meeting rooms, and technical program areas. The wireless LAN was connected to the wired network through a switch.

The data for our analysis was collected using the Porcupine, a wireless data analysis platform which was configured in passive only mode, sniffing 802.11b traffic form the air. The Porcupine stations were located in the main exhibit floor and collected around 1.8GB of packet traces on each day. We will now describe the environment in which our capture system operated followed by an overview of the Porcupine.

	Day 1	Day 2	Overall
Number of unique wireless users	1433	604	1699
Number of APs	96	82	106
Number of sessions	678,689	607,407	1,286,096
Total management traffic (in MBytes)	238.2	137.3	375.5
Total management traffic (in packets)	4,023,533	2,556,590	6,580,023
Total user data traffic (in MBytes)	1409.4	1210.6	2622.1
Total user data traffic (in packets)	9,415,414	7,003,049	16,418,463

TABLE I
Overall data statistics.

A. Wireless LAN Environment

The wireless LAN at the SC2004 exhibit floor was provided by Trapeze Networks [12]. It consisted of 82 MP-252 mobility points (aka access points or APs) to provide 802.11a, 802.11b and 802.11g wireless access to SC2004 attendees. A mobility exchange switch (MX-400) was used to connect the wireless users to the Internet.

Wireless users were prohibited from using their own APs at these radio frequencies in the entire convention center. Also, to avoid interference with other wireless users they were asked not to operate in ad-hoc or peer-to-peer mode. Users were assumed to be responsible for the security and privacy of their sessions as per their requirements and no authentication mechanisms or firewalls were used. Finally, no link level encryption mechanisms like WAP (wireless application protocol) or WEP (wireless equivalent privacy) were used.

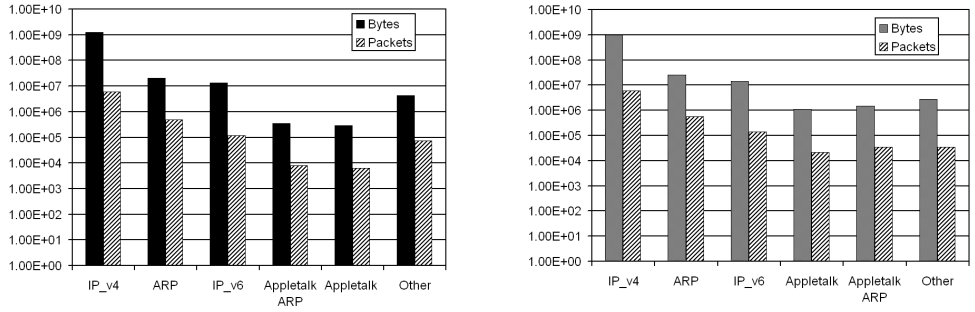
B. Porcupine Wireless Analysis Platform

The *Porcupine* [9] is an experimental wireless analysis platform developed with the goal to support research into advanced wireless network management. It is equipped with sixteen separate radios, sixteen single board computers, and sixteen directional antennas, each of which can be configured to operate together, or individually. A master console is used to synchronize the activity of the Porcupine's sixteen separate units. The Porcupine's basic platform is open source; Linux running on off-the-shelf single board computers and commodity radios.

The Porcupine can operate passively, as an active wireless client, or as a wireless AP operating on multiple channels and covering a large area with multiple high-gain antennas. In passive mode, the Porcupine can capture wireless packets and determine their origin in real-time. At SC2004, the Porcupine was demonstrated on the exhibit floor in *passive mode* and sniffed all 802.11b traffic (both user data and management frames) on November 8-11.

The advantage of passive sniffing is that it allows one to observe the unconstrained behavior of the wireless users. This mode also allows the detection of rouge APs, ad-hoc networks, and misconfigured clients that cannot be detected by any other means. The downside of passive sniffing however, is that it only captures wireless traffic that any of the Porcupine antennas are able to sniff and decode properly. In fact, for SC2004 demonstration all the sixteen Porcupine antennas were configured to simultaneously cycle through the 802.11b channels. Further, the presence of metallic stages on the exhibit floor restricted Porcupine's radio range, causing its antennas to miss traffic. Since we do not have a quantitative estimate of the percentage of the total wireless packets missed by the Porcupine the analysis presented in this paper can only provide lower bounds on the actual performance problems.

Another limitation of the wireless data passively sniffed by the Porcupine relates to the variety of 802.11 modes offered at the SC2004 exhibit floor. The Porcupine is designed to capture only 802.11b wireless packets. This implies that while the effect of the presence of 802.11g traffic would be indirectly observed in the



(a) Day 1.

(b) Day 2.

Fig. 1. Traffic due to top 6 LLC types.

performance of 802.11b, directly observing 802.11g traffic, which shares the 2.4GHz radio frequency spectrum with 802.11b is not possible. Also, no information on the 802.11a, which operates in 4.2GHz frequency spectrum is available.

Due to storage constraints the team demonstrating the Porcupine wireless analysis platform in passive mode at SC2004 was not able to save all 4 days of 802.11b traffic. We have available to us traffic trace for two full days of November 8th and November 11th (referred to as day 1 and day 2 subsequently). Table I summarizes the overall statistics of traffic collected by the Porcupine. The number of APs in table I are different from those provisioned by SC2004 due to the presence of 14 and 12 rogue APs on day 1 and day 2 respectively. Only 2 of the rogue nodes were different across the two days.

IV. Traffic Analysis

This section describes the results of the analysis performed on the SC2004 workload.

A. Network Layer Protocols

We began our analysis with a study of the prevalence of various network layer protocols. Though it is a well accepted fact that IPv4 is the protocol of choice at the network layer, we decided to test it out for out logs. Figures 1(a) and 1(b) show the total bytes and packets for the top six logical link layer (LLC) types appearing at the data link layer, as observed in our data.

As figures 1(a) and 1(b) show, IPv4 packets and bytes are almost two orders of magnitude larger than those for any other protocol. ARP is the next most prevalent protocol and IPv6 has an insignificant presence. Since the wireless LAN under study is heavily biased toward IPv4 traffic we limited the subsequent analysis to IPv4 traffic.

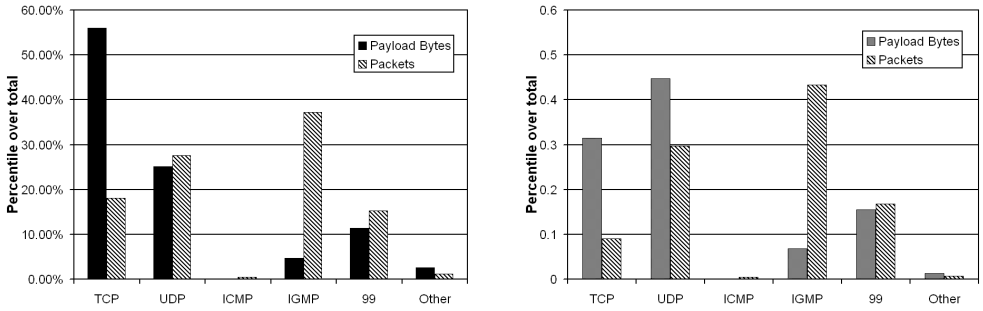
B. Transport Layer Protocols

The transport protocols differ in their behavior and hence impact the the network differently. For example, TCP [13] is connection oriented and performs congestion control while UDP [13] is connectionless and performs no congestion control. Also, while TCP and UDP are unicast, IGMP [14] is a broadcast protocol. Due to these differences, the exact mix of various transport protocols is an important factor in understanding the performance of a network.

Figures 2(a) and 2(b) show the percentage of traffic carried by various transport protocols for day 1 and day 2 respectively. Each of these figures present two different perspectives on the traffic: the *bytes* perspective and the *packets* perspective. While bytes and flows are traditional methods to analyze traffic, we introduce the *packets* perspective because is a better measure of the popularity of a protocol in a wireless medium because channel contention is resolved per packet, irrespective of the number of bytes contained in the packet. The novelty of using *packets* to understand traffic leads to several important insights into the popularity of various transport protocols. We elaborate on these next.

Two conventionally accepted beliefs about the prevalence of transport protocols in the Internet are the following: 1) most Internet traffic is carried over TCP and 2) the fraction of traffic contributed by transport protocols other than TCP and UDP is insignificant. These beliefs have been confirmed in wireless contexts by various studies (for example, [8], [6]). In fact, these are such deeply rooted beliefs that recent studies like [4] have only considered TCP and UDP traffic in their characterization studies of campus wireless usage.

As shown by figures 2(a) and 2(b), none of these beliefs hold for our data when the we consider the traffic percentages of each transport protocol in terms of the



(a) Day 1.

(b) Day 2.

Fig. 2. Transport protocols in bytes and packets.

number of packets. The first belief that most Internet traffic is carried over TCP does not hold because in figure 2(a) even though TCP accounts of 50% of bytes, it only accounts for 18% of total packets. In fact, the total number of packets for TCP are less than both UDP and IGMP. Thus, even though TCP carried more data, both UDP and IGMP required more packets to be scheduled. The second belief that traffic contributed by transport protocols other than TCP and UDP is insignificant does not hold because even though one might tend to ignore the presence of IGMP in terms of total bytes contained in its payload, the total number of IGMP datagrams (37%) outnumbered those of TCP and UDP. The ICMP traffic was negligible. However, a custom protocol, which we refer to as protocol99 (private security encapsulation allowed by IANA) because of the protocol number contained in the protocol field of the IPv4 packets consumed a significant amount of bandwidth, both in total number of bytes and packets.

As shown in figure 2(b), UDP carried more traffic than TCP both in terms of total number of packets and total bytes contained in packet payload on day 2. Also, more IGMP packets were scheduled than any other transport protocol. protocol99 also caused significant amount of traffic, next only to UDP and IGMP in terms of number of packets.

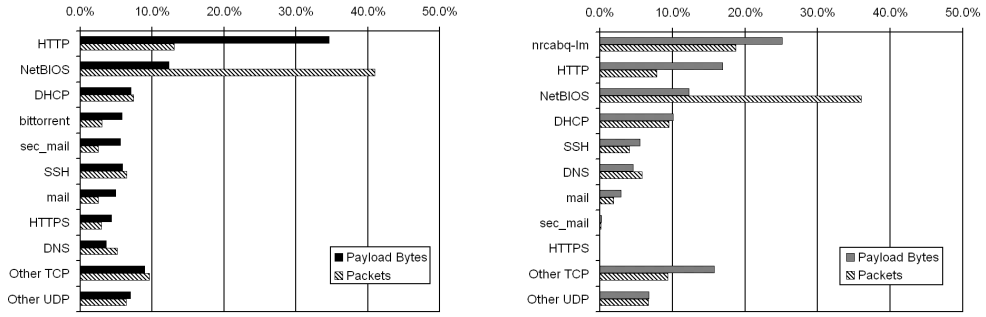
There is anecdotal evidence that SC2004 attendees experienced performance problems. The above analysis of transport protocols on both day 1 and 2 leads us to conjecture that the unusually high traffic contributed by UDP, IGMP, and protocol99 (especially in terms of the number of packets) might have caused TCP congestion control to back-off, hence causing performance degradation for applications using TCP.

1) protocol99: Protocol99 was being used in a very interesting fashion. All packets from this protocol: 1) had the same source IP address 10.0.0.1 (private IP address space), 2) were destined to the same IP address 255.255.255.255 (broadcast), and 3) came from the wired network. However, 15 distinct MAC source addresses were observed to be associated with protocol99 packets of which 14 also had a valid SC2004 IP address. We were able to understand the behavior of protocol99 only because our traffic analyzer allowed the examination of the MAC addresses. Just looking at the IP addresses would cause this protocol to be accounted as just one misconfigured host. Further, the fact that most of the participants in this custom protocol had also valid IP addresses points to the use of an obfuscated ”secure” communication channel. Any further analysis of protocol99 was not possible due to its custom nature.

C. Applications

In this section we look at the TCP and UDP application mix contained in our data. To infer names of the applications from the source and destination port numbers contained in the packets we use the following criterion: a packet belongs to a known application protocol if the lower port number contained in it matches a well known port number. For example, a TCP datagram with source port 80 and destination port 9084 will be classified as http traffic. Figures 3(a) and 3(b) show the traffic percentages of top 9 applications in bytes and packets for day 1 and 2 respectively.

While previous studies of wireless LANs conducted in $2001/2002$ [3], [8] found HTTP traffic (in bytes) to be between $50-65\%$, more recent similar characterization studies done in 2004 [4], [6] have observed the corresponding percentages to be between $27-30\%$ even though it continues to be the dominant application in the Internet. It is generally believed that the one of the reasons behind the observed decrease in the percentage of



(a) Day 1.

(b) Day 2.

Fig. 3. Application mix in bytes and packets.

HTTP traffic is due to the popularity of TCP-based peer-to-peer (P2P) applications like Gnutella [15], Kazaa [16] and Bittorrent [17]. This remains an unconfirmed fact because various recent studies have found substantially different percentages of P2P traffic in their workloads. For example, the study of campus wireless LAN by Henderson et al [4] found P2P traffic to be 19% of overall in terms of bytes but the study by Schwab and Bunt [6] found the same percentage to be only 1.5%[2]. We now investigate the presence of P2P in our logs.

From figures 3(a) and 3(b), we find that HTTP accounts for 34% and 17% of the traffic in bytes on day 1 and day 2 respectively. Also, it dominates only on day 1 in terms of the number of bytes because on day 2, a UDP-based multimedia application running on port 1458, called *nrcablq-lm*, dominates HTTP both in terms of percentage of packets as well as bytes. In terms of the number of packets, HTTP is not the dominant application on either of the days. Instead, NetBIOS[3], another UDP-based application, dominates the traffic on both days with 42% and 37% of total packets on day 1 and day 2 respectively.

While figure 3(a) does not contain enough information to comment on whether the HTTP traffic on day 1 reflects the intentions of users using HTTP for Web transfers, the reduction in the percentage of HTTP traffic by 50% on day 2 with respect to day 1, the absence of Bittorent in top 10, and a reduction in the percentage of traffic from mailer applications when UDP-based applications *nrcab-lm* and DHCP with no congestion control increased their traffic leads us to believe that the lower traffic due to TCP-based applications may not be a reflection of user intentions, but a result of a back-off by TCP congestion control upon encountering congestion.

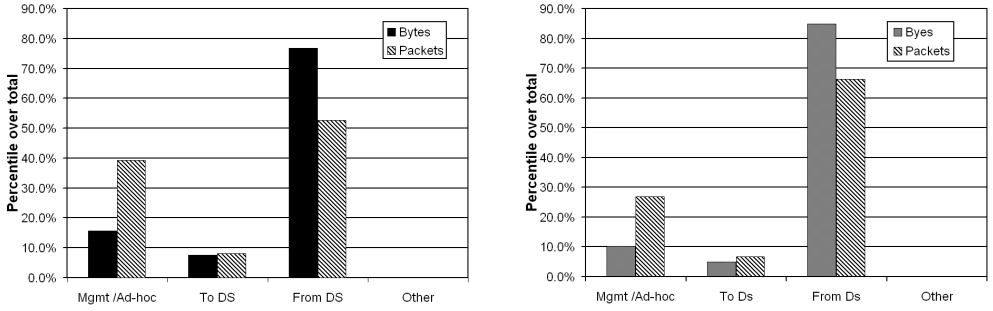
In addition to the applications discussed above, the other prominent applications shown in figures 3(a) and 3(b) include: HTTPS, mailer applications and their secure versions (latter denoted by sec_mail), SSH, Bittorrent, DHCP, and DNS. Out of these DHCP and DNS are UDP-based and the rest are TCP-based. Bittorrent contributes only about 6% of traffic in terms of bytes (even lesser in terms of the number of packets) on day 1 and is not in the top 10 applications on day 2. This seems to indicate that P2P applications were not very popular among SC2004 participants. The unusually high DHCP traffic on both days, both in terms of packets and bytes is surprising because this percentage is much higher in comparison to all the other previous wireless LAN studies. There could be several reasons for this: problems in the leasing of IP addresses in the network, presence of many associations and disassociations by the wireless clients, or the presence of unauthorized DHCP servers. We investigate if high number of client associations and disassociations indeed occurred in section IV-G.

D. Channel Errors

The study by Balachandran et. al. [8] looked at the channel error rates in a smaller conference setting and found the median packet error rate to be 2.15%. Since we only capture transmitted packets we can not directly measure the packet loss. However, we can estimate the channel error rate by identifying retransmitted data link layer frames from their original transmissions. Understanding retransmissions is an important factor in understanding performance because even though retransmissions do not impact the available bandwidth in the system, they impact the packet delay in a significant matter.

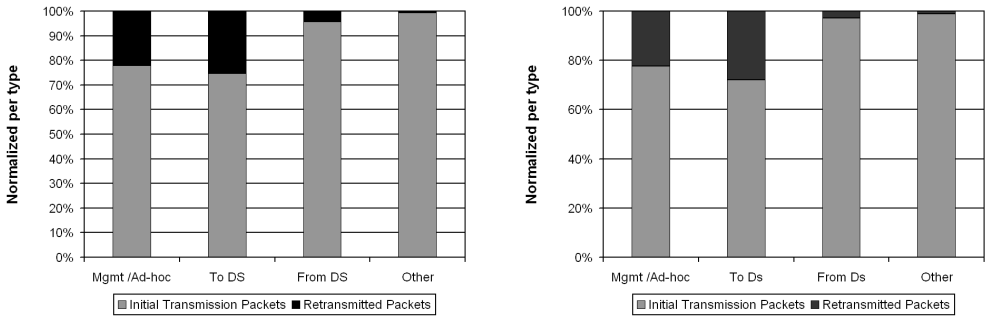
[2]This difference could be attributed to the use of non standard ports by P2P applications to escape legal hassles.

[3]A LAN extension to the BIOS for PCs.



(a) Day 1. (b) Day 2.

Fig. 4. Data link layer traffic. *To DS* denotes traffic from the wireless clients and *From DS* denotes the traffic from the APs.



(a) Day 1. (b) Day 2.

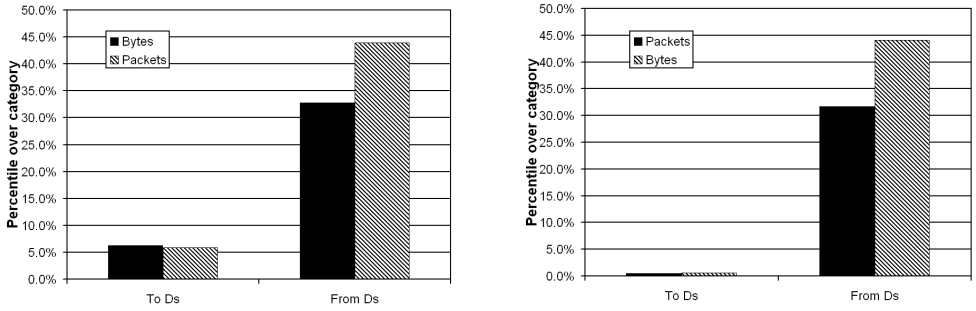
Fig. 5. Channel error rates (normalized per link layer frame type).

We find the aggregate channel error rate for the data to be 7.64% and 5.35% for day 1 and day 2 respectively. Since this error rate is higher than what was previously observed, we decided to investigate if the error rate was higher in a particular direction of traffic. This is a novel way of looking at the data and led to some surprising results. We first looked at the data link layer traffic and figures 4(a) and 4(b) show the traffic sent and received by the wireless clients on day 1 and day 2 respectively. We classify packets at the link layer by the two status bits on the 802.11 frame. The four categories that result are: *to distribution system (DS)*, *from DS*, *management* and *other*. *To DS* frames are frames that are sent from an associated client to its access point (AP). *From DS* frames sent by the AP to its clients. The frame is for an ad-hoc network or is a management frame when neither of this bits are set and we refer to such frames as *management*[4]. When both bits are set, the frame is for Ethernet bridges and we refer to them as *other*.

We notice from figures 4(a) and 4(b) that the traffic is very asymmetric both in terms of packets and bytes, with 77% (85%) of bytes and 53% (67%) of packets coming from the DS compared to the less than 10% of the bytes and packets from the wireless clients on day 1 (day 2). Also, nearly 40% (27%) of the bytes and 16% (10%) of packets are management frames on day 1 (day2). Hardly any traffic was destined for the Ethernet bridges.

One would expect the transmission error rates to be proportional to the traffic in each direction. However, as figures 5(a) and 5(b) show, this is not the case for our data. We find a much larger amount of data loss in two classes: *management* and *to DS*. Frames from the wireless clients have a 25% (24%) packet retransmission rate on day 1 (day 2), which means that one in every packets was lost. In comparison, the retransmission rate for the frames from the AP is only about 5% (4%) for day 1 (day 2). The high error rate in the management frames is also in the management frames sent by the clients. The asymmetrically high loss of client frames

[4]Less than 0.1% of the total frames in this category were ad-hoc network frames on both days.



(a) Day 1.

(b) Day 2.

Fig. 6. Presence of broadcast in bytes and packets.

could be because of *hidden nodes* due to which two clients do not realize each other transmissions, causing their frames to collide in mid air or because. Another possibility is that the *density* of wireless network made packets from different APs overlap in space and time. We conclude that channel error rate was asymmetrically high in SC2004 logs and could lead to degradation in client performance.

E. Broadcast

We saw in section IV-B that IGMP and protocol99 caused a very high percentage of overall traffic in terms of the number of packets. These protocols are broadcast-based. We now investigate the extent to which broadcast[5] was present in our data.

Figures 6(a) and 6(b) show the broadcast traffic in packets and bytes as a percentage of the total traffic for day 1 and day 2 respectively. Though broadcast traffic was a modest 6% for the clients on day 1, it comprised 33% of the bytes at the APs. The corresponding numbers for day 2 were $< 1\%$ and 32% respectively. The percentage of broadcast traffic at the APs in terms of the number of packets was even higher, 44% for both days. We do not have a comparison point for these percentages since none of the previous studies that we are aware of looked at this issue. However, this appears to be an excessive amount of broadcast and can affect client performance.

F. Available Bandwidth

We now explore the available bandwidth to various applications because it is a good measure of the performance of a network. To estimate the available bandwidth on a wireless network we chose *bulk* TCP flows with larger than 10 seconds duration and whose average packet size was greater than 1.1KBytes. The goal was to ensure the selection of those flows that required high bandwidth for a long enough duration. We focussed only on TCP traffic because the sequence numbers on TCP traffic enabled us to compute the actual bytes transfered. The available bandwidth for each flow was calculated by using the TCP header sequence numbers of the last and fist packets for each flow.

Figures 7(a) and 7(b) show the maximum bandwidth, average bandwidth, and the standard deviation of the bandwidth at the top six most used APs (based on the amount of data captured by the Porcupine antenna) on day 1 and day2 respectively. The average and standard deviation of the available bandwidths differed across APs but were very similar for individual APs, with values ranging from $40 - 120$kbps. However, the maximum available bandwidth at each AP varied significantly from the average. Given that all the TCP flows considered were bandwidth hungry, this is indeed surprising.

For a random variable X with only non-negative values, following Markov's inequality holds for all a:

$$P[X \geq a] \leq \frac{E[X]}{a}$$

Making the following generous substitution

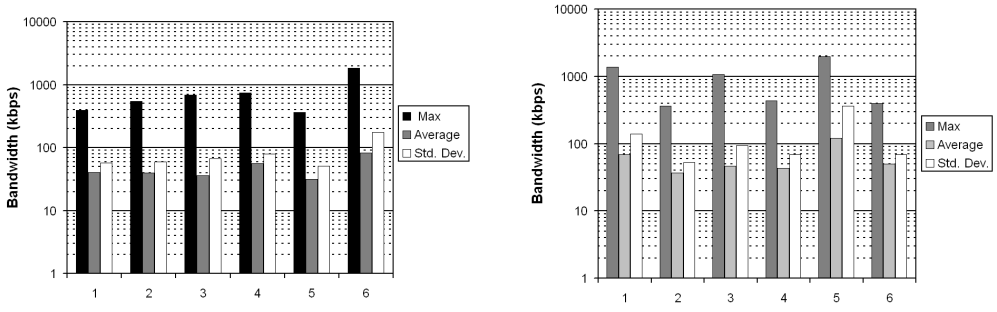
$$a = 2E[X]$$

the above gives the definition of median as shown below

$$P[X \geq 2E[X]] = \frac{1}{2}$$

Thus, using Markov's inequality we determined the median available bandwidth for the TCP flows depicted in figures 7(a) and 7(b) to be a mere 130kbps. This indicates that the performance experienced by SC2004

[5]Not to be confused with the link layer broadcast that is inherent to wireless networks.



(a) Day 1. (b) Day 2.

Fig. 7. Available bandwidth (top 6 APs).

wireless LAN users was not satisfactory because 802.11b wireless LANs could achieve a bandwidth of up to 11Mbps.

G. Mobility

As discussed in section IV-C SC2004 wireless LAN witnessed an unusually high percentage of DHCP traffic on both days. We investigate if there were many client associations and disassociations. Such a confirmation will help in establish the presence of high levels of DHCP traffic.

Figures 8(a) and 8(b) show the number of APs visited by clients on day 1 and day 2 respectively. As these figures show, the distribution of the number of APs visited per client is exponential for both days. A large number of clients (> 100 on day 1 and ≈ 80 on day 2) changed more than 5 APs over the course of the day. This observed mobility is higher than that reported by previous studies of this nature [4], [6]. Also, as shown by the variation in the standard deviation of the number of clients associated with an AP (figures 9(a) and 9(b)), the number of clients per AP varied significantly throughout the day for both days. Both of these observations indicate that the SC2004 clients tended to move, causing many associations and disassociations at APs. In the lack of information about any troubles in leasing IP addresses or the presence of rogue DHCP servers, it is hard to say whether associations and disassociations were the sole reason for the observed high DHCP traffic.

The time varying distribution of clients associated per AP (on the top 5 more used AP) is very constant, with 31 distinct clients on average in a 10 minute interval.

V. Concluding Remarks

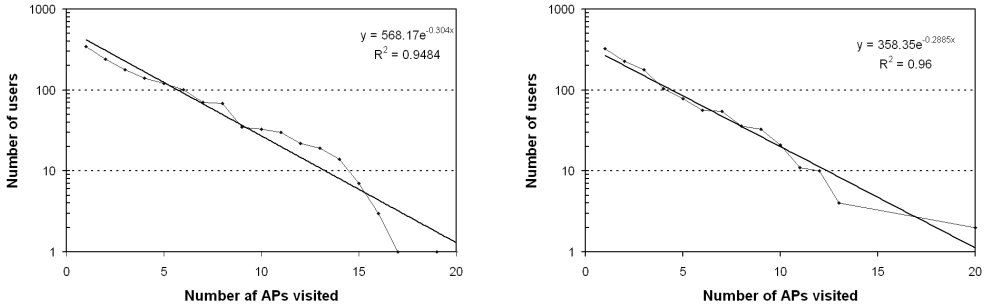
The micro-scale analysis described in this paper revealed several characteristics of SC2004 dense wireless LAN traffic that could have caused performance loss for its users. Beyond the results presented here, this research raises an important question relating to the characterization of workloads from the Web, and other wired and wireless network contexts in general. The question is: *should protocol distributions and usage trends be trusted without a micro-scale analysis confirming that application and user behavior was not altered as a result of misbehaving applications or due to misconfigurations?*. This is because TCP-based applications are subject to congestion control which may cause them to significantly reduce their sending rates in response to aggressive UDP or broadcast based protocols, or in the presence of misconfigurations which may cause undesired traffic from protocols like DHCP.

This research also illustrated a need for some form of congestion control in the Internet for non-TCP protocols. Otherwise, TCP-based applications will suffer immensely in the event the transport layer protocol mix observed in SC2004 traffic becomes a reality for the Internet.

Finally, the high channel error rates observed in our workload point to the need for a better understanding of the physical layout for dense wireless LANs. This is very important for wireless LANs set up for conferences etc. because performance problems may go unnoticed if the users decide to tolerate the inadequacies in the short term.

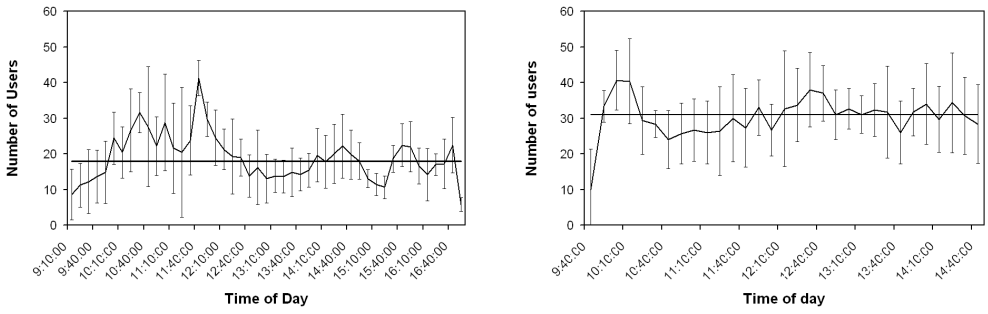
References

[1] D. Tang and M. Baker, "Analysis of a local-area wireless network," in *ACM MobiCom*, Aug. 2000.
[2] R. Hutchins and E. W. Zegura, "Measurements from a campus wireless network," in *IEEE ICC*, Apr. 2002.
[3] D. Kotz and K. Essien, "Characterizing usage of a campus-wide wireless network," in *ACM MobiCom*, Sept. 2002.
[4] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," in *ACM MobiCom*, Sept. 2004.



(a) Day 1. (b) Day 2.

Fig. 8. Client mobility.



(a) Day 1. (b) Day 2.

Fig. 9. Number of clients per AP throughout the day (aggregate data for top 5 APs).

[5] M. Balazinska and P. Castro, "Characterizing mobility and network usage in a corporate wireless local-area network," in *ACM/USENIX MobiSys*, May 2003.

[6] D. Schwab and R. Bunt, "Characterizing the use of a campus wireless network," in *IEEE INFOCOM*, Mar. 2004.

[7] F. Chinchilla, M. Lindsey, and M. Papadopouli, "Analysis of wireless information locality and association patterns in a campus," in *IEEE INFOCOM*, Mar. 2004.

[8] A. Balachandran, G. M. Voelker, P. Bahl, and P. V. Rangan, "Characterizing user behavior and network performance in a public wireless LAN," in *IEEE INFOCOM*, Mar. 2004.

[9] "Porcupine web page," `http://porcupine.iu.edu/home.html`.

[10] "fprobe Web page," `http://sourceforge.net/projects/fprobe/`.

[11] "ARGUS (Audit Record Generation and Utilization System) Web page," `http://www.qosient.com/argus/`.

[12] "Trapeze Networks web page," `http://www.trapezenetworks.com`.

[13] L. Peterson and B. Davie, *Computer Networks, A Systems Approach*, Morgan Kaufmann Publishers, 3 edition, 2003.

[14] W. Fenner, "Internet group management protocol, version 2," RFC 2236, Nov. 1997.

[15] "Gnutella Web page," `http://www.gnutella.com/`.

[16] "Kazaa Web page," `http://www.kazaa.com/`.

[17] "BitTorrent Web page," `http://www.bittorrent.com/`.