The Changing Usage of a Mature Campus-wide Wireless Network

Tristan Henderson
Dept. of Computer Science
Dartmouth College
Hanover, NH 03755, USA
tristan@cs.dartmouth.edu

David Kotz
Dept. of Computer Science
Dartmouth College
Hanover, NH 03755, USA
dfk@cs.dartmouth.edu

Ilya Abyzov
Dept. of Computer Science
Dartmouth College
Hanover, NH 03755, USA
ilyab@cs.dartmouth.edu

ABSTRACT

Wireless Local Area Networks (WLANs) are now commonplace on many academic and corporate campuses. As "Wi-Fi" technology becomes ubiquitous, it is increasingly important to understand trends in the usage of these networks.

This paper analyzes an extensive network trace from a mature 802.11 WLAN, including more than 550 access points and 7000 users over seventeen weeks. We employ several measurement techniques, including syslogs, telephone records, SNMP polling and tcpdump packet sniffing. This is the largest WLAN study to date, and the first to look at a large, mature WLAN and consider geographic mobility. We compare this trace to a trace taken after the network's initial deployment two years ago.

We found that the applications used on the WLAN changed dramatically. Initial WLAN usage was dominated by Web traffic; our new trace shows significant increases in peer-to-peer, streaming multimedia, and voice over IP (VoIP) traffic. On-campus traffic now exceeds off-campus traffic, a reversal of the situation at the WLAN's initial deployment. Our study indicates that VoIP has been used little on the wireless network thus far, and most VoIP calls are made on the wired network. Most calls last less than a minute.

We saw greater heterogeneity in the types of clients used, with more embedded wireless devices such as PDAs and mobile VoIP clients. We define a new metric for mobility, the "session diameter." We use this metric to show that embedded devices have different mobility characteristics than laptops, and travel further and roam to more access points. Overall, users were surprisingly non-mobile, with half remaining close to home about 98% of the time.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Wireless communication; C.2.3 [**Network Operations**]: Network monitoring

General Terms

Measurement

MobiCom'04, Sept. 26-Oct. 1, 2004, Philadelphia, Pennsylvania, USA.

Keywords

Wireless network, Wi-Fi, 802.11, voice, VoIP, telephony, WLAN

1. INTRODUCTION

Wireless Local Area Networks (WLANs) have become commonplace, especially on university and corporate campuses, and increasingly in public "Wi-Fi hotspots" as well. Most modern laptops are equipped with a network adapter that can access one or more types of IEEE 802.11 network, but wireless devices are rapidly diversifying to include PDAs, printers, audio players, and more. These new devices lead to changes in the way that WLANs are used. For instance, we might expect a wireless PDA to have different usage patterns than a wireless printer; a PDA might be more mobile as its user traverses a WLAN-enabled campus, whereas the printer may remain in one place to serve wireless clients.

The growing popularity of WLANs encourages the development of new applications, which may also exhibit new usage characteristics. Real-time multimedia applications, for example, have quality-of-service (QoS) requirements that may be difficult to fulfill in a shared-medium WLAN. Some of these new applications and devices may emerge simultaneously; for instance many wireless PDAs are sold equipped with streaming audio or video software.

Understanding the usage, and trends in usage, of these new devices and applications is important for providers who deploy and manage WLANs, for designers who develop new high-throughput and multimedia-friendly wireless networking standards, and for software developers who create new wireless and location-aware applications.

In this paper we study a large trace of network activity in a mature production wireless LAN. Dartmouth College has had 802.11b coverage for three years in and around nearly every building on campus, including all administrative, academic, and residential buildings, as well as most social and athletic facilities. We collected extensive trace information from the entire network throughout the Fall and Winter terms of 2003/2004.

Our work expands significantly upon previous studies. Tang and Baker [19] traced 74 computer-science clients in one building for 12 weeks. There are two more recent studies; Schwab and Bunt [18] examine 134 users over one week, and Chinchilla et al. track over 7,000 wireless cards for 11 weeks at UNC, examining web-browsing activity (for 4 weeks) and location prediction. Our earlier study, conducted at Dartmouth in 2001 [11], looked at more than 1700 users over 11 weeks. In this new 2003/04 study, we observed over 7,000 unique wireless cards using over 550 access points over the course of a 17-week trace period.

In particular, our study extends previous work by examining trends in behavior of a mature WLAN, and by examining geo-

graphic mobility within a large WLAN. We compare this 2003/4 trace to our earlier trace from Fall 2001, taken shortly after the initial installation of our campus WLAN. We found that the workload has changed significantly since 2001, and is significantly different than in other previous studies. We saw new embedded wireless devices, and new applications such as peer-to-peer services and streaming multimedia.

We next describe the environment of our study, the Dartmouth College campus, and then detail our tracing methodology in Section 3. In Section 4 we present and compare the most interesting characteristics of the data to those taken from an earlier study during the initial WLAN deployment. In Section 5 we examine three particular applications in detail: peer-to-peer file sharing, streaming media, and voice over IP. In Section 6 we analyze some of the mobility characteristics of the new devices and applications that we observed. Section 7 compares our results with those of earlier studies, and Section 8 draws overall conclusions and lists recommendations for developers and deployers of wireless network technology.

2. THE TEST ENVIRONMENT

The Dartmouth College campus has over 190 buildings on 200 acres. 476 Cisco 802.11b access points (APs) were installed in 2001 to cover most of the campus. Since then, APs have been added to increase coverage and to cover new construction, and there are currently 566 APs. The compact nature of the campus means that the signal range of interior APs extends to cover most of the campus' outdoor areas.

All APs share the same SSID, allowing wireless clients to roam seamlessly between APs. On the other hand, a building's APs are connected to the building's existing subnet. The 188 buildings with wireless coverage span 115 subnets, so clients roaming between buildings may be forced to obtain new IP addresses. (During our study, Dartmouth began to move its WLAN to a small set of separate VLANs, reducing the number of subnets).

Dartmouth College has about 5500 students and 1200 faculty, and during our study there were approximately 3200-3300 undergraduates on campus. Students are required to own a computer, and most purchase a computer through the campus computer store. Wireless laptops increasingly dominate those purchases, making up 45% of the total in 2000, 70% in 2001, 88% in 2002, and 97% in 2003. Assuming that students obtaining computers elsewhere choose laptops in the same proportion, we estimate that over 75% of the undergraduates owned laptops at the time of our study.

2.1 Voice over IP

In the summer of 2003 Dartmouth began to migrate its telephone system from a traditional analog Private Branch Exchange (PBX) to a Voice over IP (VoIP) system. A new Cisco VoIP system includes a "CallManager" call processing server, which serves to connect callers and callees, and bridge to the PBX and the local telephone company. A second, independent VoIP system by Vocera [20] serves wearable voice-controlled Wi-Fi badges; its server connects Vocera callers to other Vocera users, and bridges to the PBX, CallManager, and telephone company. Note that only our internal telephone network has migrated to IP; all off-campus calls route to the telephone company and beyond, and these other telephony providers may or may not use VoIP.

The VoIP roll-out was still underway during this study. Eventually, all undergraduates will be issued free telephony software; at the time of our study only approximately 500 licenses (for Cisco's SoftPhone) had been issued. Vocera devices are available for rent at subsidized rates. Wired and wireless Cisco VoIP phones are also available, along with a client for wireless PocketPCs.

Table 1: Devices seen on the wireless network

Guessed OS/Device	*Number of MAC addresses*	
Windows	3627	50.8%
MacOS	1838	25.8%
Unidentified	1468	20.6%
Vocera	70	0.98%
PalmOS	41	0.057%
Cisco 7920 VoIP phone	27	0.038%
Linux	27	0.038%
Dualboot Windows/Linux	24	0.034%
PocketPC	11	0.015%
Dualboot MacOS/Linux	1	0.00014%
total	7134	100.0%

2.2 Client devices

Since most students own laptops, we expected most of the devices on our WLAN to be Windows or Macintosh laptops. As the WLAN has matured and a larger variety of client devices has become available, however, we also expected to see more non-laptop devices on the network.

To determine the types of devices in use, we used the OS fingerprinting tool p0f [15] on our tcpdump traces (see Section 3 for details of our collection infrastructure) to identify the operating systems used by a given device. p0f uses differences in TCP/IP stacks and implementation flaws (e.g., timestamp values, initial window sizes, ACK values and TCP options), to derive an OS signature by scanning packet traces, much as nmap [6] and TBIT [16] do. We chose p0f for its extensive list of OS signatures.

For each card (MAC address) seen in our syslog and SNMP traces, we ran p0f on all of its TCP flows recorded by our sniffers. If all guesses for a card were the same OS (ignoring version numbers), then we assigned that OS to the card. If all guesses could run on the same CPU (e.g., Linux and Windows both run on x86), then we assumed that card was a dual-boot machine.[1] We left the card as "unidentified" if p0f guessed OSes that run on different CPUs, such as MacOS and Windows; these cards may have been used in multiple devices, or been in a host emulating another OS.

For cards that p0f could not identify, we looked at the OUI (Organizationally Unique Identifier) of the MAC address. We classified the card appropriate to the OUI if it matched an "unambiguous" vendor, i.e., one that does not sell standalone 802.11 cards that could be inserted into multiple devices. For example, Vocera is an unambiguous vendor, because the only devices with a Vocera OUI are the Vocera badges.

Table 1 shows that, unsurprisingly, Windows machines were most common, representing over 64% of the 5666 identified MAC addresses (the 1468 unknown entries include MAC addresses that we did not see on our sniffers, or for which we obtained several conflicting guesses). We also saw a large number of MacOS machines: 32% of our identifiable clients. Linux users made up a tiny proportion of our population. There were approximately 150 embedded 802.11 PDAs and VoIP devices.

3. TRACE COLLECTION

In this paper we focus on data collected during the Fall 2003 and Winter 2004 terms, a 17 week period from 2 November 2003 to 28 February 2004, inclusive.

[1]We assume that these cards represent dual-boot laptops. They could be cards that have been inserted in different machines. This distinction, however, does not affect our analysis.

We used four techniques to trace WLAN usage: syslog events, SNMP polls, network sniffers, and VoIP records.

3.1 Syslog

The APs were configured to send syslog messages to a central server whenever clients authenticated, associated, roamed, disassociated or deauthenticated. We have been continuously collecting syslogs since the installation of our WLAN in 2001.

Unfortunately we have three holes in our syslog data due to server failures. Two holes are just under four hours long, and the third is 43 hours long.

3.2 SNMP

We used the Simple Network Management Protocol (SNMP) to poll each AP every five minutes, querying AP and client-specific counters. AP-specific variables included inbound/outbound bytes, packets and errors, and the clients associated with a given AP. Client-specific variables included MAC and IP addresses, signal strength and quality.

We have two holes in our SNMP data: one week over the Christmas break, when we disabled our polls to aid network maintenance, and one day in February, where network problems on our poller caused many polls to fail (we ignore this day in our analysis).

3.3 Ethernet sniffers

We used network "sniffers" to obtain detailed network-level traces. Due to the volume of traffic on the wireless network, it was impractical to capture all the traffic. Moreover, the structure of our WLAN, with several subnets, meant that there was no convenient central point for capturing wireless traffic. Instead, we installed 18 sniffers in 14 different buildings; in some large buildings, we needed multiple sniffers to monitor all of the building's APs. The buildings were among the most popular wireless locations in 2001, and included libraries, dormitories, academic departments and social areas. In total, our 18 sniffers covered 121 APs.

Each sniffer was a Linux box with two Ethernet interfaces. One interface was used for remote access, to maintain the sniffer and to obtain the data for analysis. The other interface was used for collecting ("sniffing") data. In each of the 18 switchrooms we attached the APs to a switch, and set another port on the switch to "mirror" mode, so that all the traffic on that switch would be sent to this port. The sniffer's second interface was attached to this mirrored port. We used tcpdump to capture any wireless traffic that came through these APs and their wired interfaces. We missed any traffic between two clients associated with the same AP, as this would not be sent via the AP's wired interface, but we believe this occurred rarely.

3.4 VoIP CDR data

To understand the usage of our campus VoIP system, we configured the Cisco Call Manager server to export the details of every VoIP telephone call. These Call Detail Records (CDR) include the time and duration of the call, the caller's, callee's and final numbers (the latter represents the final reached number, e.g., if a call is diverted to voice-mail) caller and/or callee IP addresses, and reasons for call termination (e.g., a normal hang-up or a diverted call).

We have a nine-day hole at the start of our trace period due to delays in configuring the Call Manager. We lack Vocera server logs, so we have no record of Vocera calls, unless they involve a Cisco device and were logged by the CCM.

For comparison, we also look at CDR data from our analog PBX system. This data does not include on-campus calls, as these internal calls are not billed for and are thus not logged.

3.5 Definitions

One of our goals is to understand user behavior. We imagine "sessions" where a user joins the network, uses the network, possibly roams to other APs, and disconnects. We use the following definitions:

Card: A wireless NIC, identified by MAC address.[2]

Session: A session consists of an associate event, followed by zero or more roam events, and ends with a disassociate or deauthenticate event, or at the beginning of one of the holes mentioned in Section 3.1.

Active Card: A card that is involved in a session, during a given time period or at a given place.

Active AP: An AP with which one or more cards are associated, during a given time period.

Roam: A card switches APs within a session. An Associate or Reassociate message that occurs within 30 seconds after any previous event for that card is considered a roam rather than the start of a new session. (Some cards never send Reassociate messages, but only send Associate messages. It is difficult to identify precisely which of these Associate messages represent a new "session," and which are roams within the current session. We chose 30 seconds, assuming that anything shorter is not a new "session" in the eyes of the user.)

Roaming Session: A session containing roams.

Roamer Card: A card involved in one or more Roaming Sessions.

We use card-oriented definitions of "in" and "out" [11, 19]:

Inbound: Traffic sent by the AP to the card.[3]

Outbound: Traffic sent by the card to the AP.

3.6 Defining mobility

We are interested in user mobility; i.e., how often, and how far, a user moves during a session. We cannot directly measure *physical* mobility; we must infer it from users' roaming patterns. Unfortunately, roaming does not imply physical motion; we often saw cards 'ping-pong', associating and reassociating with several APs many times in succession. Although Kotz and Essien [11] define a "mobile session" as one where a card visits APs in more than one building, we found that stationary cards may ping-pong between APs in different buildings.

We define a *mobile session* to be one whose diameter is larger than a minimum size *D*. The *diameter* of a session is the maximum horizontal distance between any two APs visited during the session.[4] We used a map of the campus to determine the position of each AP.[5] Note that we consider all pairs of APs, not simply the first and last AP, because a session may wander far, only to loop back to the start by the end of the session. We cannot only consider the distance of each roam in the session, since a user may walk across campus, making short hops from AP to AP. Nor do we consider the sum of the distances of each roam in the session, because

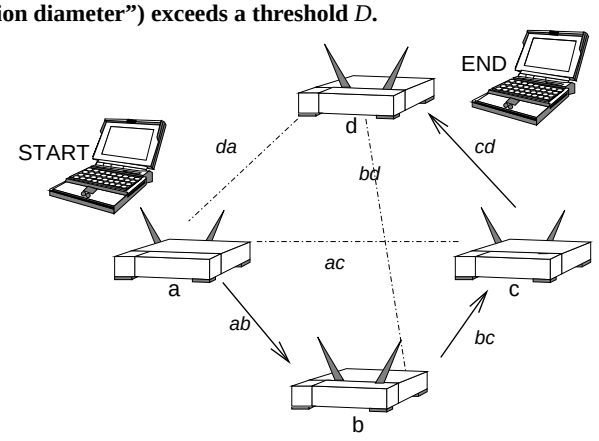
[2] Our WLAN has no MAC or IP layer authentication. Any card can associate with any AP, and obtain an IP address via DHCP. Thus we cannot identify any of the clients in our traces. We have chosen to equate a MAC address with a single user. Although some users may have multiple cards, or some cards may be shared by multiple users, we believe that this is a good approximation.

[3] If a sniffer sees a frame with a wireless source *and* destination, we counted it as "inbound," rather than double-counting it as inbound and outbound. In the SNMP data, we believe the AP counted such traffic twice. In practice, such frames were rare.

[4] We ignore the APs' altitude; our campus is relatively flat.

[5] Some APs were located off the map, e.g., off-campus student housing or athletics facilities. We ignored the few (5%) sessions that visit these APs when calculating mobility.

Figure 1: A mobile session's maximum inter-AP distance ("session diameter") exceeds a threshold D.



a stationary user can ping-pong between nearby APs many times. Figure 1 shows a session where a user starts at a, visits b and c, and ends the session associated to d. Even if ab, bc, cd and da are all shorter than D, this session is mobile if ac or bd are longer than D. Intuitively, the session diameter indicates the size of the area in which the user traveled during that session. We refer to a card that is involved in a mobile session over a given time period as a *mobile card*.

The Cisco specifications for our APs state that indoor and outdoor range at 11 Mbps is 39.6m and 244m respectively. Most APs are located indoors, although they may cover outdoor areas, so an appropriate D would be slightly greater than the indoor range. After experimentation and studying data from clients that we knew to be non-mobile, we chose $D = 50$m.

4. CHANGES

Our data collection resulted in an extremely large dataset, and it is impossible to present all of the interesting characteristics of this data in this paper. Over the 17 weeks of our trace we saw 7134 unique cards associate with an AP (Table 2). We received 32,742,757 syslog messages, conducted 16,868,747 SNMP polls and sniffed 4.6TB of data.

In this section we present some general aspects of our dataset and compare this to our Fall 2001 trace. For each figure or table, we identify the source as one or more of [syslog], [SNMP], [tcpdump] or [CDR].

We classify APs by the type of building in which they are located: 221 residential, 147 academic, 72 administrative, 59 library, 45 social and 22 athletic. Residences include dormitories, fraternities, business school and faculty housing. Social buildings include dining areas, an arts center and a museum. Athletic facilities include skating rinks, football fields, boathouses and a ski lodge.

4.1 Clients

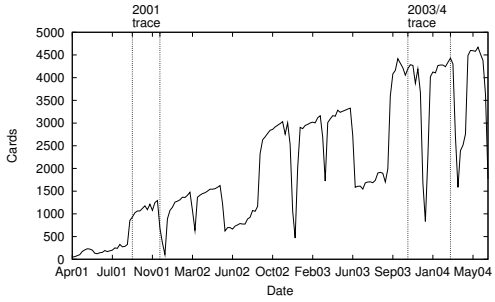
We are interested in understanding changes in the number of users on our WLAN. Has the population grown? Have usage patterns changed? Where do users visit?

The user population increased. Figure 2 shows the number of unique cards that have associated with an AP on our WLAN each week, since the installation of the network in April 2001. As each new incoming class arrives equipped with wireless laptops, and the outgoing non-wireless classes leave, the number of clients has grown steadily. The short dips represent Christmas and Spring

Table 2: Overall client observations

Total cards	7134
Peak simultaneous cards	2146
Peak simultaneous cards on an AP	91
Peak simultaneous cards in a building	193
Peak simultaneous active APs	429
Peak simultaneous active buildings	145

Figure 2: [syslog] Number of active cards per week. Note that this graph is derived from ongoing continuous data monitoring from April 2001, whereas in most of this paper we only discuss two traces from Fall 2001 and 2003/4. The vertical grid lines indicate our two trace periods.



breaks, while the longer dips are summer terms, when fewer students were on campus.

Figure 3 shows our two trace periods in further detail. The dip in Figure 3(a) in late November indicates the Thanksgiving holiday, and the two week dip in late December indicates the Christmas break, when most students and faculty were not on campus. We can again see that the population has increased dramatically. In the 2001 trace, the WLAN was still new, and consequently the population grew over time, from around 800 cards per day to 1000 cards by December 2001. In the 2003/4 trace, we saw 3000–3500 cards every day. There were slightly fewer cards in the Winter term (Jan–Feb 2004), which may reflect the smaller student population that term. In both traces, about half of the population was active on a given day.

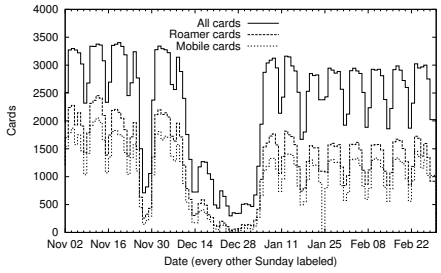
Roaming increased. The proportion of mobile and roaming cards (Figure 3) increased from approximately one-third in 2001, to one-half of the cards in 2003/4.

Usage remained diurnal. As might be expected from an academic campus where most students and some staff live on campus, we see diurnal usage patterns in Figure 4, but usage does not drop to zero during the night. These diurnal patterns have not changed significantly — we see usage peaking in the afternoon, and usage dropping from midnight to 6 a.m. The proportion of cards that remain active overnight has risen, most likely due to devices left on overnight.

The proportion of heavy users remained static. Figure 5 shows the distribution of the average time spent per day by a card on the network. This distribution is almost linear. Surprisingly, the distribution hardly changed between 2001 and 2003/4. This is confirmed by looking at a quantile-quantile plot (Figure 6). Although our user population grew significantly, the proportion of heavy users (those who spend a long time on the network each day) remained constant. Similarly, the distribution of the average number of active days per week per card has shown little change (Figure 7).

Figure 3: [syslog] Number of active, mobile, and roamer cards per day. A date's data appears to the right of its tick-mark. Note that the scales differ between 2001 and 2003/4.

(a) Fall/Winter 2003/4



(b) Fall 2001

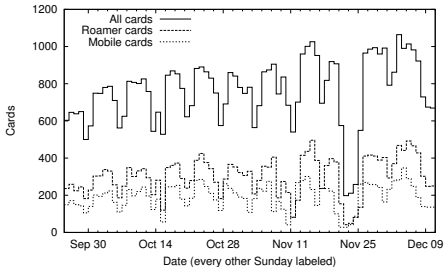
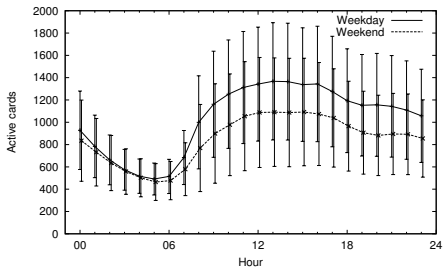


Figure 4: [syslog] Number of active cards per hour. The number of active cards for each hour of the day, separately for weekdays and weekends. The curve shows the mean, while the bars show standard deviation. The two curves are slightly offset so the bars are distinguishable.

(a) Fall/Winter 2003/4



(b) Fall 2001

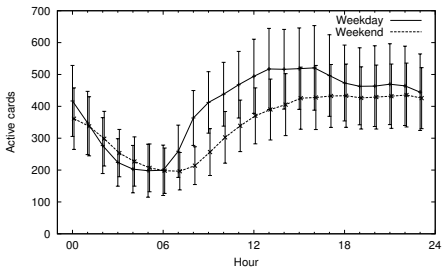
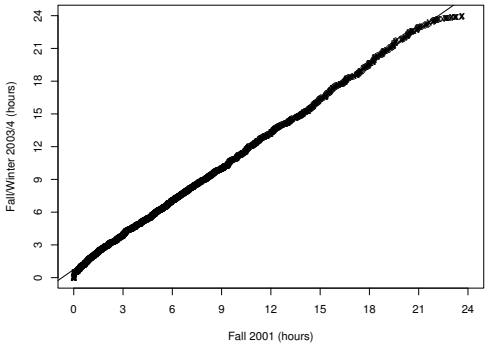


Figure 6: Quantile-Quantile plot, average time per day per user.



AP utilization increased. In Figure 8 we examine the number of APs that see a user association each day. Our network has grown from 476 APs in 2001 to 566 APs today (Figure 8(b) includes data from only 430 APs that reported syslog records). The average percentage of active APs has risen from 66.4% to 76.4%, despite the quiet Christmas break in our 2003/4 trace. Interestingly, the number of active APs during the Christmas break does not decrease by the same proportion as the number of active cards (Figure 3(a)). Many of the cards that we see during the break may have been devices that are always left turned on, and it appears that these are widely distributed across campus. The fact that the proportion of active APs has increased may indicate that the 136 new APs have been added to locations that not only lacked coverage, but locations where potential wireless users existed. Despite the increase in APs, there was a larger increase in the population of wireless users; thus, we saw a rise in the density of users on each AP: Figure 9 shows the average cards per AP in our two traces. It can be seen that the number of clients on each AP has increased markedly, and peak density in our 2001 trace is comparable to the off-peak (vacation) density in 2003/4.

Figures 10–14 illustrate the most popular locations on campus. The AP and building names have been anonymized with a name indicating the building's type, e.g., "ResBldg1" is a residential building.

The busiest types of building remained the same. We see in Figure 10 that academic buildings and libraries continued to see the largest population of cards. This result is not surprising, given that these are communal areas visited by many, if not most, students.

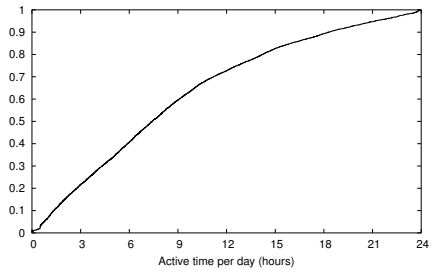
4.2 Traffic

In this section we look at traffic changes on our WLAN.

Overall traffic increased. Unsurprisingly, given the increased population, we saw an increase in the daily amount of traffic, with peaks of over 400GB in 2003/4, compared to 150–250GB in 2001 (Figure 11). Nonetheless, the average daily traffic *per active card* rose from 27.0MB in 2001 to 71.2MB in 2003/4. Today's wireless users are far more active on the network than before.

Figure 5: [Syslog] Average active time per day per user, distribution across users. Only days where a user is active on the network are considered.

(a) Fall/Winter 2003/4



(b) Fall 2001

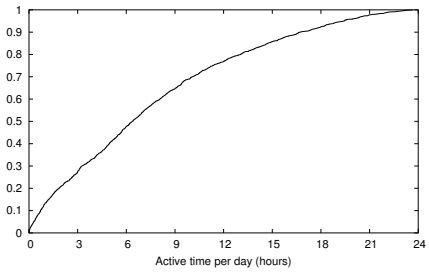
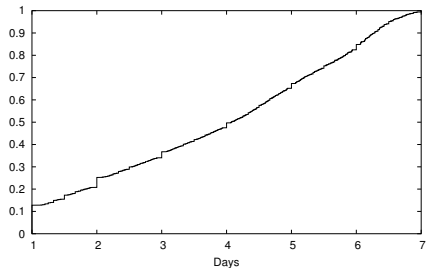


Figure 7: [Syslog] Average active days per week per user, distribution across users.

(a) Fall/Winter 2003/4



(b) Fall 2001

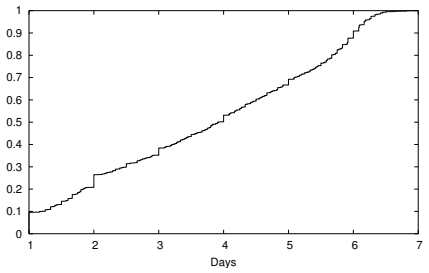
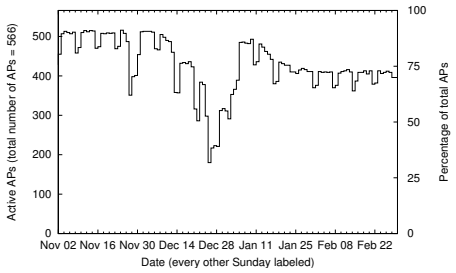


Figure 8: [syslog] Number of active APs per day. The y-axis range is from 0 to the total number of APs.

(a) Fall/Winter 2003/4



(b) Fall 2001

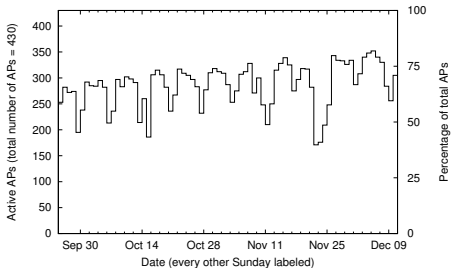
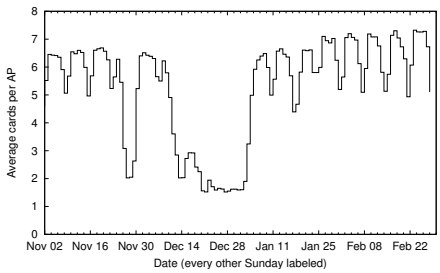


Figure 9: [syslog] Average number of active cards per active AP per day.

(a) Fall/Winter 2003/4



(b) Fall 2001

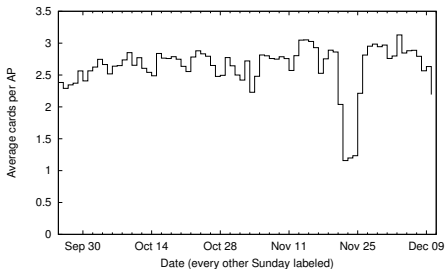
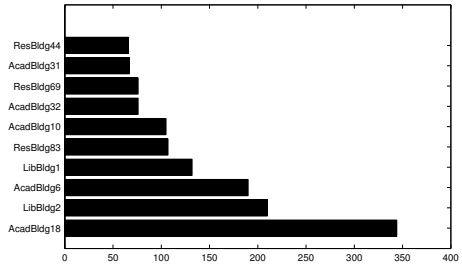


Figure 10: [syslog] Maximum cards per hour, for the busiest buildings. Ranked by their busiest hour (in number of active cards).

(a) Fall/Winter 2003/4



(b) Fall 2001

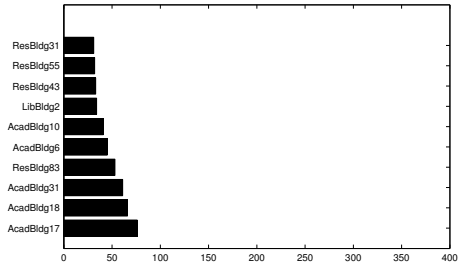
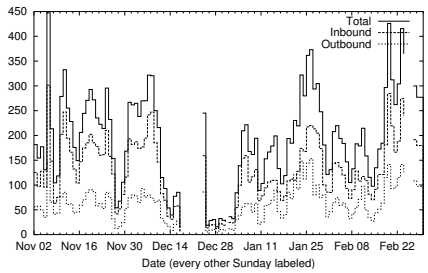
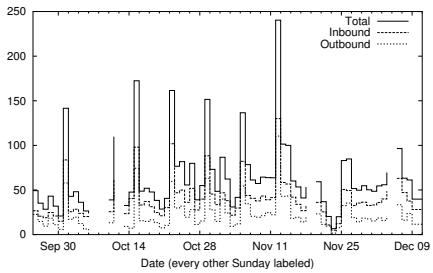


Figure 11: [SNMP] Daily traffic (GB). A date's bar appears to the right of its tick-mark. Gaps in the plot represent holes in our data.

(a) Fall/Winter 2003/4



(b) Fall 2001



We now consider the applications used on the WLAN. To identify applications, we compared the TCP or UDP port number to a customized "services" file, based on the official IANA list, but with several changes to include well-known applications that lack assigned numbers, such as games, peer-to-peer (P2P) applications and malware.

To identify Cisco VoIP traffic, which uses randomly assigned port numbers, we identified and parsed SCCP call setup packets directed to and from the CCM servers to determine the host addresses and ports for each call. We classify all UDP traffic within the Vocera port range of 5300–5400 sent to and from the central Vocera server as Vocera VoIP.

The port numbers that we saw represented thousands of applications. To summarize these, we grouped the applications by type. We based our groupings on the SLAC monitoring project [12], but with changes to reflect some of the most popular applications on campus (Table 3). Two applications are Dartmouth-specific: DND is a directory service, and BlitzMail is a popular e-mail and news client.

For those comparing this paper to our earlier work [11], note that this application classification is different than the more specific view of the data presented in our earlier work. Also, the tcpdump-based plots in the Mobicom paper were corrected by the TR and then expanded significantly in the MONET paper.

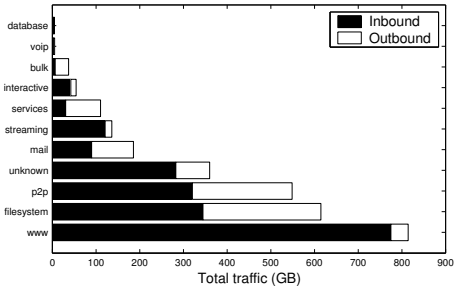
The applications used on the network changed significantly. Figure 12 shows the total amount of traffic observed to (inbound) and from (outbound) hosts on the WLAN. Note that both plots show only the traffic observed at our sniffers, which covered 121 out of 566 APs in 2003/4, and 22 out of 476 APs in 2001. Also note that Figure 12(b) does not contain a bar for VoIP, since this dataset predates the installation of the VoIP system. The proportion

Table 3: Classification of applications

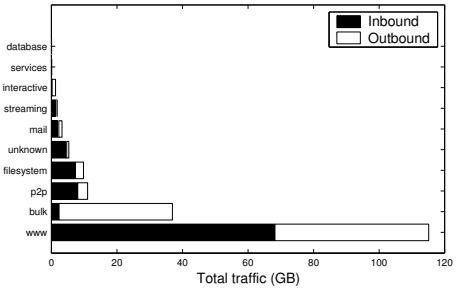
Category	*Applications*
bulk	FTP, backup
database	Oracle, PostgreSQL, SQLnet
interactive	IRC, AIM, iChat, klogin, rlogin, ssh, telnet
mail	POP, SMTP, IMAP, NNTP, BlitzMail
p2p	DirectConnect, Gnutella, Kazaa, BitTorrent, eDonkey, Napster
services	X11, DNS, finger, ident, DND, Kerberos, LDAP, NTP, printer, BOOTP, Rendezvous/ZeroConf
filesystem	SMB/CIFS, NetBIOS, AppleShare, NFS, AFS
streaming	RealAudio, QuickTime, ShoutCast, RTSP, Windows Media
voip	Cisco CallManager, SCCP, Vocera
www	HTTP, HTTPS
unknown	All unnamed and unidentified ports

Figure 12: [tcpdump] Total traffic (GB), by TCP or UDP protocol.

(a) Fall/Winter 2003/4



(b) Fall 2001



of web traffic (marked www) decreased significantly, from 62.9% of the traffic in 2001, to 28.6% in 2003/4. Three types of application saw the largest increases: P2P (from 5.2% in 2001 to 19.3% in 2003/4), filesystems (from 5.3% to 21.5%) and streaming (from 0.9% to 4.6%). We saw 5.16GB of VoIP traffic, representing 0.2% of the total traffic.

Traffic destinations changed. Figure 13 shows the proportion of near (on-campus) traffic to far (remote, off-campus) traffic. In 2001, off-campus traffic made up 64.5% of the total bytes seen on the WLAN. In 2003/4 this situation reversed, and off-campus traffic only represented 30.4% of the traffic. This reversal may be explained by the shift from a web-dominated workload in 2001 to a P2P-dominated workload in 2003/4, due to heavy local peer-to-peer usage, as we discuss in Section 5.2. This shift came very soon after the installation of our campus WLAN; we noticed it in a Spring 2002 trace [11], though without identifying the strong shift to P2P traffic.

Residences continued to generate the most traffic. Figure 14 shows the average daily traffic levels on each AP. It can be seen that the increase in traffic was not due to additional wireless coverage; as increased user population and traffic per user increased, the traffic per AP increased. We also see that residential buildings remained the most active. The ordering of the less popular categories (social, adminstrative, and athletic buildings) changed, but the majority of wireless network traffic continued to occur in residential, academic and library buildings.

5. SPECIFIC APPLICATIONS

In Section 4, we present the changes that we have seen in WLAN usage, and note significant increases in the amount of peer-to-peer and streaming multimedia traffic. In this section we analyze these applications in more detail. We begin with a look at VoIP usage.

5.1 VoIP

Our VoIP usage data came from Call Manager CDR records, which included data for both wired and wireless users. Since a SoftPhone user could be wired or wireless, depending on the user's network connection at the time of the call, we used our SNMP data to determine whether a given call was made on the wireless network. If either IP address in a CDR record was seen in an SNMP poll during the period of the call, we consider the call to be wireless.

VoIP usage mirrors general network usage. VoIP usage shows diurnal patterns (Figure 15), and these are similar to those for overall WLAN usage (Figure 4).

Figure 15: [CDR] Number of calls made by hour. The line shows the mean, and the bars show standard deviation. The values are slightly offset so that the bars are distinguishable. The wireless curve is on the bottom.

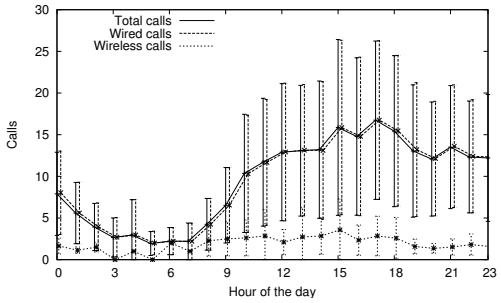


Figure 16: [CDR] Number of devices that made a call each day. The wireless curve is much smaller than the wired curve.

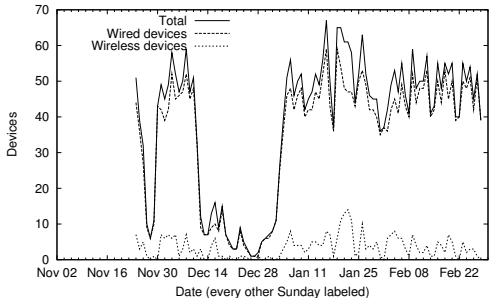
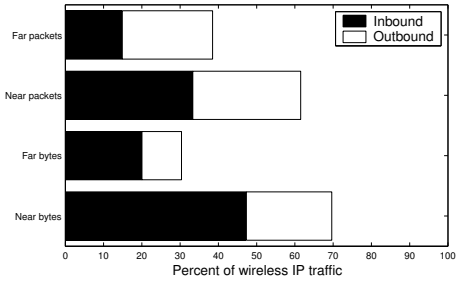


Figure 13: [tcpdump] Proportion of near and far traffic. "Near" traffic is to or from `dartmouth.edu`, all else is "Far."

(a) Fall/Winter 2003/4



(b) Fall 2001

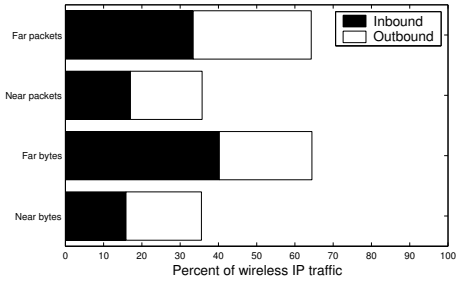
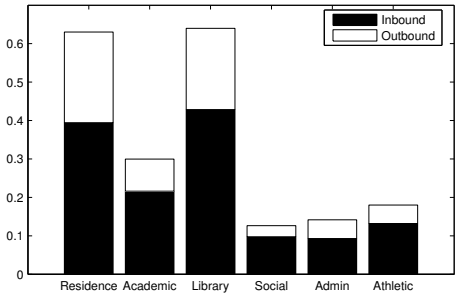


Figure 14: [SNMP] Average daily traffic per AP (GB), by category.

(a) Fall/Winter 2003/4



(b) Fall 2001

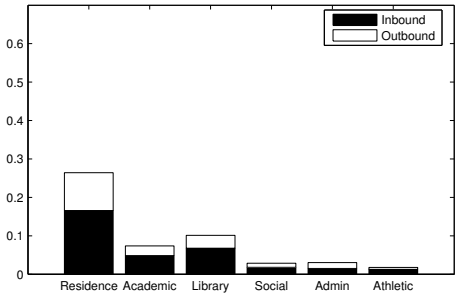


Figure 17: [CDR] log-log CCDF (Complementary Cumulative Distribution Function) of call duration. We only consider calls of duration $\geq$ 1 second and $\leq$ 6 hours.

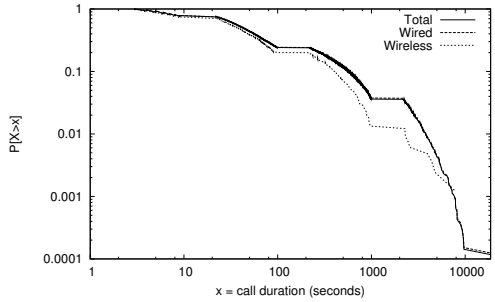
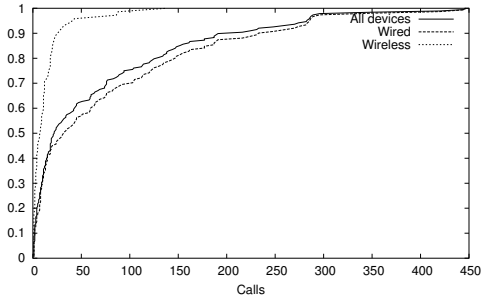


Figure 18: [CDR] CDF of the number of calls made by a VoIP device.



VoIP population was static. The number of regular VoIP users shows little growth over the course of our trace (Figure 16). We again see two dips for Thanksgiving and Christmas break. The total number of calls made each day also showed similar static levels.

VoIP users made short calls. We found that the median call duration was 41 seconds (Figure 17). For calls from wired devices, the median duration was 42 seconds, whereas for wireless devices, the median duration was 31 seconds. A Kolmogorov-Smirnov (K-S) test indicates that the difference in distributions is insignificant; VoIP calls tended to be short.

The VoIP calls are much shorter than the non-VoIP calls. The median duration of the off-campus VoIP calls[6] was 63 seconds, whereas the median duration for off-campus non-VoIP calls was 103 seconds. A K-S test indicates that non-VoIP calls are significantly longer. It is not clear why the VoIP calls, both wired and wireless, would be shorter than the PBX phone calls; the PBX population is much larger and more diverse. We hope to collect more VoIP data once the bulk of the PBX population shifts to VoIP and then we can examine this issue more deeply.

Wireless users made fewer calls. During our trace, we observed that wired devices tended to make more calls than wireless devices (Figure 18). Many wireless devices were only used once or twice, or not at all. Unfortunately, we lack detailed QoS data, but this low usage may be due to the difficulty of delivering VoIP in 802.11b networks.

VoIP calls were long-distance. Just over half of our VoIP calls, both wired and wireless, were made to long-distance destinations

[6]Our non-VoIP data only includes off-campus calls.

Table 4: VoIP calls, by destination

Destination	*Total*		*Wired*		*Wireless*	
Campus	2385	(17.6%)	2122	(16.9%)	263	(26.4%)
Local	1574	(11.6%)	1461	(11.6%)	113	(11.3%)
Regional	844	(6.2%)	759	(6.0%)	85	(8.5%)
Long-distance	7515	(55.4%)	7003	(55.7%)	512	(51.3%)
411/911	7	(0.05%)	7	(0.06%)	0	(0.00%)
Voicemail	1242	(9.2%)	1217	(9.7%)	25	(2.5%)
Total	13567	(100.0%)	12569	(100.0%)	998	(100.0%)

Figure 19: [tcpdump] Total p2p traffic (GB), by TCP or UDP protocol.

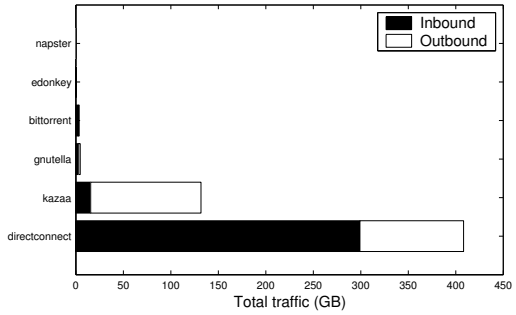
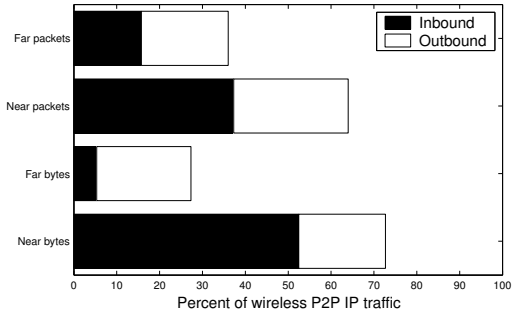


Figure 20: [tcpdump] Proportion of near and far traffic for P2P users. "Near" traffic is to/from `dartmouth.edu`.



(Table 4). Campus and local calls were the next most popular destinations. This skew may be an effect of a recent decision by our network administrators to make all domestic telephone calls free to the end-user. We also saw a high proportion of long-distance traffic in the non-VoIP calls, with 72.5% of off-campus non-VoIP calls made to long-distance destinations.

5.2 Peer-to-peer applications

Peer-to-peer (P2P) traffic increased from 5.3% of the total traffic in 2001 to 19.3% in 2003/4. The absolute increase was from 9.7GB to 548.8GB, although we had fewer sniffers installed for our 2001 trace.

In this section, we analyze the P2P file sharing that we observed on our WLAN. Note that we only consider the applications listed as "p2p" in Table 3, and not filesystems such as SMB/CIFS.

Wireless P2P users both downloaded and uploaded files. Figure 19 shows that the most popular P2P application on our WLAN was "DirectConnect". This P2P application differs from many others in that it enforces sharing: to connect to a DirectConnect "hub", a client has to be willing to offer a hub-specific amount of files to share with other users. Thus we did not see the general free-riding behavior seen in other P2P populations, where most users download files and only a few users share and upload [1]. Surprisingly, with another P2P application, Kazaa, which does not enforce sharing, we saw more outbound than inbound traffic. The reasons for this result are unclear, but it may be the presence of a packet shaper on our border router. This packet shaper limited the bandwidth for applications on certain ports, and it may have been configured to only limit Kazaa downloaders (inbound traffic).

Peer-to-peer traffic was predominantly internal; 72.7% of the wireless P2P traffic was between on-campus hosts (Figure 20). This may be due to our packet shaper. Our campus, however, is not atypical in its use of such a shaper; the Campus Computing Project [4] reports that over two-thirds of universities have some policy for limiting file transfers of audio and video files. We thus expect that this P2P behavior would be observed in many academic campus environments. The outbound remote traffic that we do see is mainly Kazaa traffic.

A few users were responsible for most of the P2P throughput. Examining the extremes of Figure 21 shows that a small number of cards send and receive a large amount of P2P data. In fact, of the 147 cards that saw more than 1MB of P2P traffic, a mere 10 cards (6.8% of the population) were responsible for over 50% of the traffic. This behavior has been observed elsewhere [17].

5.3 Streaming media

The proportion of wireless streaming audio/video traffic increased by 405% between 2001 and 2003/4, and we saw over 129GB of streaming traffic in our 2003/4 trace.

Most, but not all, streaming media was inbound. Figure 22 shows that this traffic was made up mainly of two applications: iTunes and RealAudio. Most streaming traffic was inbound: applications such as RealAudio and Quicktime are intended for large streaming media operators such as news websites, and so there tend to be a few servers, and these are rarely wireless laptops. The exception is iTunes, which allows users to easily stream music to each other. Thus we see some wireless cards sharing their iTunes music with other users, and 28% of the iTunes traffic was outbound.

Most streaming traffic was within campus. We see that most (79.6%) of the streaming traffic was to or from hosts on campus (Figure 23). This may be surprising given the number of mainstream off-campus websites that offer streaming audio and video. Within our campus, however, streaming media is used heavily for teaching, e.g., in language courses. Some of these teaching files are very large, reaching hundreds of megabytes in size, and this content may account for much of the on-campus traffic. By default, iTunes will only stream music to users on the same subnet, and hence almost all of the iTunes outbound traffic is on-campus.

Figure 21: [tcpdump] log-log CCDF of traffic per card by P2P users. Cards that saw less than 1MB are ignored.

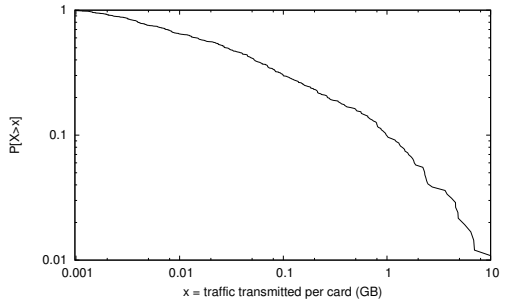


Figure 22: [tcpdump] Total streaming traffic (GB), by TCP or UDP protocol.

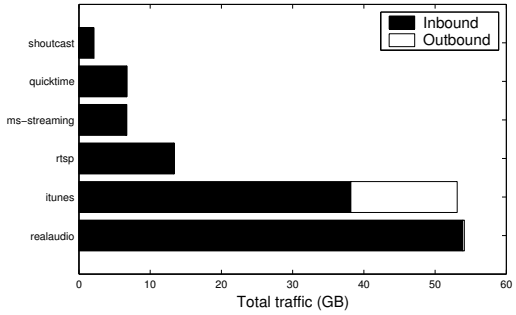


Figure 23: [tcpdump] Proportion of near and far traffic for streaming users. "Near" traffic is to/from `dartmouth.edu`.

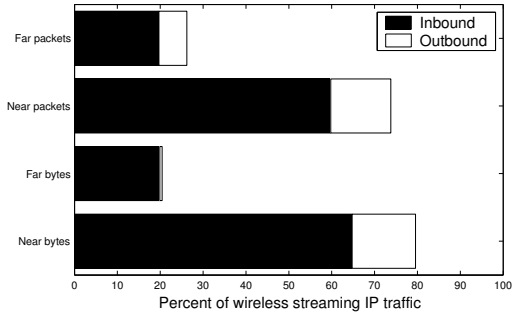
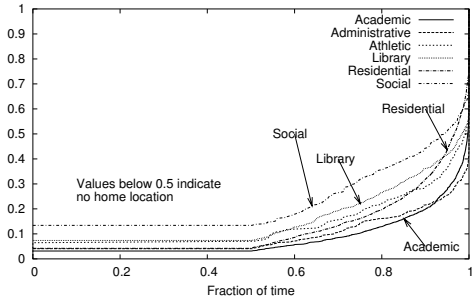


Figure 24: [syslog] Fraction of time that users spend at their home location, by the building type of their home location.



6. MOBILITY

In this section we analyze the mobility of the users in our trace. We used only the syslog records for mobility analysis, as they contain the most detailed and comprehensive record of user location.

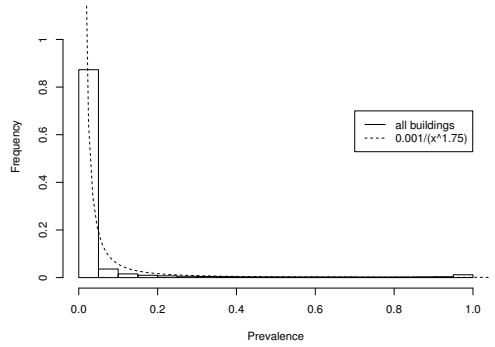
Users spent almost all their time in their home location. Figure 24 indicates the amount of time that a user spent at their "home location." We base our definition of home location on that of Balazinska and Castro [3], who choose the AP at which a client spent more than 50% of their total time on the network. We modify this definition, however, to account for our 50m session diameter. For each card, we find all the APs with which they associated over the course of our trace. Using our syslog data, we take the AP where they spend the most time associated, and consider all APs within 50m of this to represent the card's home location. Like Balazinska and Castro, we do not consider users who spend less than 50% of their time at APs in their home location, due to the difficulty of accurately determining a "home" for such users. Thus, only the right half of Figure 24 is meaningful.

We have dramatically different results than Balazinska and Castro, who found that 50% of their users spent 60% of the time in their home location. Our population is far less mobile: 95.1% of our users have a home location, and 50% of those users spend 98.7% of their time there. This striking difference was only partly due to our redefinition of "home location." If we follow Balazinska and choose just one AP as a home location, we still found that 50% of our users spend 74.0% of their time associated with a single AP. This result seems surprising, as Balazinska and Castro study a corporate campus, and one might expect higher mobility on an academic campus, with students traveling between classes. On the other hand, our trace covers residential users, who spend more time in their home location, especially if devices are left switched on overnight. Figure 24 shows that those users with a home location in a social or library building spent less time there than those with home locations in residential, academic or administrative buildings. Overnight usage is not the only reason for our low mobility, however. If we remove overnight (12 midnight to 6 am) from our data, then we still find that 50% of our users spend 69.2% of their time associated with a single AP.

Our results may also differ from the corporate data because we use syslog records, with a one-second timestamp resolution, whereas Balazinska and Castro use SNMP with a five-minute poll period. Their use of five-minute intervals led them to overestimate the time spent at a location (missing all short-term stays), and thus the two sets of results differ further.

Prevalence indicates the time that a user spends on a given AP, as a fraction of the total amount of time that they spend on the network [3]. Figure 25 again shows that our users were less mobile

Figure 25: [syslog] CDF of prevalence values for all buildings. Zero-values are discarded.



(had lower prevalence) than corporate users: the dashed line in Figure 25 represents the line of best fit for the corporate data [3]. This difference in prevalence may actually be larger, since the SNMP-collected corporate data missed short visits to APs and thus tends to overestimate prevalence.

Users persisted at a single location for longer. Another metric for demonstrating mobility is user persistence: the amount of time that a user stays associated with an AP before moving on to the next AP or leaving the network [3]. We again consider persistence using our 50m session diameter. We keep a list of all the APs that a user visits; whenever a user visits a new AP, we calculate the session diameter of this list of APs, and if the diameter is greater than 50m, we output a persistence value and clear the list.

The line in Figure 26 marked $0.92/x$ is the line of best fit from [3]. It is clear that our data is different, and that users tended to remain in a single location for longer. This difference may be due, however, to our redefinition of "location" to match our notion of a session diameter. In Figure 26 we have also calculated persistence as originally defined (the line marked "All (by AP)"). These persistence values are lower, as they include roams within a 50m diameter that may not be due to physical mobility. Nonetheless, they are still far higher than the values for corporate users; our users move less often. Moreover, since their SNMP approach tends to overestimate persistence, the fact that we saw longer persistence in our data is not an artifact of the different measurement techniques; if anything, the difference is stronger than it appears.

Different devices traveled more widely. Figure 27 shows the total number of APs visited by a device, over the course of our trace. The median number of APs visited by a user has risen from 9 in 2001 to 12 in 2003/4. In general, VoIP devices visited the largest number of APs, because these devices are "always on" and ready to receive a call. Thus a VoIP device is likely to associate with almost every AP that its owner passes, whereas a laptop will only associate with those APs where a user stops, opens the laptop and connects to the network.

A similar effect can be seen in Figure 28, which shows the session diameter for different types of devices. The always-on VoIP devices tend to travel further than laptops and PDAs.

Different devices had different session characteristics. Some of the mobility differences between devices can be attributed to the different session types for different devices. Figure 29 shows the distribution of session durations for different types of devices. As many sessions lasted almost the length of our trace period (stationary devices that were never switched off), the inset plot shows those

Figure 26: [syslog] log-log CCDF of user persistence values. We show values calculated using our session diameter metric and persistence on a per-AP basis for comparison.

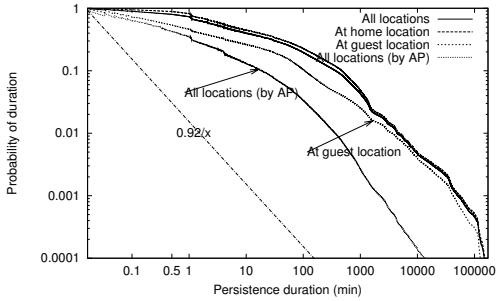


Figure 27: [syslog] CDF of the number of APs visited by a user.

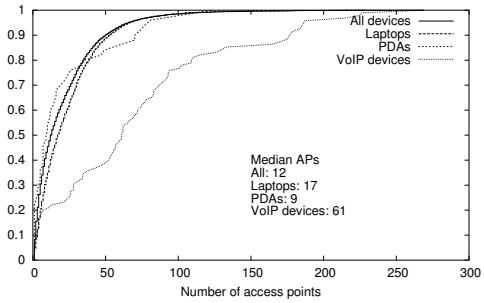


Figure 28: [syslog] Session diameter, distribution across sessions, by device. The vertical dashed line indicates 50m, our threshold for a mobile session.

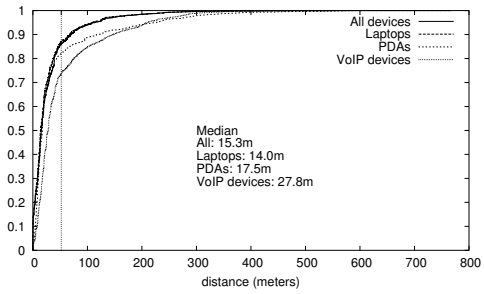


Figure 29: [syslog] Session duration, distribution across sessions, by device. The inset plot shows durations $\leq$ one hour.

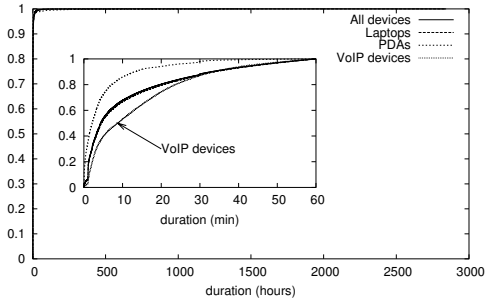
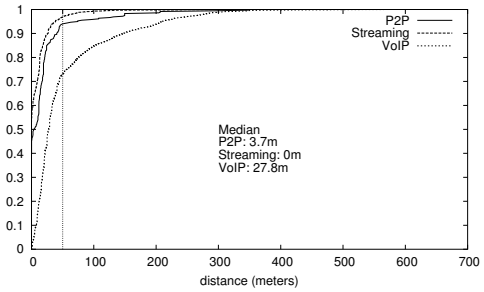


Figure 30: [syslog] Session diameter, distribution across sessions, by application. The vertical dashed line indicates 50m, our threshold for a mobile session.



durations of less than one hour for clarity. All of the device types have a short median session duration, less than 10 minutes. The short median, consistent with our earlier results, is detectable in the syslog data but would be difficult to observe with a 5-minute SNMP polling interval.

PDAs, shown in the leftmost curve, have much shorter durations than other types of devices. These short sessions are due to the way a PDA is used: kept in a pocket until needed, and switched on sporadically for short periods of time to access information. Always-on devices, however, are already becoming more common on our campus; indeed, PDAs and laptops are becoming always-on as they are used as VoIP clients. The session behavior that we show here for VoIP devices may thus be a broader indicator of future usage trends.

Different applications had different mobility characteristics. In Section 5 we focus on three of the newest wireless applications: VoIP, P2P, and streaming media. In Figure 30 we look at the distance traveled during a VoIP, P2P, or streaming session. We classify a session as containing a given application if, during that session, a host was seen by one of our sniffers, and was seen to send or receive traffic of that application category. We again see that VoIP sessions tend to travel further. Streaming sessions were less mobile than P2P sessions, perhaps because a streaming video application tends to involve active user participation, and so mobility is impeded by the need to continuously look at a device. A P2P application, however, can run in the background; a user could easily share files while moving, perhaps with a laptop left in a bag while connected to the network.

7. RELATED WORK

Our study is the largest and most comprehensive characterization of WLAN users to date. One of the earliest analyses of WLAN usage was by Tang and Baker, who use tcpdump and SNMP to trace 74 users in the Stanford CS Department over a 12-week period in 2000 [19]. While this study is similar to our own, our population is much larger and more diverse. Their top five applications (http, netbios, ftp, unknown, ssh+telnet), vary from ours, and indicate both a CS workload, and one that predates the popularity of P2P file sharing.

Balachandran et al. [2] traced 195 wireless users during the ACM SIGCOMM 2001 conference. They use SNMP to poll each of their four APs every minute. Such a small interval would have been impractical in our scenario, as it took about 90 seconds to receive SNMP responses from all of our APs. As they study a conference, user behavior is homogeneous, with clients following the conference schedule. Most sessions were short ($<$ 10 min), and longer sessions tended to be idle. About 46% of the TCP traffic was http, and 18% ssh, again indicating a CS workload.

Hutchins and Zegura used sniffers, SNMP and Kerberos authentication logs to trace 444 clients over a subset of the Georgia Tech WLAN, totaling 109 APs spread across 18 buildings, for two months in 2001 [9]. Authentication data means that they can more accurately identify sessions. As they only examine non-residential areas of campus, they find stronger diurnal usage patterns. One-third of their users do not move, although their measurements are less accurate than ours due to a 15 minute poll interval.

We have already mentioned Balazinska and Castro's study [3]. They traced 1366 corporate users on 117 APs over four weeks. They developed two metrics for mobility, *prevalence* and user *persistence*. As they used SNMP, with a five minute poll period, their data lack the precision of our syslog trace, and Section 6 shows that our results were very different.

Saroiu et al. [17] traced all HTTP and P2P traffic at the University of Washington border routers for nine days in 2002. P2P dominates, accounting for 43% of the traffic, compared to 14% for web traffic. We found slightly more web than P2P traffic, although we were examining traffic within the campus and they were examining the border. They look at both wired and wireless traffic, and a WLAN's lower throughput may have led some heavy P2P clients to use the wired network instead.

McNett and Voelker [13] install a tool on wireless PDAs, and used this to collect detailed mobility and session-level data for 272 residential users over an 11 week period on the University of California San Diego WLAN. This approach was impractical for our study, given the variety of operating systems and devices on our WLAN. They found similar session behavior to our study: mostly short sessions. As with our embedded device users, their PDA users associated with many APs.

One of the most recent campus WLAN studies comes from Schwab and Bunt at the University of Saskatchewan [18]. Their network uses a central RADIUS authentication server, allowing for accurate session determination. Their trace is significantly smaller than ours, covering 136 users on 18 APs over a one-week period. Their WLAN does not cover residential areas, and so their diurnal usage patterns differ from ours. The largest identified protocol on the wireless network was HTTP, at 28% of packets. They were unable to identify 35% of TCP packets, most of which is probably due to P2P applications (the fact that they identified only Gnutella, at 1.5%, indicates that they likely did not search for the other major P2P protocols).

Chinchilla et al analyzed WWW users on the University of North Carolina campus WLAN [5]. They tracked syslog from 222 APs and 7694 users over a 11 week period. As in our study, student residences saw the most wireless associations. Clients had fewer roams between APs, but this may have been due to lower AP density, and thus a smaller likelihood of overlapping AP coverage.

8. CONCLUSIONS AND RECOMMENDATIONS

This paper presents the results of the largest WLAN trace to date, and the first analysis of a large, mature WLAN to measure geographic mobility as well as network mobility. Most importantly, this is the first study that revisits a WLAN. We consider the changes in usage of the WLAN since its initial deployment, by re-examining usage after the WLAN has matured, and the user-base has grown beyond the early adopters. We found dramatic increases in usage, and changes in the applications and devices used on the network.

Our study has several implications for wireless network designers, network modelers, and software developers.

Although roaming increased from our previous study, our users were not very mobile, and tended to stay, or persist, at one home location for most of the time. This behavior can be exploited by network designers, for instance in the use of network caches, or prediction-based mobility schemes.

Although most users stayed predominantly in one location overall, different devices and applications had different mobility characteristics. In particular, always-on VoIP devices tend to associate with more APs and have longer-lived and farther-ranging sessions. Always-on devices are becoming more popular, and as a result WLANs will see an increase in the number of devices associated with individual APs, even though each device may not be sending or receiving large quantities of data. Designers should be conscious of this behavior, for instance, when allocating memory for association tables. Application developers may wish to consider higher levels of mobility, as it may be some time before new standards such as Mobile IP or IPv6 are widely deployed.

The higher mobility of always-on devices over laptops suggests that different devices may benefit from different policies. For example, a WLAN designer might choose to place VoIP phones and PDAs on a separate VLAN. This VLAN might be Mobile IP-enabled, or comprise one subnet that spanned an entire campus, whereas laptops could reside on building-specific subnets, on the assumption that they tend not to move around as much. This setup might also be preferable for non-mobility related reasons, such as security, since many embedded wireless devices lack software for newer wireless security standards like IEEE 802.1x.

There was a large increase in the amount of peer-to-peer traffic on our WLAN, despite the presence of a high-speed wired Ethernet network throughout our campus, and particularly in the dorms where much peer-to-peer activity takes place. Evidently the convenience of a wireless solution outweighs the limited bandwidth of an 802.11b network. As 802.11 is a shared medium, large peer-to-peer file transfers may impact other users in different ways to the wired network, and wireless-specific traffic management may be desirable. WLAN designers cannot assume that the WLAN will only be used when users are on the move, away from their home location. Instead, the WLAN has replaced the wired LAN as the primary means of network connectivity for many users.

Wireless VoIP appeared and is likely to become much more common. The wireless VoIP calls that we saw were short, with a median duration of 31 seconds, significantly shorter than calls on the old non-VoIP phone network. If such short calls are representative of typical wireless VoIP usage, this may impact the design of WLAN protocols: it may not be cost-effective to implement complex reservation schemes for such short calls.

The short VoIP calls could be a result of the difficulties of provisioning for VoIP in an 802.11b WLAN [8]; if users lack the required QoS, they may be hanging up calls in frustration. The short calls, however, were observed on both the wireless and wired network, and one would expect that our wired network is capable of handling VoIP traffic.

As well as highlighting changes between our two traces, it is important to look at those usage aspects that did not change. We found that the proportion of heavy users on our WLAN remained static, despite the shift from early adopters to a more general population. The number of hours that each client spent on the network each day was also similar between the two trace periods. This information could be useful for provisioning a WLAN. Usage remained diurnal, although given our residential campus, the diurnal variations were lower than those observed elsewhere. Residences continued to be the largest WLAN users.

Although our study is large, our results must be interpreted in context. We highlight differences in mobility between our users and previous studies of corporate users, and our academic population may not reflect activity in other venues. We believe that academic campuses are important WLAN venues, however. WLANs have been deployed at many academic institutions [4], and business surveys have started to examine academic wireless usage in addition to public usage [10]. Indeed, a university campus contains elements of an enterprise, a residential community, public hotspots (libraries), research labs, and educational workloads.

Another caveat to be considered is that our results only look at the wireless portion of our campus LAN. Some of the changes that we have observed, for example, the increase in P2P applications, may not be wireless-specific, and may have occurred on the wired LAN as well. Unfortunately it was impractical for us to measure the wired LAN due to the structure of the wired network and the quantities of data that would need to be monitored.

8.1 Future work

Our monitoring efforts are ongoing. Dartmouth College is currently in the midst of upgrading the entire WLAN to a tri-mode 802.11/a/b/g network. Soon, the campus cable television network will be migrated to an IP-based streaming video platform. As a result, we expect to see more streaming media usage on the wireless network in the future, and in particular higher-quality and higher-bandwidth video on the 802.11a network that is difficult to provide over 802.11b.

Our existing sniffers, SNMP and syslog measurement infrastructure only looks at the wired side of our wireless APs. We are currently extending our sniffing capability to include wireless sniffers, so as to monitor the 802.11 MAC layer. Whilst some researchers have taken 802.11 wireless measurements [21, 7, 14], these have typically taken place in laboratory conditions, and there is little wireless monitoring of production WLANs. As the quantity of data collected by wireless sniffing is much greater than for wired sniffing, we again intend to only monitor the most popular parts of campus. We expect, however, that this data will provide further insights into WLAN usage, and the effects of new applications on the network.

Owing to the large amount of data that was collected, we have only shown selected characteristics of the wireless traffic in this paper. There remain many questions that require further analysis of our traces. For instance, we observed high numbers of small SMB/CIFS packets involving many hosts; these are likely to be worm and virus traffic, and we are presently analyzing the effects of these worms on our WLAN.

We also welcome other researchers to make use of our data, and anonymized versions of both our 2001 and 2003/4 traces are publicly available for the community.

Acknowledgments

The authors are thankful for the help of Charles Clark and Udayan Deshpande in setting up the sniffers. We thank Kobby Essien for collecting the Fall 2001 trace.

We are grateful for the assistance of the staff of Dartmouth Computing Services, particularly Jim Baker, Craig Bisson, Steve Campbell, Robert Johnson and Brad Noblet; of Computer Science, particularly Wayne Cripps and Tim Tregubov; and of Engineering, Ted Cooley and DJ Merrill.

Finally, we thank Cisco Systems for their funding, equipment, and technical assistance. This project was funded by the Cisco Systems University Research Program.

This project was supported under Award Number 2000-DT-CX-K001 from the Office for Domestic Preparedness, U.S. Department of Homeland Security. Points of view in this document are those of the author(s) and do not necessarily represent the official position of the U.S. Department of Homeland Security.

9. REFERENCES

[1] E. Adar and B. A. Huberman. Free riding on Gnutella. *First Monday*, 5(10), Oct. 2000.

[2] A. Balachandran, G. M. Voelker, P. Bahl, and P. V. Rangan. Characterizing user behavior and network performance in a public wireless LAN. In *Proceedings of ACM SIGMETRICS 2002*, pages 195–205, Marina Del Rey, CA, June 2002.

[3] M. Balazinska and P. Castro. Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network. In *Proceedings of MobiSys 2003*, pages 303–316, San Francisco, CA, May 2003.

[4] Campus Computing Project. The 2003 National Survey of Information Technology in US Higher Education, Oct. 2003. Available online at http://www.campuscomputing.net/pdf/2003-CCP.pdf.

[5] F. Chinchilla, M. Lindsey, and M. Papadopouli. Analysis of wireless information locality and association patterns in a campus. In *Proceedings of IEEE Infocom 2004*, Hong Kong, China, Mar. 2004.

[6] Fyodor. Remote OS detection via TCP/IP stack fingerprinting. *Phrack*, 54(8), Dec. 1998. Available online at http://www.insecure.org/nmap/nmap-fingerprinting-article.html.

[7] G. Gaertner and V. Cahill. Understanding link quality in 802.11 mobile ad hoc networks. *IEEE Internet Computing*, 8(1):55–60, Jan/Feb 2004.

[8] S. Garg and M. Kappes. Can I add a VoIP call? In *Proceedings of IEEE ICC 2003*, pages 779–783, Anchorage, AK, May 2003.

[9] R. Hutchins and E. W. Zegura. Measurements from a campus wireless network. In *Proceedings of IEEE ICC 2002*, volume 5, pages 3161–3167, New York, NY, Apr. 2002.

[10] Intel Corporation. Most unwired college campuses survey, Apr. 2004. Available online at http://www.intel.com/products/mobiletechnology/unwiredcolleges.htm.

[11] D. Kotz and K. Essien. Analysis of a campus-wide wireless network. *Mobile Networks and Applications*, 2003. Accepted for publication. An earlier version appeared in ACM MobiCom 2002, and as Dartmouth College Technical Report TR2002-432.

[12] C. Logg. Characterization of the traffic between SLAC and the Internet. Technical report, Stanford Linear Accelerator Center, Menlo Park, CA, July 2003. Available online at http://www.slac.stanford.edu/comp/net/slac-netflow/html/SLAC-netflow.html.

[13] M. McNett and G. M. Voelker. Access and mobility of wireless PDA users. Technical Report CS2004-0780, Department of Computer Science and Engineering, University of California, San Diego, Feb. 2004.

[14] A. Mishra, M. Shin, and W. A. Arbaugh. An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process. *ACM SIGCOMM Computer Communication Review*, 33(2):93–102, Apr. 2003.

[15] p0f. Available online at http://lcamtuf.coredump.cx/p0f.shtml.

[16] J. Padhye and S. Floyd. On inferring TCP behavior. In *Proceedings of ACM SIGCOMM 2001*, pages 287–298, San Diego, CA, Aug. 2001.

[17] S. Saroiu, K. P. Gummadi, R. J. Dunn, S. D. Gribble, and H. M. Levy. An analysis of Internet content delivery systems. In *Proceedings of OSDI 2002*, pages 315–328, Boston, MA, Dec. 2002.

[18] D. Schwab and R. Bunt. Characterising the use of a campus wireless network. In *Proceedings of IEEE Infocom 2004*, Hong Kong, China, Mar. 2004.

[19] D. Tang and M. Baker. Analysis of a local-area wireless network. In *Proceedings of ACM MobiCom 2000*, pages 1–10, Boston, MA, Aug. 2000.

[20] Vocera. http://www.vocera.com.

[21] J. Yeo, S. Banerjee, and A. Agrawala. Measuring traffic on the wireless medium: Experience and pitfalls. Technical Report CS-TR 4421, Department of Computer Science, University of Maryland, Dec. 2002.