Dartmouth College
Dartmouth Digital Commons

Open Dartmouth: Faculty Open Access Articles

3-19-2004

The Changing Usage of a Mature Campus-wide Wireless Network

Tristan Henderson
Dartmouth College

David Kotz
Dartmouth College

Ilya Abyzov
Dartmouth College

Follow this and additional works at: https://digitalcommons.dartmouth.edu/facoa

 Part of the Computer Sciences Commons

Recommended Citation
Henderson, Tristan; Kotz, David; and Abyzov, Ilya, "The Changing Usage of a Mature Campus-wide Wireless Network" (2004). *Open Dartmouth: Faculty Open Access Articles*. 3259.
https://digitalcommons.dartmouth.edu/facoa/3259

This Article is brought to you for free and open access by Dartmouth Digital Commons. It has been accepted for inclusion in Open Dartmouth: Faculty Open Access Articles by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.

The Changing Usage of a Mature Campus-wide Wireless Network

Tristan Henderson, David Kotz, Ilya Abyzov
Department of Computer Science,
Dartmouth College, Hanover NH 03755
{tristan,dfk,ilyab}@cs.dartmouth.edu

Dartmouth Computer Science Technical Report TR2004-496
March 19, 2004

Abstract

Wireless Local Area Networks (WLANs) are now common on 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 more 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.

1 Introduction

Wireless Local Area Networks (WLANs) have become common, 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 the types of devices is rapidly expanding 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 [13] traced 74 computer-science clients in one building for 12 weeks, and a more recent study by Schwab and Bunt [12] examines 134 users over one week. The largest study to date, conducted at Dartmouth in 2001 [7], looked at more than 1700 users over 11 weeks. In this study, we observed over 7000 unique wire-

less cards 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 geographic mobility within a large WLAN for the first time. We compare this 2003/4 trace to an 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 is compact, with over 190 buildings on 200 acres. Every building is wired to the campus backbone network. Every office, dorm room, and lecture hall, and in some places every seat in a lecture hall, has wired Ethernet. In 2001, Dartmouth College installed 476 Cisco access points (APs) to provide 11 Mbps coverage to nearly the entire campus. Since then, additional APs have been added to reduce holes in coverage and to cover new construction, bringing the current number of APs to 566. Each AP has an indoor range of around 30–40m, so there are several APs in all but the smallest buildings. Although there was no specific effort to cover outdoor spaces, the compact campus means that interior APs cover most outdoor areas.

All APs share the same network name (SSID), allowing wireless clients to roam seamlessly from one AP to another. On the other hand, a building's APs are connected through a switch or hub to the building's existing subnet. The 188 buildings with wireless coverage span 115 subnets, so a wireless client roaming from one building to another will often be forced to obtain a new IP address. (During the course of this study, Dartmouth College began to move its WLAN to a small set of separate VLANs, reducing the number of subnets and the chance of roaming).

Dartmouth College has about 5500 students and 1215 full-time professors. Approximately 3345 undergraduate students lived on campus Fall 2003, and 3233 in Winter 2004. All undergraduate students are required to own a computer. Each year, approximately 1,000 undergraduate students enter Dartmouth College, and most purchase a computer through the campus computer store. 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, at least 75% of the undergraduates owned laptops at the time of our study in Fall 2003. Since summer 2001, all of the laptops that have been offered for purchase have integrated wireless support. The school of business requires all of its 280 students to own wireless laptops.

2.1 Voice over IP

In the summer of 2003 Dartmouth College 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[1] 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 VoIP is only used for internal calls; all off-campus calls route to the telephone company.

The VoIP rollout was still underway during this study. Eventually, all undergraduates will be issued with a free copy of the Cisco SoftPhone software, although during our study only approximately 500 licenses 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.

2.2 Client devices

Since most students own laptops, we expected that most of the devices on our WLAN are 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[2] on our tcpdump traces (see Section 3 for details of our collection infrastructure) to approximately identify the operating systems used by the devices associated with a given MAC address. p0f uses differences in TCP/IP stacks and implementation flaws (e.g., timestamp values, initial window sizes, ACK values

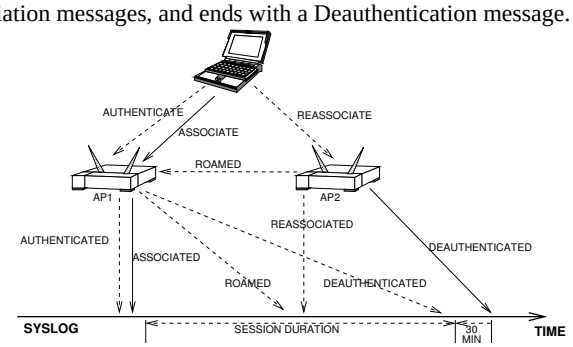
[1] http://www.vocera.com

[2] p0f is available from http://lcamtuf.coredump.cx/p0f.shtml

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%

Figure 1: A normal session begins with an Authentication message and an Association message, followed by optional Reassociation messages, and ends with a Deauthentication message.



and TCP options), to derive an OS signature by scanning packet traces, much as nmap [4] and TBIT [10] do. We chose p0f for its extensive list of OS signatures.

For each wireless card (MAC address) ever observed in our syslog and SNMP traces, we apply p0f to each of its TCP flows recorded by our sniffers. If p0f guessed a set of OSes from the same "genre" (e.g., Windows XP and Windows 2000 are both of the genre "Windows"), then we assigned that genre to that card. If p0f guessed a set of OSes that run on the same CPU (e.g., Linux and Windows both run on x86), then we declared that card to be a dual-boot machine. If p0f guessed OSes that run on different CPUs (such as MacOS and Windows), then we categorized the card as "unknown;" 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, and if this address matched an "unambiguous" vendor, we classified the card appropriate to that vendor. We define an unambiguous vendor as one that does not manufacture standalone 802.11 NICs that could be inserted into multiple devices. For instance, Vocera is an unambiguous vendor, because the only devices with a Vocera OUI are the Vocera 802.11b badges.

Table 1 shows that, unsurprisingly, Windows machines were most common, representing over 64% of the 5666 identified MAC addresses (the 1468 unknown entries represent 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 November 2 to February 28, inclusive.

At the beginning of our trace period, there were 561 APs (APs). Six more APs were installed and one AP was removed during the data collection period, bringing the final total to 566 APs.

We used five techniques to collect data about wireless network usage: syslog and DHCP events, SNMP polling, tcpdump sniffers, and VoIP CDR records.

3.1 Syslog

We configured the APs to transmit syslog messages about the comings and goings of clients. The syslog messages arrived via UDP at a server in our lab, which recorded them for later analysis. We have been continuously collecting syslog messages since the installation of our WLAN in April 2001.

Syslog messages were recorded with a timestamp and the AP name. Each message contained a client's MAC address and the type of message (Figure 1). We use the same methodology as in our previous study [7]:

Authenticated. Before a card may use the network, it must authenticate. We ignore this message.

Associated. After authentication, a card chooses an AP within range and associates with it; all traffic to and from the card goes through that AP.

Reassociated. The card monitors periodic beacons from the APs and, based on signal strength or other factors, may choose to reassociate with another AP. This feature is not present in all cards, and some cards always associate rather than reassociate (further discussion of this behaviour can be found in our previous study [7]).

Roamed. When a card reassociates with a new AP, the new AP broadcasts that fact on the Ethernet; upon receipt, the old AP emits a syslog "Roamed" message. We ignore these messages.

Disassociated. When the card no longer needs the network, it disassociates with its current AP. We found that

these messages were rare, and most disassociate messages do not report a successful disassociate, but rather indicate an error with a card, e.g., an attempt to use an AP with which they are not associated, or to use WEP (Wired Equivalent Privacy) encryption on an AP with no WEP.

Deauthenticated. While it is possible for the card to request deauthentication, this almost never happened in our log. Normally, the associated AP deauthenticates the card after 30 minutes of inactivity. For a widely roaming card, it is common to see several deauthentication messages; one from each subnet visited in the session, and we ignore all but the message from the most recent AP. We also saw deauthenticate error messages similar to those for disassociations.

Our wireless network requires no MAC or IP layer authentication. Any card can associate with any AP on campus, and request an IP address via DHCP. Due to the lack of authentication, we do not know the identity of 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 behavior is rare, and throughout this paper we use the term "card" for precision, although with the intention that cards approximate users.

Unfortunately we have three holes in our syslog and DHCP traces due to failures in the central campus syslog daemon. Two of the holes are just under four hours long, and the third is 43 hours long. For the purposes of calculating the length of a user's session from our syslog data, we end a session when we see one of these three holes.

3.2 DHCP

Both our wired and wireless network use the Dynamic Host Configuration Protocol (DHCP) to allocate IP addresses. Most wireless clients use DHCP due to the large number of subnets on the wireless network. We collected the logs of the campus central DHCP server, which allocates addresses for the entire campus network, except for the Computer Science and Engineering Departments, which maintain their own DHCP servers.

3.3 SNMP

We used the Simple Network Management Protocol (SNMP) to poll all of the APs every five minutes. Each poll retrieved values for both AP-specific and client-specific counters. The AP-specific variables included the number of inbound and outbound bytes, packets and errors, the number of client stations that were associated or had recently associated with a given AP, and the speed of the AP's wireless interface. Client-specific variables included the client's MAC and IP addresses, signal strength and quality, and the number of inbound and outbound bytes, packets and errors. All of the SNMP counters that we query have only 32 bits, and it was necessary to consider counter roll-over when analyzing the retrieved values. For most polls we consider the difference between the result of the current poll and the previous poll. For the purposes of counting traffic, however, we ignore the result in three instances: a) when the time between successful polls is more than 12 minutes (twice the polling interval plus a little slack); b) when the resulting number of bytes is more than the wireless interface could have sent or received in the time since the last poll; c) when the AP's uptime is lower than that at the previous poll (indicating a reset or a roll-over in the uptime counter). In the first case, the AP was unreachable for more than one poll, and we were unsure how many times the counter may have rolled during those missed polls. In the second and third cases, the AP (and its counters) were likely reset due to maintenance or a power failure.

Because the SNMP records contain a list of the cards associated with the AP at each poll, we combined this information with the list of MAC addresses obtained via the syslog messages. This allowed us to create a master list of all of the client MAC addresses seen on our wireless network during our trace. We used this list to identify wireless packets within the tcpdump traces.

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

3.4 Ethernet sniffers

The syslog, DHCP and SNMP logs provided us with high-level information about the number of clients, the amount of traffic and their location on the wireless network.

Due to the volume of traffic on the wireless network, it was impractical to capture all the traffic. Moreover, the structure of our network, with several subnets, meant that there was no convenient central point for capturing wireless traffic. Instead, we installed 18 network "sniffers" in 14 different buildings; in some large buildings such as the main library, we needed more than one sniffer to monitor all of the APs in that building. The 14 buildings were among the most popular wireless locations in 2001, and included libraries, dormitories, academic departments and buildings used for social activities such as eating and arts events. In total, our 18 sniffers covered 121 APs.

Each sniffer was a Linux box running tcpdump in promiscuous mode, listening to the Ethernet on the wired side of the APs. Thus we captured any wireless traffic that came through that AP and its wired interface, even traffic between APs. We could not capture any traffic between two clients associated with the same AP, as these

packets would not be sent via the AP's wired interface (this traffic is included in the SNMP counters and thus our total traffic statistics), but we believe this occurred rarely.

3.5 VoIP CDR data

To understand the usage of our campus VoIP system, we configured the Cisco CallManager 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 number, the called and final number (the latter represents the final reached number, for instance, if a call is diverted to voicemail), the name of the caller and callee as registered in the CallManager system, their IP addresses, and the reasons for call termination (e.g, a normal hang-up, a diverted call, and so forth). Unfortunately, the records contain no data about call quality.

We have a nine-day hole at the beginning of our collection period caused by delays in configuring the CallManager server. We do not have logs from the Vocera server, so we have no record of Vocera-Vocera calls, or Vocera-phone calls unless they involve a VoIP phone and were logged by the CallManager.

3.6 Definitions

One goal of this study is to understand user behavior. We imagine user "sessions" in which a user (card) joins the network, uses the network, possibly roams to other APs, and leaves the network. We use the following definitions:

Card: A wireless NIC, identified by MAC address.

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

Session: A session consists of an associate event, followed by zero or more roam events, and ends with a dissassociate or deauthenticate event (Figure 1). We interpret an Associate or Reassociate message that occurs soon after any previous event for that card to indicate a roam rather than the start of a new session.[3]

Roam: A card switches APs within a session, identified by a Reassociate message to a new AP, or by an Associate message that is treated as a roam (as described above).

We use the conventional card-oriented definitions of "in" and "out" [7, 13]:[4]

Inbound: Traffic sent by the AP to the card.

Outbound: Traffic sent by the card to the AP.

A note about the timestamps in the syslog. Although the messages may be delayed or reordered as they pass through the campus network to our server, the delays are small relative to our timestamp granularity (one second) and any reordering that affects causality should be rare. It is difficult to quantify the effect of any possible timestamp reordering, but we believe that these occurrences are rare in the context of our large trace.

3.7 Defining mobility

We are interested in the mobility of a user's session; that is, how often, and how far, a user moves during a given session. Above, we say that a card "roams" when it switches APs; we say that a session containing one or more roams is a *roaming session* and a card involved in one or more roaming sessions is a *roamer card*.

We cannot directly measure users' *physical* mobility; we must infer it from their roaming pattern. Unfortunately, roaming does not imply physical motion; we often saw cards associate with one AP, and then disassociate and reassociate with another nearby AP many times in succession. Although in previous work we have defined a "mobile session" as one in which the card visited APs in more than one building [7], we found that a stationary card may roam back and forth 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.[5] We used a detailed map of the campus to determine the position of each AP.[6] 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. Nor do we consider the distance of each roam in the session, because a user can 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 a stationary user can hop back and forth between nearby APs many times. Figure 2 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 a mobile session if ac or bd are longer than D. Intuitively, the session diameter indicates the size of the area in which the user wandered during that session. A card involved a mobile session is a *mobile card.*

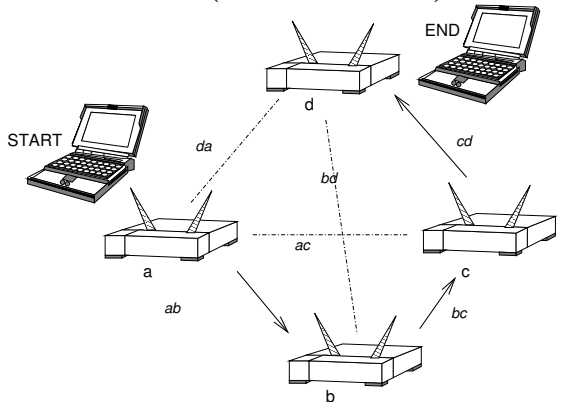
[3] "Soon" means within 30 seconds. Some cards never send Reassociate messages, but only send Associate messages. It is difficult to identify precisely which of these Associate messages should define a new "session," and which really represent a roam within the current session. We chose 30 seconds, under the assumption that anything shorter is certainly not a new "session" in the eyes of the user.

[4] 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 analysis, we believe the AP counted such traffic twice. In practice, such frames were rare.

[5] We ignore the altitude of the APs; our campus is relatively flat.

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

Figure 2: A session with many small hops may be erroneously considered non-mobile, if all the inter-AP distances are less than our threshold D. Similarly a mobile session may be considered non-mobile if the distance between the start and end of a session is shorter than D. Instead, we say a session is mobile if its maximum inter-AP distance (the "session diameter") exceeds D.



Cisco specifications for our APs indicate that indoor and outdoor range at 11 Mbps can be up to 39.6m and 244m respectively. Almost all of our campus APs are located inside buildings, although their range may extend to outdoor areas. An appropriate threshold would thus be slightly greater than the indoor range. After some manual experimentation and studying the syslog records of clients that we knew to be non-mobile, we chose D =50m.

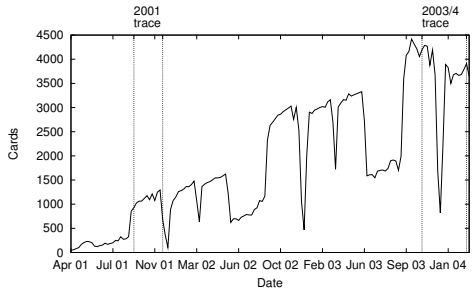
4 Changes

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. We received 32,742,757 syslog messages and 34,053,270 DHCP messages, conducted 16,868,747 SNMP polls and sniffed 4.6TB of data.

In this section we analyze some of the general characteristics of our dataset, and compare this dataset to data collected in Fall 2001, shortly after installation of the WLAN. For each figure or table, we identify the data source as one or more of [syslog], [DHCP], [SNMP], [tcpdump] or [CDR]. We first look at some of the overall changes in the client population, and then changes in the traffic mix.

We summarize the 566 APs by dividing them into six categories depending on the type of building in which they are located: 221 residential, 147 academic, 72 administrative, 59 library, 45 social and 22 athletic. The residential buildings include undergraduate dormitories, fraternities, sororities, business school and faculty housing. Social buildings include dining areas, an arts center and a museum. Athletic facilities with wireless APs include skating rinks, football fields, boathouses and a ski area.

Figure 3: [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.



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

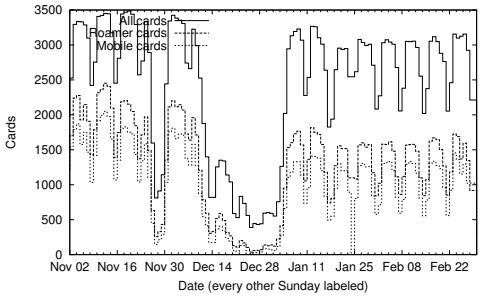
Figure 4 shows our two trace periods in further detail. The dip in Figure 4(a) in late November indicates the Thanksgiving holiday, and the two week dip in late December indicates the Christmas break, when most students and faculty were not on campus. We can again see that the population has increased dramatically. In the 2001 trace, the WLAN was still new, and consequently the population grows 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 are 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 is active on a given day.

Roaming increased. The proportion of mobile and roaming cards (Figure 4) 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 6, but usage does not drop to zero during the night. These diurnal patterns have not changed significantly— usage peaks

Figure 4: [syslog] **Number of active, mobile, and roamer cards per day.** A date's data appears to the right of its tick-mark.

(a) Fall/Winter 2003/4



(b) Fall 2001

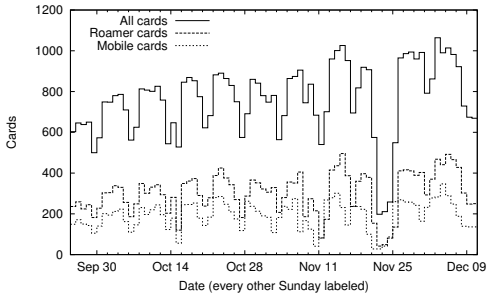
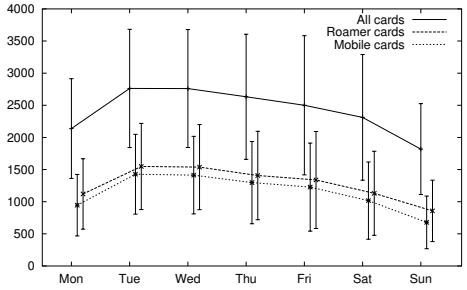


Figure 5: [syslog] **Number of active, mobile, or roamer cards per weekday.** The curve shows the mean, while the bars show standard deviation. The three curves are slightly offset so the bars are distinguishable.

(a) Fall/Winter 2003/4



(b) Fall 2001

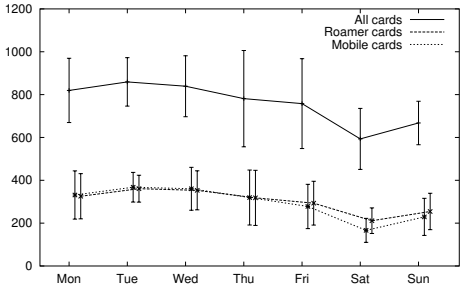
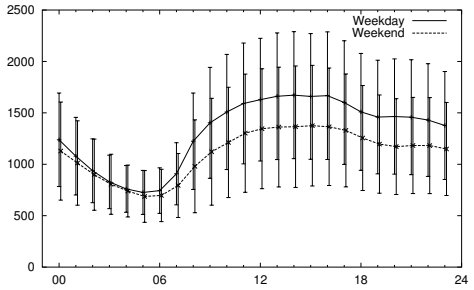


Figure 6: [syslog] **Number of active cards per hour.** The number of active cards for each hour of the day, separately for weekdays and weekends. The curve shows the mean, while the bars show standard deviation. The two curves are slightly offset so the bars are distinguishable.

(a) Fall/Winter 2003/4



(b) Fall 2001

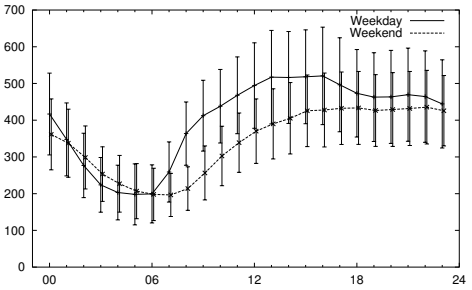
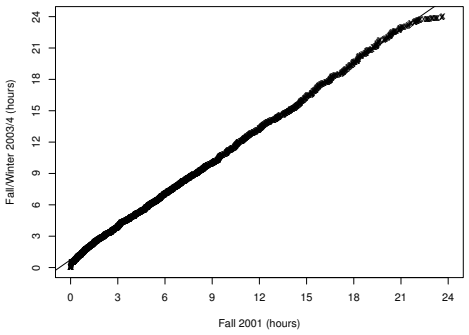


Figure 8: **Quantile-Quantile plot, average time per day per user.**



in the afternoon, and usage drops 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 7 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 8). 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 9).

AP utilization increased. In Figure 10 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 10(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 4(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. It may also be a function of the increased density of active users: in 2001 we had 476 APs serving approximately 800 clients (1.7 clients/AP), and in 2003/4 there were 576 APs for around 3500 clients (6.2 clients/AP).

Figures 12–14 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 12 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.

Residences continued to generate the most traffic. Figures 13 and 14 indicate that residential buildings remained the most active, in terms of traffic per building and per AP. 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.

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 15). The increase in traffic is proportionally smaller than the increase in clients, however, and the average daily traffic per card has fallen from 87.0MB in 2001 to 77.8MB in 2003/4. The average proportion of inbound traffic remained similar, though: 65.1% in 2001 and 68.0% in 2003/4.

We now consider the destination of traffic (on- or off-campus) and the applications used. To determine the applications, we compared the TCP or UDP port number to a customized "services" file, based on the IANA assigned list of applications, but which contains several changes and additions to include well-known applications that lack assigned numbers, such as games, peer-to-peer applications and malware.

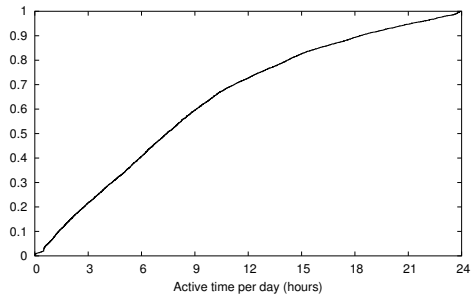
To identify Cisco VoIP traffic, which typically uses randomly assigned port numbers, we identified SCCP call setup packets directed to and from the CallManager servers, and parsed the SCCP headers in these packets to determine the host addresses and ports for each call. Since the Vocera VoIP traffic is sent via one central server, we classify all UDP traffic within the Vocera default port range of 5300 to 5400 sent to and from this server as VoIP.

The port numbers that we saw represented thousands of different applications. To better visualize and summarize the data, we grouped the applications by type. We based our groupings on the SLAC NetFlow monitoring project [8]. We modified these categories and added some of the most popular applications that we observed on campus. The categories and respective applications are displayed in Table 2. Two of the applications listed need further explanation as they are not in common use on the Internet: DND is a locally-developed directory service, and BlitzMail is a locally-developed e-mail and news application, which is heavily used by most students and staff.

The applications used on the network changed signifi-

Figure 7: [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

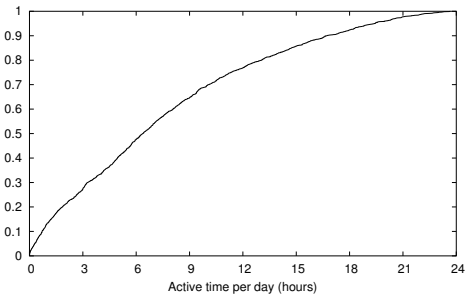
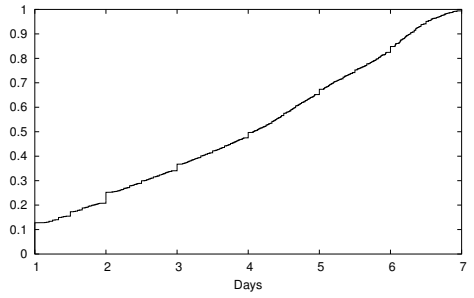


Figure 9: [Syslog] **Average active days per week per user, distribution across users.**

(a) Fall/Winter 2003/4



(b) Fall 2001

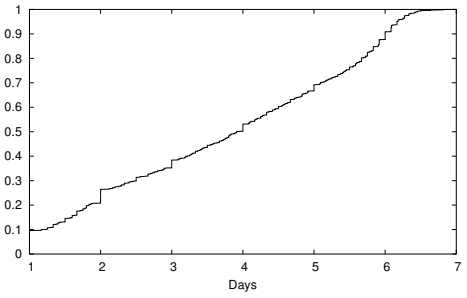
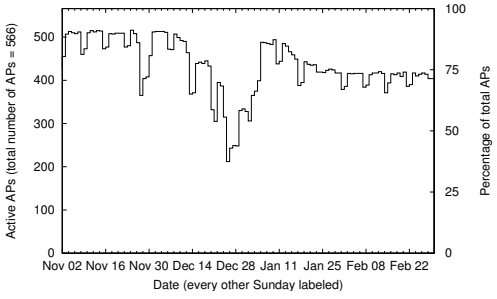


Figure 10: [syslog] **Number of active APs per day.** The y-axis range is from 0 to the total number of APs.

(a) Fall/Winter 2003/4



(b) Fall 2001

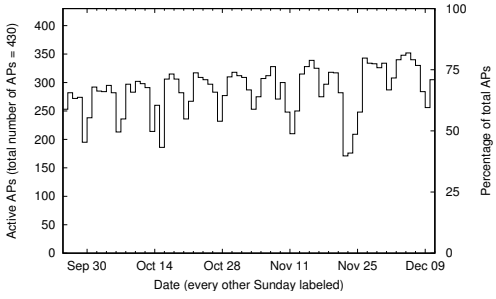
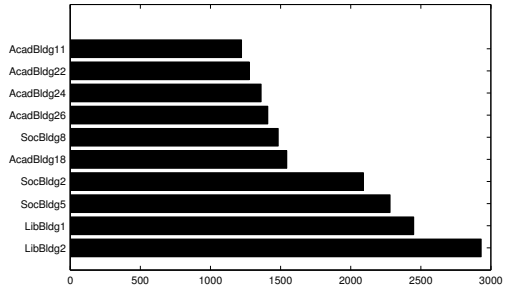


Figure 11: [syslog] **Number of active cards per building, for the ten most popular buildings.** Ranked by the number of unique cards visiting that building, over the whole trace.

(a) Fall/Winter 2003/4



(b) Fall 2001

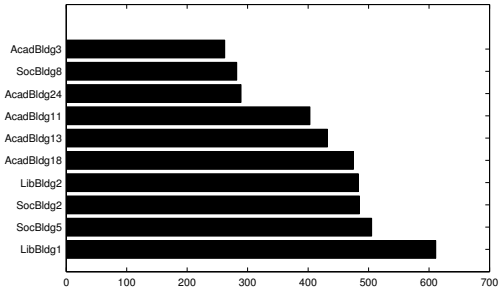
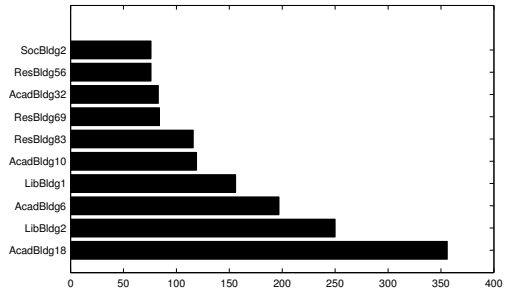


Figure 12: [syslog] **Maximum cards per hour, for the busiest buildings.** Ranked by their busiest hour (in terms of active cards).

(a) Fall/Winter 2003/4



(b) Fall 2001

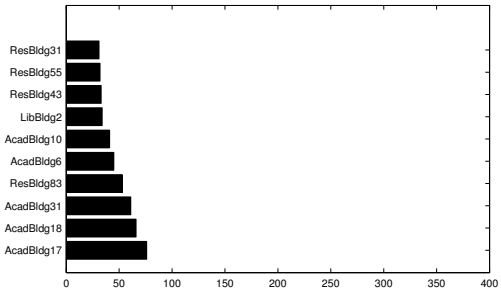
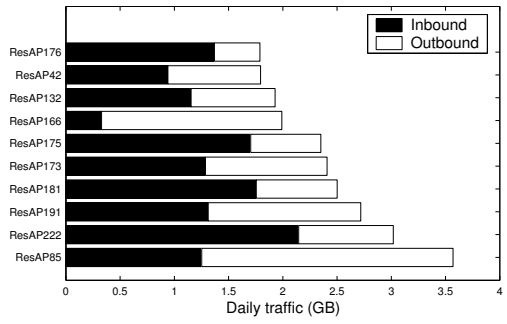


Figure 13: [SNMP] **Average daily traffic (GB), for the busiest APs.** Ranked by daily traffic.

(a) Fall/Winter 2003/4



(b) Fall 2001

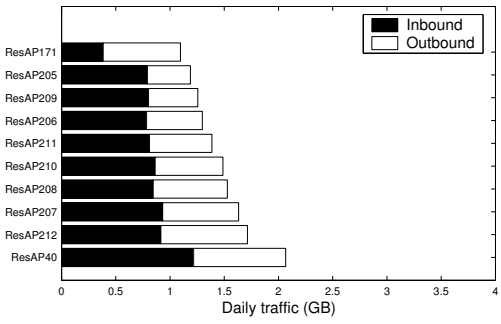
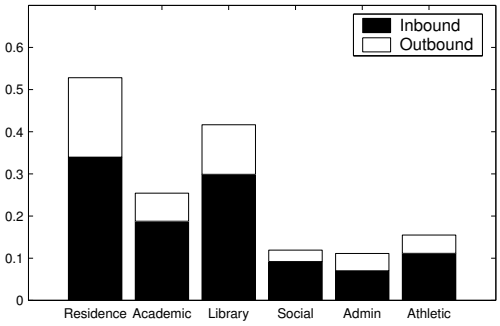


Figure 14: [SNMP] **Average daily traffic (GB), by category.**

(a) Fall/Winter 2003/4



(b) Fall 2001

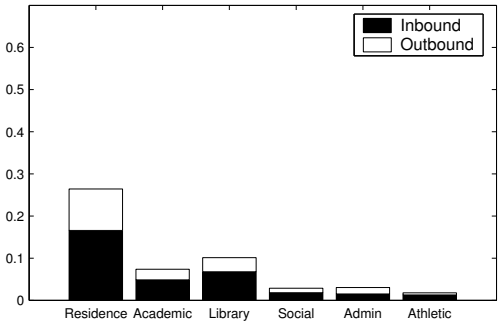
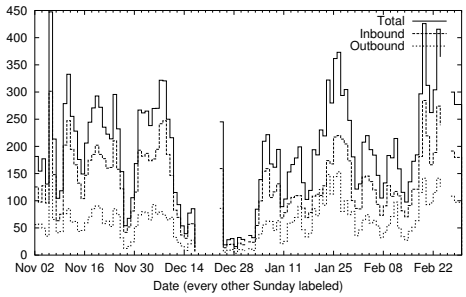


Figure 15: [SNMP] **Daily traffic (GB)**. A date's bar appears to the right of its tick-mark. Gaps in the plot represent holes in our data.

(a) Fall/Winter 2003/4



(b) Fall 2001

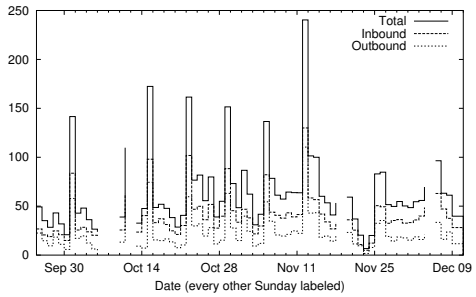
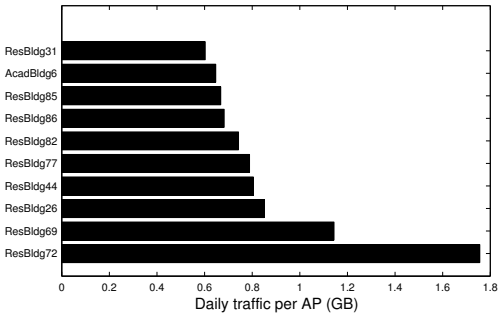


Figure 16: [SNMP] **Average daily traffic *per AP* (GB), for the busiest buildings.** Ranked by daily traffic, per AP.

(a) Fall/Winter 2003/4



(b) Fall 2001

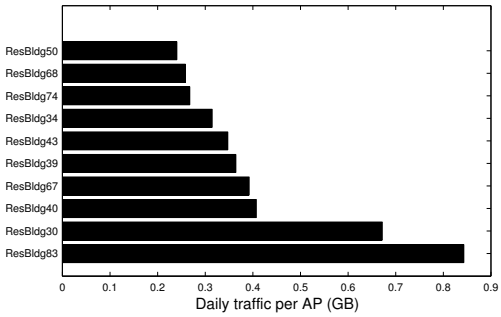
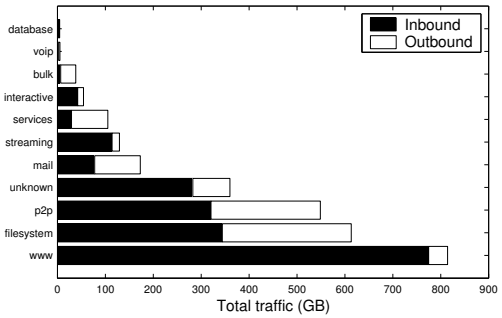


Figure 17: [tcpdump] **Total traffic (GB), by TCP or UDP protocol.** There was 4738GB traffic in total in 2003/4, and 241GB in 2001.

(a) Fall/Winter 2003/4



(b) Fall 2001

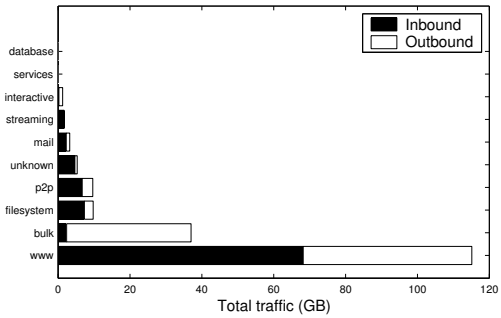


Table 2: Classification of applications

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

cantly. Figure 17 shows the total amount of traffic observed to (inbound) and from (outbound) hosts on the WLAN. Note that both plots show only 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 17(b) does not contain a bar for VoIP, since this dataset predates the installation of the VoIP system. 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 was distributed across building types. Figure 18 shows the total number of *packets* seen on our sniffers, aggregated by both application and building type. Note that there are only four building categories listed, as we did not place any sniffers in Administrative or Athletic buildings. In 2001 the bulk of the web and P2P traffic was located in residences. In 2003/4, traffic was more evenly distributed across different types of buildings. We saw almost as many filesystem packets as web packets, which is surprising since we expect that most filesystem connections involve large data files, and packets with lengths nearing an Ethernet MTU. On the other hand, our campus was hit hard by several worms, including Welchia and Blaster, during the trace period. Such worms often use the Microsoft file-sharing ports to propagate, and could be a cause of the small packets. To minimize the possibility of misclassifying legitimate Windows filesystem traffic, we do not attempt to distinguish worm traffic.

Traffic destinations changed. Figure 19 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, and was also noticed in earlier studies [7].

5 Specific applications

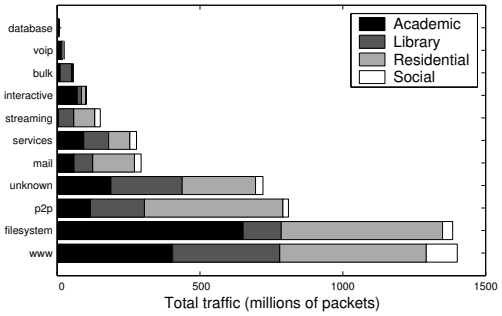
In Section 4, we present the changes that we have seen in WLAN usage, and note significant increases in the amount of peer-to-peer and streaming multimedia traffic. In this section we analyze these applications in more detail. We begin with a look at VoIP usage.

5.1 VoIP

Our primary VoIP usage data came from the CDR records provided by the CallManager, which provided data for both wired and wireless users. For Vocera users, we only had data for calls between Vocera devices and Cisco VoIP users. Vocera to Vocera calls are handled by the Vocera server itself, for which we had no logs.

Figure 18: [tcpdump] **Total traffic (packets), by TCP or UDP protocol.**

(a) Fall/Winter 2003/4



(b) Fall 2001

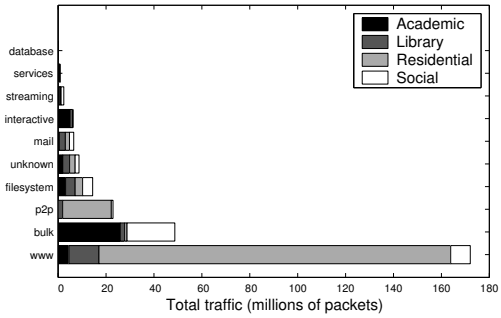
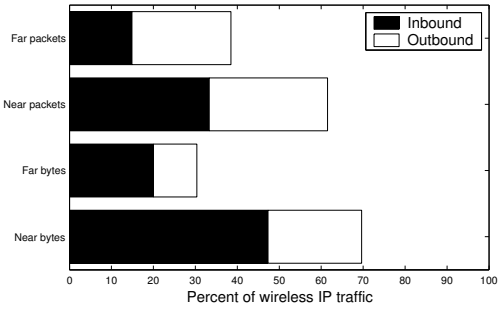


Figure 19: [tcpdump] **Proportion of near and far traffic.** "Near" traffic is to or from `hidden.edu`, all else is "Far."

(a) Fall/Winter 2003/4



(b) Fall 2001

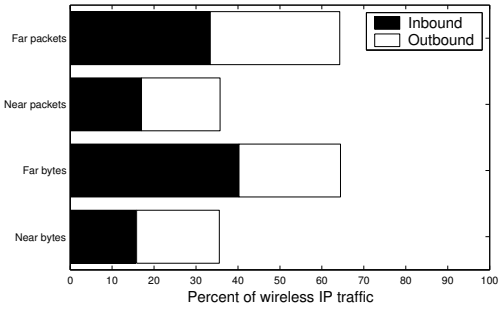


Table 3: VoIP devices

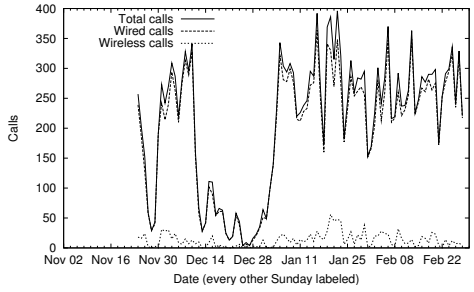
Device	*Count*
Wired Cisco VoIP telephone	80
Wireless Cisco VoIP telephone	20
Cisco SoftPhone	86
Telesym PocketPC SoftPhone	6
Total	192

As a Cisco SoftPhone user could be wired or wireless, depending on the network connection of the user's host when the call is made, we used our SNMP data to determine whether a particular call was made on the wired or wireless network. If the IP address listed in a CDR record was seen as a client in an SNMP poll during the time of this call, we consider it to be a wireless call.

Table 3 lists the number of devices that made a call during our trace period. It appears that some devices, while active on the network, made no calls at all during the trace period, and so there are discrepancies with the total number of devices listed in Table 1. We do not know the number of Vocera devices that made a call, as they all appear with the same identifier in the CallManager logs.

Figure 20 shows the number of calls made each day

Figure 20: [CDR] **Calls made by wired and wireless devices over time.** The wireless curve is much smaller than the wired curve.



over our monitoring period. We again see two dips for Thanksgiving and Christmas break.

VoIP usage mirrors general network usage. VoIP usage shows diurnal patterns (Figure 21), and these are similar to those for overall WLAN usage (Figure 6).

VoIP population was static. The number of regular VoIP users shows little growth over the course of our trace (Figure 22). 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 me-

Figure 21: [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.

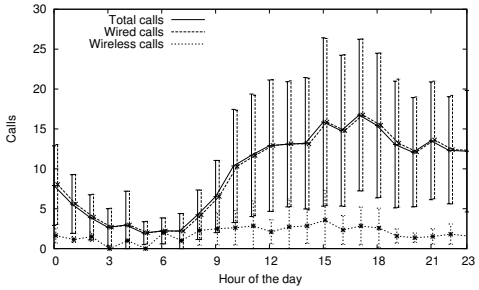


Figure 22: [CDR] **Number of devices that made a call each day.** The wireless curve is much smaller than the wired curve.

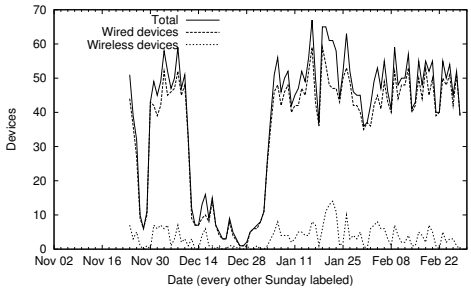


Figure 23: [CDR] **CCDF (Complementary Cumulative Distribution Function) of call duration.** We only consider calls of duration ≥ 1 second and ≤ 6 hours. Note that the axes are on a logarithmic scale.

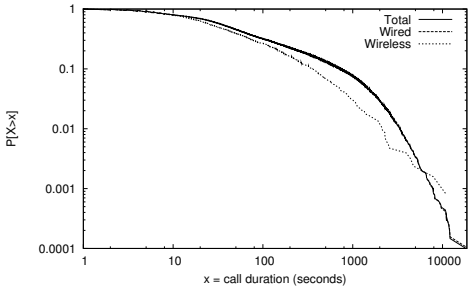
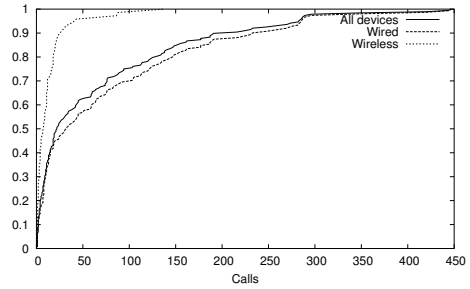


Figure 24: [CDR] **Distribution of the number of calls made by a VoIP device.**



dian call duration was 41 seconds (Figure 23). For calls from wired devices, the median duration was 42 seconds, whereas for wireless devices, the median duration was 31 seconds. A Kolmogorov-Smirnov test indicates that the difference in distributions is insignificant.

Wireless users made fewer calls. During our trace, we observed that wired devices tended to make more calls than wireless devices (Figure 24). 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 problems 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 (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 adminstrators to make all domestic telephone calls free to the end-user.

5.2 Peer-to-peer applications

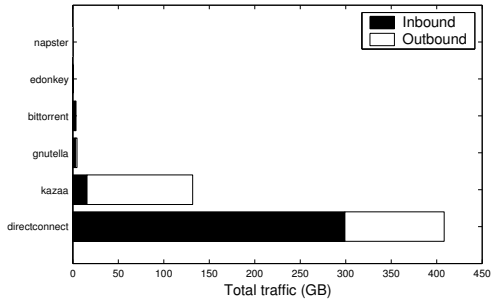
Peer-to-peer (P2P) traffic increased from 5.3% of the total traffic in 2001 to 19.3% in 2003/4. The absolute increase was from 9.7GB to 548.8GB, although we had fewer sniffers installed for our 2001 trace. We saw 0.4GB of P2P traffic per sniffer in 2001, and 4.5GB in 2003/4. In this section, we analyze the P2P file sharing that we observed on our WLAN. Note that we only consider the applications listed as "p2p" in Table 2, and not filesystems such as SMB/CIFS.

Wireless P2P users both downloaded and uploaded files. Figure 25 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

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 25: [tcpdump] **Total P2P traffic (GB), by TCP or UDP protocol.**



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 26). This situation may be specific to our campus however, due to our packet shaper. 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 27 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 [11].

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 28 shows that this traffic was made up mainly of two applications: iTunes and RealAudio. Most of the streaming traffic was inbound: applications such as RealAudio and Quicktime are intended for large streaming media operators such as news websites, and thus there tend to be a few servers, and such servers are rarely wireless lap-

Figure 26: [tcpdump] **Proportion of near and far traffic for P2P users.** "Near" traffic is to/from `hidden.edu`.

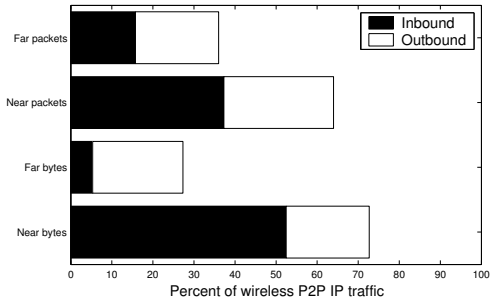


Figure 27: [tcpdump] **CCDF of traffic per card by P2P users.** Cards that saw less than 1MB are ignored. Note that the axes are on a logarithmic scale.

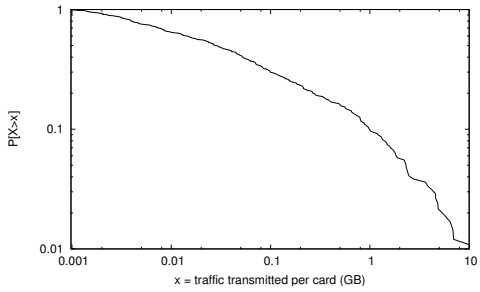
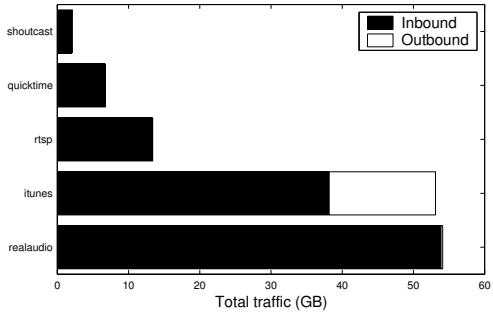


Figure 28: [tcpdump] **Total streaming traffic (GB), by TCP or UDP protocol.**



tops. The exception is iTunes, which is designed to allow iTunes users to easily stream music between each other. Thus we see that some wireless cards were sharing their iTunes music with other clients, and 28% of the iTunes traffic was outbound.

Most streaming traffic was within campus. We see that most (79.56%) of the streaming traffic was to or from hosts on campus (Figure 29). 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, 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.

6 Mobility

In this section we analyze the mobility of the users in our trace. We only used 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 30 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 31 is meaningful.

We have dramatically different results than Balazinska

Figure 29: [tcpdump] **Proportion of near and far traffic for streaming users.** "Near" traffic is to/from `hidden.edu`.

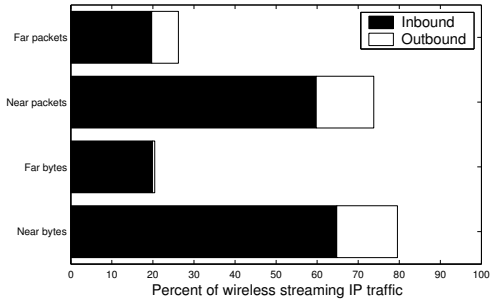


Figure 30: [syslog] **Fraction of time that users spend at their home location, by device.**

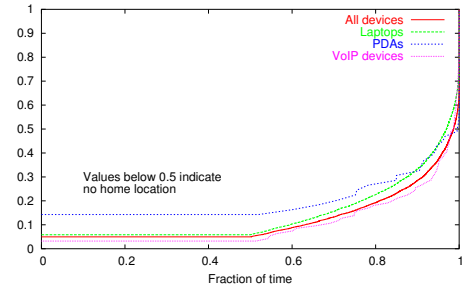


Figure 31: [syslog] **Fraction of time that users spend at their home location, by the building type of their home location.**

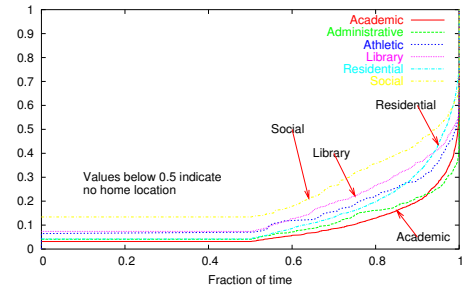


Figure 32: [syslog] **Probability distribution of prevalence values for all buildings**. Zero-values are discarded.

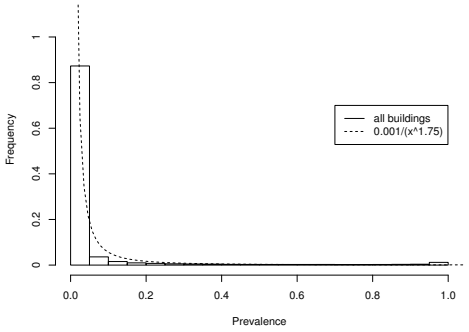
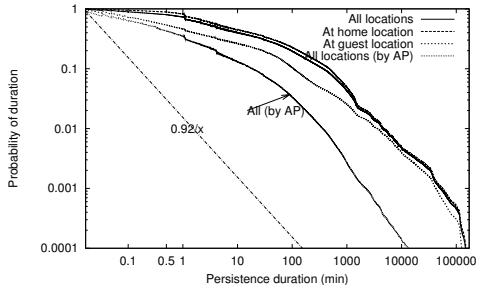


Figure 33: [syslog] **CCDF of user persistence values.** We show values calculated using our session diameter metric and persistence on a per-AP basis for comparison. Note that both axes are on a logarithmic scale.



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 31 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.

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 32 again shows that our users were less mobile (had lower prevalence) than corporate users: the dashed line in Figure 32 represents the line of best fit for the corporate data [3].

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 33 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 33 we have also calculated persistence as originally defined (the line marked "All locations (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.

Different devices traveled more widely. Figure 34 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 35, 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 36 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 durations of less than one hour for clarity. 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,

Figure 34: [syslog] **Distribution of the number of APs visited by a user.**

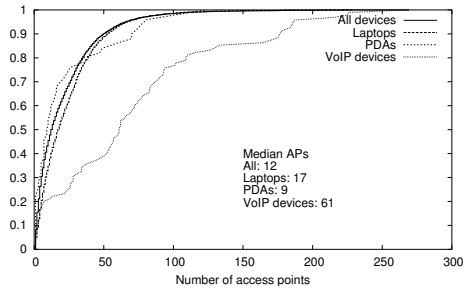


Figure 35: [syslog] **Session diameter, distribution across sessions, by device.** The vertical dashed line indicates 50m, our threshold for a mobile session.

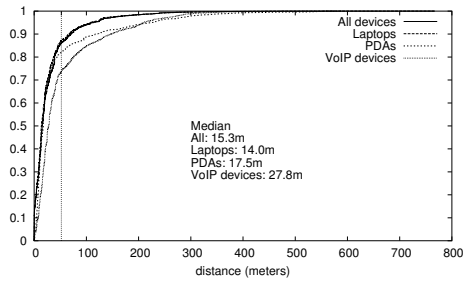


Figure 36: [syslog] **Session duration, distribution across sessions, by device.** The inset plot shows durations $\leq$ one hour.

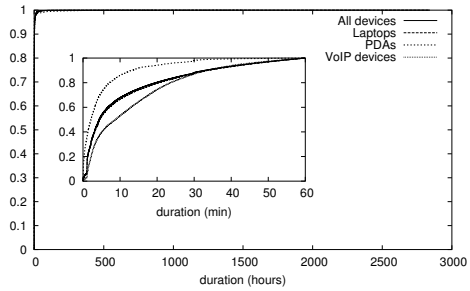
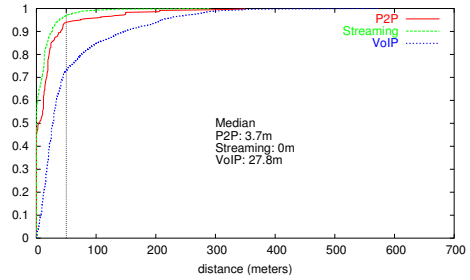


Figure 37: [syslog] **Session diameter, distribution across sessions, by application.** The vertical dashed line indicates 50m, our threshold for a mobile session.



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 37 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 large-scale analyses of wireless-network usage was conducted by Tang and Baker in 1999, who examined 24,773 users of the Ricochet network, a wireless metropolitan-area network service [14], over seven weeks. Given the nature of the data available, their analysis focuses on session behavior and client mobility, defining mobility as movement between 'poletops' (the Ricochet equivalent of APs). In 2000, they use tcpdump and SNMP to trace the activity of 74 users in the Stanford Computer Science Department over a 12-week period [13]. 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), are different from ours, and are indicative both of a CS workload, and one that predates the popularity of peer-to-peer 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 poll interval would have been impractical in our scenario, as it took approximately 90 seconds to receive SNMP responses from all 566 of our APs. Because they study a conference, user behavior is quite homogeneous, with clients following the conference schedule. Most sessions were short, less than 10 minutes, and longer sessions tended to be idle. About 46% of their TCP traffic was http, and 18% was ssh, which again indicates a CS workload.

Hutchins and Zegura used sniffing, SNMP and authentication logs to trace 444 clients over a subset of the Georgia Tech campus WLAN, totalling 109 APs spread across 18 buildings, for two months in 2001 [6]. With their Kerberos authentication data, they can identify sessions more accurately than in our study. 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 the work of Balazinska and Castro [3]. They traced 1366 users on 117 APs over four weeks on a corporate campus. They develop two new metrics for network mobility, *prevalence* and user *persistence*. As they only used SNMP, with a five minute poll period, their results lack the precision of our syslog trace, and we show in Section 6 that our results were quite different.

Saroiu et al. [11] traced all HTTP and P2P traffic at the University of Washington border routers for nine days in spring 2002. They find that peer-to-peer traffic dominates the network, accounting for 43% of the bandwidth, compared to only 14% for web traffic. We found slightly more web than P2P traffic, although their traces look at both wired and wireless traffic, and the decreased throughput available in a WLAN may have led some of our heaviest P2P clients to use the wired network instead.

In later work, Gummadi et al. look at a 200 day trace of P2P traffic [5], again from the UW border routers, and quantify the bandwidth that could be saved by locality-aware routing. We found that our P2P traffic is already mainly located on-campus. We cannot directly compare our results to the UW traces, as their data only contains off-campus traffic. Locality-aware routing, however, might be useful in a bandwidth-constrained WLAN.

One of the more recent studies of an 802.11 WLAN comes from Schwab and Bunt at the University of Saskatchewan [12]. 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 measurement period. Their WLAN does not cover residential areas, and so their diurnal usage patterns differ from ours.

Another recent study is by McNett and Voelker [9], who look at 275 PDA users over an 11-week period. They install a measuring tool on each of the PDAs, which allows them to collect detailed and accurate mobility and session-level data for each of the clients. This approach was impractical for our study, as we were unable to install software on all of the clients on our WLAN, especially given the variety of operating systems and embedded devices. They see similar session behavior to our study: mostly short sessions. They also observe that, over the course of their trace, PDA usage drops, and speculate that users may have become bored with their new devices.

All of these studies, including our own, are located on the wired side of the wireless network. That is, these studies all look at infrastructure 802.11 networks, and the monitoring takes place on the wired Ethernet into which the wireless APs have been connected. Yeo et al. [15] present some initial insights into RF monitoring, which allows the capture of all 802.11 MAC-layer traffic. They find large differences in the ability of equipment from different vendors to take effective wireless measurements.

8 Conclusions & 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. In considering the changes in usage of the WLAN since its initial deployment, 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.

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. The number of hours that each client spent on the network each day remained static between the two trace periods; this information could be useful for provisioning a WLAN.

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.

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.

Wireless VoIP appeared and is likely to become much more common. In some respects, wireless VoIP behavior mirrored wired VoIP behavior, and networks with existing wired VoIP systems may be able to use usage data from these wired systems to provision wireless VoIP systems. The wireless VoIP calls that we saw were short, with a median duration of 31 seconds. 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.

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.

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. At the same time, the campus cable television network is being 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.

We are also currently extending our sniffing capability to include wireless sniffers, so as to monitor the 802.11 MAC layer. 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. However, we expect that this data will provide further insights into WLAN usage, and the effects of new applications on the network.

Acknowledgements

The authors acknowledge the contribution of Charles Clark and Udayan Deshpande in helping to set up the sniffers. We thank Kobby Essien for conducting 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 study was funded by the Cisco Systems University Research Program.

References

[1] E. Adar and B. A. Huberman. Free riding on Gnutella. *First Monday*, 5(10), October 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 the 2002 ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems*, pages 195–205, Marina Del Rey, CA, June 2002. ACM Press.

[3] M. Balazinska and P. Castro. Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network. In *Proceedings of the 2003 International Conference on Mobile Systems, Applications, and Services (MobiSys)*, pages 303–316, San Francisco, CA, May 2003. USENIX Association.

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

[5] K. P. Gummadi, R. J. Dunn, S. Saroiu, S. D. Gribble, H. M. Levy, and J. Zahorjan. Measurement, modeling, and analysis of a peer-to-peer file-sharing workload. In *Proceedings of the 19th ACM Symposium on Operating Systems Principles (SOSP-19)*, pages 314–329, Bolton Landing, NY, October 2003.

[6] R. Hutchins and E. W. Zegura. Measurements from a campus wireless network. In *Proceedings of the IEEE International Conference on Communications (ICC)*, volume 5, pages 3161–3167, New York, NY, April 2002. IEEE Computer Society Press.

[7] 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.

[8] 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.

[9] 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, February 2004.

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

[11] 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 the 2002 Symposium on Operating Systems Design and Implementation*, pages 315–328, Boston, MA, December 2002. USENIX Association.

[12] D. Schwab and R. Bunt. Characterising the use of a campus wireless network. In *Proceedings of the 23rd Conference of the IEEE Communications Society (Infocom)*, Hong Kong, China, March 2004. IEEE.

[13] D. Tang and M. Baker. Analysis of a local-area wireless network. In *Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking*, pages 1–10, Boston, MA, August 2000. ACM Press.

[14] D. Tang and M. Baker. Analysis of a metropolitan-area wireless network. *Wireless Networks*, 8(2–3):107–120, March–May 2002.

[15] 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, December 2002.