Computer Networks 52 (2008) 2690–2712

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

The changing usage of a mature campus-wide wireless network

Tristan Henderson [a,*], David Kotz [b], Ilya Abyzov [b]

[a] School of Computer Science, University of St. Andrews, St. Andrews, Fife KY16 9SX, United Kingdom
[b] Department of Computer Science, Dartmouth College, Hanover, NH 03755, United States

ARTICLE INFO

Article history:
Received 8 August 2007
Received in revised form 5 February 2008
Accepted 16 May 2008
Available online 24 May 2008

Responsible Editor: Prof. E. Gregori

Keywords:
Wireless network
Wi-Fi
802.11
Voice
VoIP
Telephony
WLAN
Measurement

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 syslog messages, telephone records, SNMP polling and tcpdump packet captures. This is the largest WLAN study to date, and the first to look at a mature WLAN. We compare this trace to a trace taken after the network's initial deployment two years prior.

We found that the applications used on the WLAN changed dramatically, with significant increases in peer-to-peer and streaming multimedia traffic. Despite the introduction of a Voice over IP (VoIP) system that includes wireless handsets, 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.

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.

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 an IEEE 802.11 wireless network adapter, 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 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

* Corresponding author.
E-mail addresses: tristan@cs.st-andrews.ac.uk (T. Henderson), dfk@cs.dartmouth.edu (D. Kotz), ilya.abyzov@alum.dartmouth.org (I. Abyzov).
URLs: http://www.cs.st-andrews.ac.uk/~tristan (T. Henderson), http://www.cs.dartmouth.edu/~dfk (D. Kotz).

1389-1286/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2008.05.003

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. At the time, DartmouthCollege had 802.11b coverage for three years in and around nearly every building on campus. We collected extensive traces from the entire network throughout the Fall and Winter terms of 2003/2004.

Our study extends previous work in two ways. First, this is the largest study to date, with over 7000 unique wireless cards observed over 17 weeks. Second, we examine trends in behavior of a mature WLAN, and examine geographic mobility within a large WLAN. We compare this 2003/2004 trace to our earlier trace from Fall 2001, taken shortly after the initial installation of our campus WLAN. We found that the workload had changed significantly since 2001, and was significantly different than in other previous studies. We saw new embedded wireless devices, and new applications such as peer-to-peer file sharing and streaming multimedia.

The main difference between this paper and the original MobiCom conference version of this paper [12] is the additional data analysis in this version. In particular, we extend our analysis of application usage by considering usage at home and non-home locations in Section 6. We have also improved our method for detecting peer-to-peer file sharing applications, which has led to different results in Section 5.

We next describe the environment of our study, and then detail our methodology in Section 3. In Section 4, we compare the most interesting characteristics of the data to those taken from an earlier study. Section 5 examines 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 discusses related work, and Section 8 draws conclusions and lists recommendations for WLAN developers and deployers.

2. The test environment

The Dartmouth College campus has over 190 buildings on 200 acres. In 2001, Dartmouth installed 476 Cisco 802.11b access points (APs) to cover most of the campus. Since then, APs have been added to increase coverage and to cover new construction, and at the time of this study, there were 566 APs. The compact nature of the campus means that the range of indoor APs covers most 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 roaming clients 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 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. In 2008, at the time of this writing, the number is close to 100%.

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 included a "CallManager" call processing server, which served to connect callers and callees, and bridge to the PBX and the local telephone company. A second, independent VoIP system by Vocera [28] served wearable voice-controlled Wi-Fi badges; its server connected Vocera callers to other Vocera users, and bridged 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 not use VoIP.

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

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 [21] 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 [9] and TBIT [22] 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 ran on different CPUs,

[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.

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 cards (the unknown entries include cards that we did not see on our sniffers, or for which we obtained 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 wireless 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. 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 syslog messages 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 4 h long, and the third is 43 h long.

3.2. SNMP

We used the Simple Network Management Protocol (SNMP) to poll each AP every 5 min, querying AP and client-specific counters. AP-specific variables included in/outbound bytes, packets and errors. Client-specific variables included MAC and IP addresses, and signal strength.

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	0.57%
Cisco 7920 VoIP phone	27	0.38%
Linux	27	0.38%
Dualboot Windows/Linux	24	0.34%
PocketPC	11	0.15%
Dualboot MacOS/Linux	1	0.0014%
Total	7134	100.0%

We have two SNMP holes: one week over the Christmas break, when we disabled our polls to aid network maintenance, and one day in February, where network problems 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 WLAN, it was impractical to capture all the traffic. Moreover, the network topology, 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, the sniffers covered 121 APs.

Each sniffer was a Linux box with two Ethernet interfaces. One interface was used for remote access 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" all of the traffic on that switch. The sniffer's second interface was attached to this mirrored port. We used tcpdump to capture the first 200 bytes of any Ethernet frame 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 was rare. We logged the standard error messages from tcpdump and do not believe that tcpdump dropped any frames. We did observe some errors, however, as discussed in Section 4.2.

3.4. VoIP CDR data

To record usage of our VoIP system, we configured the Call Manager server to export the details of every VoIP 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 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 Call Manager.

For comparison, we also look at CDR data from our analog PBX system. These data do 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. Associate or Reassociate messages occurring within 30 s after any previous event for that card are considered roams rather than the start of a new session. (Some cards only ever associate. It is hard to identify which of these Associate messages represent a new "session", and which are roams within the current session. In a preliminary analysis of the data we found that 30 s was an appropriate, though approximate, cut-off time to separate reassociations from new sessions.)
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" [15,27]:

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; how often, and how far, a user moves during a session. We cannot directly measure *physical* mobility; we must infer it from 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 [15] 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 located 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 a stationary user can ping-pong between nearby APs many times. Fig. 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 specifications for our APs state that indoor and outdoor range at 11 Mbps is 39.6 m and 244 m, respectively. Most APs are located indoors, although they may cover outdoor areas, so an appropriate D would be slightly greater than the indoor range. Moreover, AP range can vary, given obstacles and mobile users or other sources of interference. We studied data from clients that we knew to be non-mobile, and 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 in this paper. Over the 17 weeks of our trace we saw 7134 unique cards (Table 2). We received 32,742,757 syslog messages, conducted 16,868,747 SNMP polls and sniffed 4.6 TB of traffic.

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, and faculty housing. Social buildings include dining areas, an arts center and a museum. Athletic facilities in-

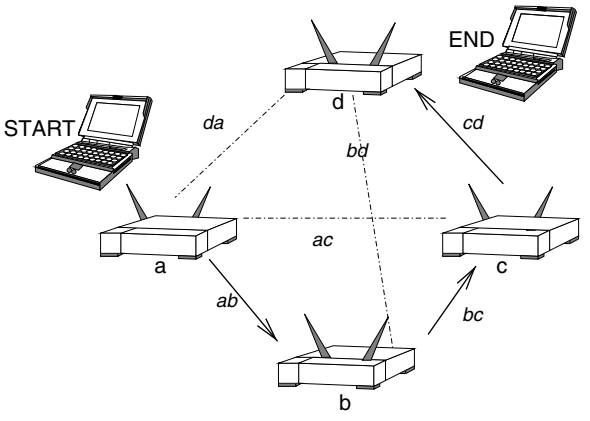


Fig. 1. A mobile session's maximum inter-AP distance ("session diameter") exceeds a threshold D.

[2] Our WLAN had 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. We ignored the few (5%) sessions that visit these APs when calculating mobility.

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

clude 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. Fig. 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 of students 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 were on campus.

Fig. 3 shows our two trace periods in further detail. The late November dip in Fig. 3.1 indicates Thanksgiving, and the dip in late December indicates Christmas, when most students and faculty were not on campus. We can again see that the population has increased dramatically. In 2001, 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 2003/2004 trace, we saw 3000–3500 cards every weekday. There were slightly fewer cards in the Winter term (January–February 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 (Fig. 3) increased from approximately one-third in 2001, to one-half of the cards in 2003/2004. This plot also shows that most cards that roamed were also involved in a mobile session, that is, at least one session that day had diameter over 50 m.

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 Fig. 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 0600. 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. Fig. 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 2003/2004. This is confirmed by looking at a quantile–quantile plot (Fig. 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 shows little change (Fig. 7).

AP utilization increased. In Fig. 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 in 2003/2004 (Fig. 8.2 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 2003/2004 trace. Interestingly, the number of active APs during the Christmas break does not decrease by the same proportion as the number of active cards (Fig. 3.1). 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

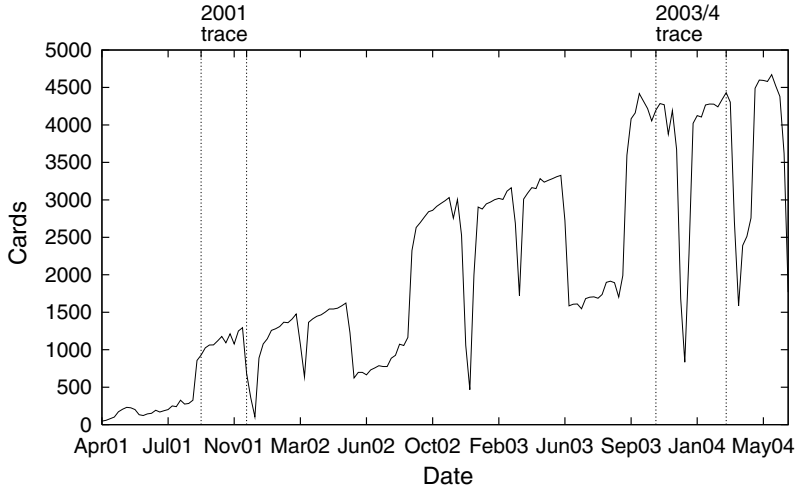
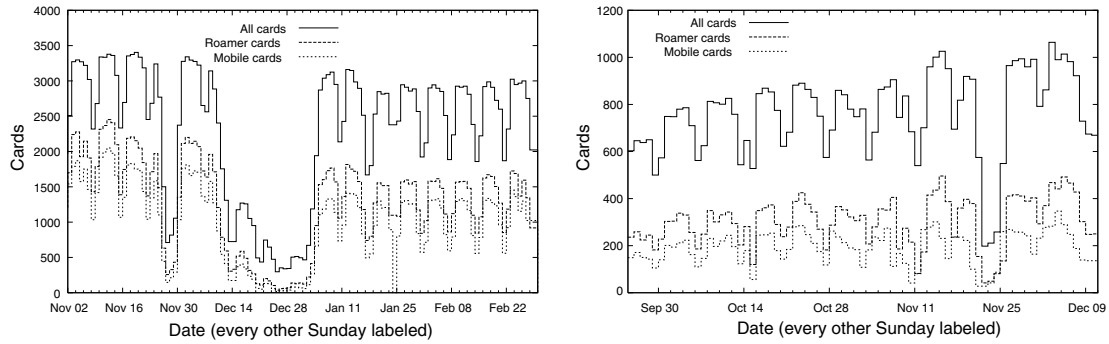
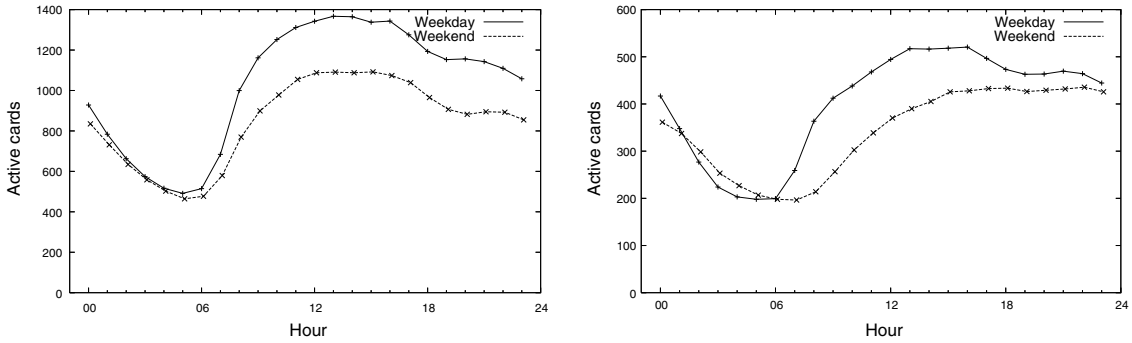


Fig. 2. [syslog] Number of active cards per week. Note that this graph is derived from ongoing continuous data monitoring from April 2001, whereas in most of this paper we only discuss two traces from Fall 2001 and 2003/2004. The vertical grid lines indicate our two trace periods.



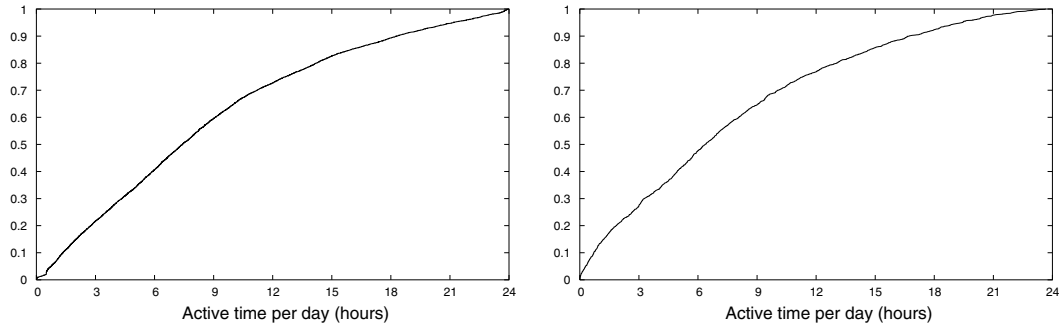
3.1: Fall/Winter 2003/4 3.2: Fall 2001

Fig. 3. [syslog] Number of active, mobile, and roamer cards per day. A date's data appear to the right of its tick-mark. Note that the scales differ between 2001 and 2003/2004.



4.1: Fall/Winter 2003/4 4.2: Fall 2001

Fig. 4. [syslog] Number of active cards per hour. The average number of active cards for each hour of the day, separately for weekdays and weekends.



5.1: Fall/Winter 2003/4 5.2: Fall 2001

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

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: Fig. 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 2003/2004.

The busiest types of building remained the same. Fig. 10 illustrates the most popular locations on campus. The AP and building names have been sanitized with a name that indicates the type of that building, e.g., "ResBldg1" is a

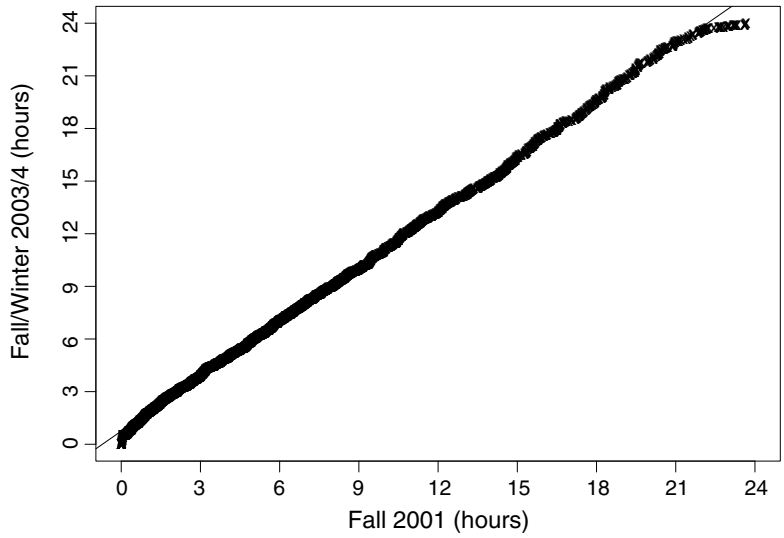
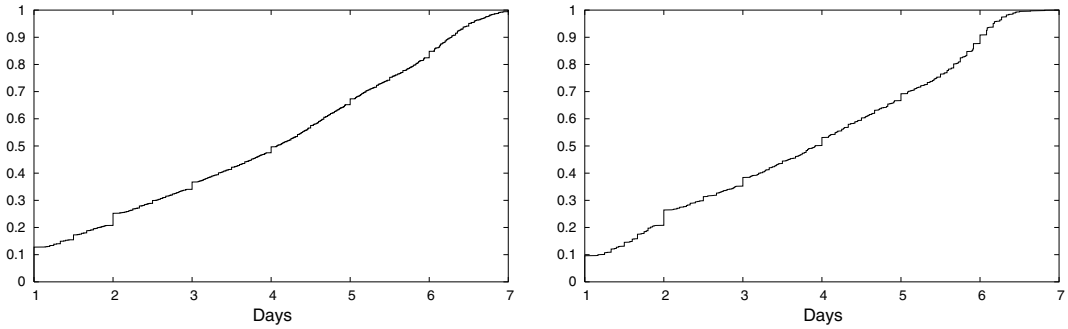
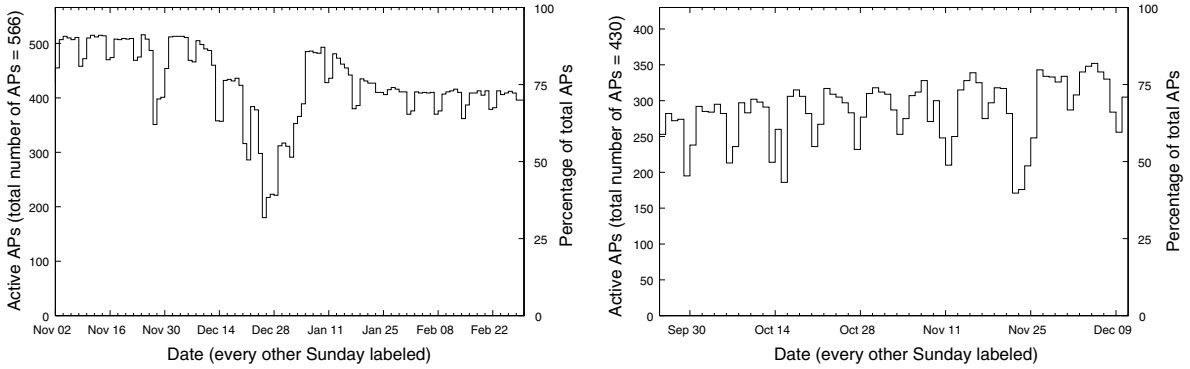


Fig. 6. Quantile–quantile plot, average time per day per user.



7.1: Fall/Winter 2003/4 7.2: Fall 2001

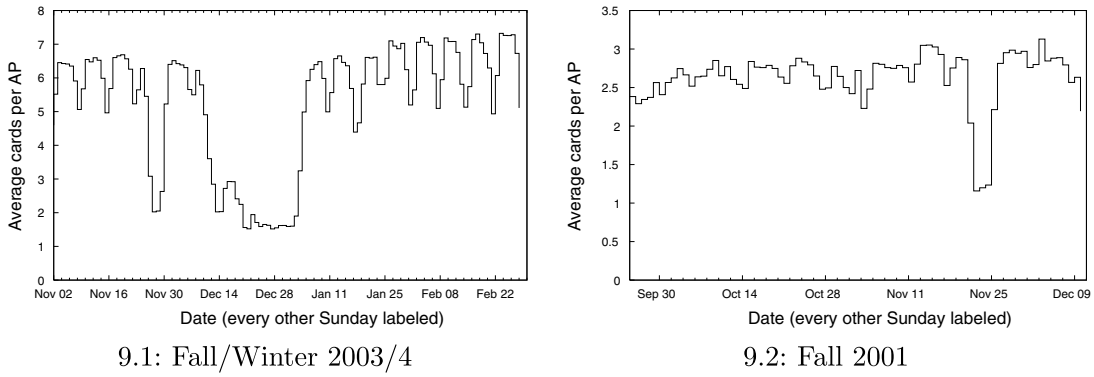
Fig. 7. [syslog] Average active days per week per user, distribution across users.



8.1: Fall/Winter 2003/4 8.2: Fall 2001

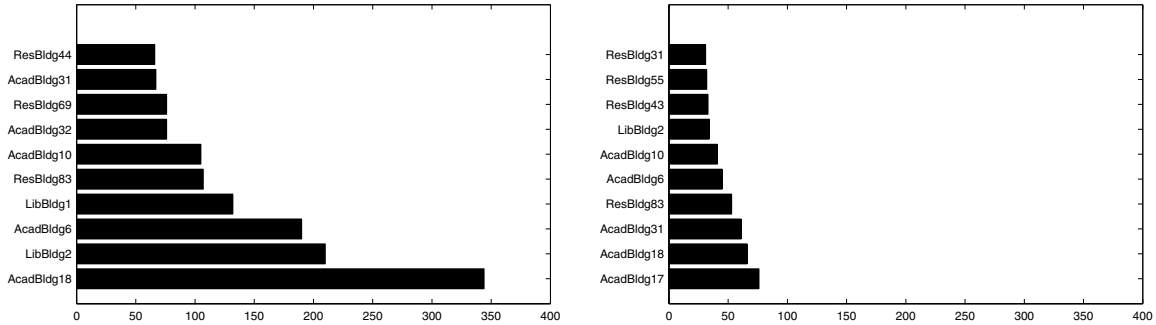
Fig. 8. [syslog] Number of active APs per day. The *y*-axis range is from 0 to the total number of APs.

residential building. We see that academic buildings and libraries continued to see the largest number of cards. This is not surprising, given that these are communal areas visited by many, if not most, students. The peak population was much larger in 2003/2004, due to the larger population of wireless cards.



9.1: Fall/Winter 2003/4 9.2: Fall 2001

Fig. 9. [syslog] Average number of active cards per active AP per day. Note the different *y*-axis scales.



10.1: Fall/Winter 2003/4 10.2: Fall 2001

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

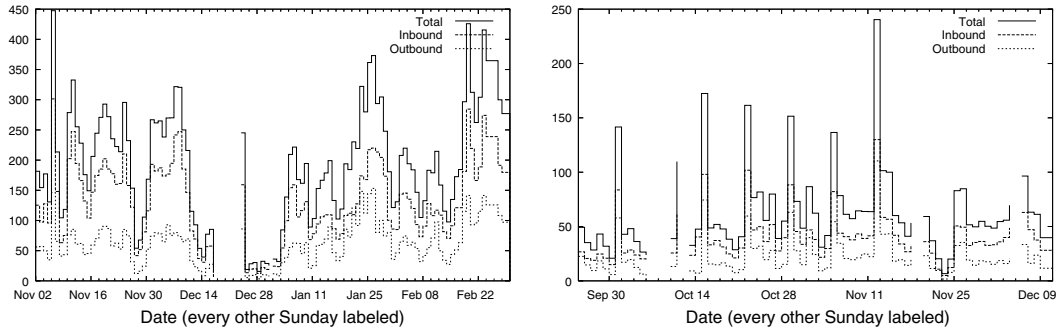
4.2. Traffic

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 400 GB in 2003/2004, compared to 150–250 GB in 2001 (Fig. 11). Nonetheless, the average daily traffic *per active card* rose from 27.0 MB in 2001 to 71.2 MB in 2003/2004. Today's wireless users are far more active on the network than before.

We now consider the applications used on the WLAN. To identify applications, we categorize each TCP and UDP packet by source–destination flow-pair. We then compare 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 identify and parse SCCP



11.1: Fall/Winter 2003/4 11.2: Fall 2001

Fig. 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.

call-setup packets directed to and from the Call Manager servers to determine 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.

Classifying applications by port number is insufficient, since applications may use randomly-assigned port numbers, or applications may masquerade behind different port numbers (e.g., a P2P application may operate on port 80). In the earlier version of this paper we observed a large amount of P2P traffic. Thus we chose to explicitly look further for specific P2P applications. Since we captured the first 200 bytes of each packet, this includes some TCP or UDP payload. We parse this payload to look for packets belonging to four popular P2P applications: BitTorrent, DirectConnect, Gnutella and Kazaa. To save processing time, we only examine the first 10 packets of each TCP or UDP flow-pair. If a flow-pair cannot be identified as a P2P application in the first ten packets in either direction, we use the services file to identify the flow-pair by port number. If we are unable to identify the flow-pair, we consider it "unknown".

We further filter our flow-pairs by ignoring any TCP flow that does not contain at least one ACK segment. In an earlier analysis of this dataset [12], we observed a surprising level of filesystem traffic. We suspected that a large proportion of this traffic comes from worms such as Welchia or Nachi, which conduct scans on the Windows file sharing ports. We do not consider these scans to be filesystem traffic, and so by removing unacknowledged TCP SYN segments, we ignore these scans in this study. In practice these unacknowledged flows did not account for much of the total filesystem traffic in 2003/2004: 1.5% by bytecount, and 3.0% by packetcount (the 2001 figures are 8.2% and 20.6%, but these numbers are inflated as explained below).

In ignoring these TCP flows without any ACK segments, we discovered an anomaly in our 2001 dataset. We discovered a large number of TCP flows containing no ACKs. Many of these flows were long-lived and included many hundreds of megabytes of data, and the destination MAC address was often set to 00:00:00:00:00:00. We suspect that a malfunctioning switch, upstream of one of our sniffers, was flooding all its interfaces with these frames. We have removed all of these unacknowledged flows, which accounted for 70.2% of the 2001 traffic by bytecount (76.1% by packetcount). 0.2% of the 2003/2004 data were unacknowledged flows by bytecount (0.6% by packetcount).

The port numbers that we saw represented thousands of applications. To summarize the traffic, we grouped the applications by type. We based our groupings on the SLAC monitoring project [17], but with changes to reflect some of the most popular applications on campus (Table 3). Two applications are Dartmouth-specific: DND (Dartmouth Name Directory) is a directory service, and BlitzMail is a popular e-mail and news client.

For those comparing this paper to our earlier study [15], note that this application classification differs from the more specific view of the data presented there. Furthermore, please note that the flow-analysis method used in this paper means that our results differ from the MobiCom version of this study [12].

Table 3
Classification of applications

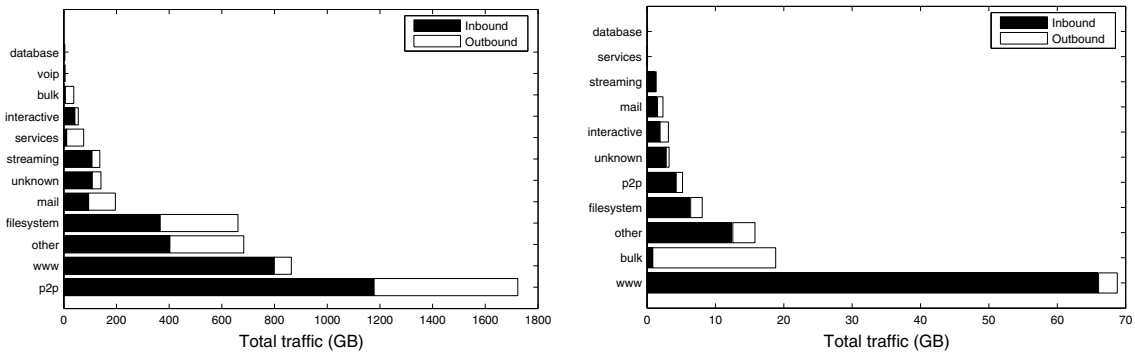
Category	Applications
Bulk	FTP, backup
Database	Oracle, PostgreSQL, SQLnet
Interactive	IRC, AIM, iChat, klogin, rlogin, ssh, telnet
Mail	POP, SMTP, IMAP, NNTP, BlitzMail
P2P	DirectConnect, Gnutella, Kazaa, BitTorrent, eDonkey, Napster
Services	X11, DNS, finger, ident, DND, Kerberos, LDAP, NTP, printer, BOOTP, Rendezvous/ZeroConfig, BGP, portmapper, Service Location
filesystem	SMB/CIFS, NetBIOS, AppleShare, NFS, AFS
Streaming	RealAudio, QuickTime, ShoutCast, RTSP, Windows Media
VoIP	Cisco CallManager, SCCP, SIP, Vocera
WWW	HTTP, HTTPS
Other	All named ports that do not fit into the above categories
Unknown	All unnamed and unidentified ports

The applications used on the network changed significantly. Fig. 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 2003/2004, and 22 out of 476 APs in 2001. Also note that Fig. 12.2 does not contain a bar for VoIP, since this dataset predates the installation of the VoIP system. The proportion of web traffic (marked www) decreased significantly, from 54.3% of the traffic in 2001, to 18.8% in 2003/2004. Of particular interest are the increases in P2P file sharing (from 4.1% in 2001 to 37.6% in 2003/2004), filesystem traffic (from 6.4% to 14.4%) and streaming audio/video (from 1.0% to 3.0%). 0.1% (5.16 GB) of the total traffic was VoIP.

Traffic destinations remained the same. Fig. 13 shows the proportion of near (on-campus) traffic to far (off-campus) traffic. In 2001, on-campus traffic made up 74.2% of the total wireless traffic. In 2003/2004 this situation remained similar, with on-campus traffic comprising 66.9% of the traffic. Note that in a previous version of this paper [12], we stated that off-campus traffic exceeded on-campus traffic in our 2001 trace. We now believe this to be incorrect, having removed the unacknowledged flows as mentioned above.

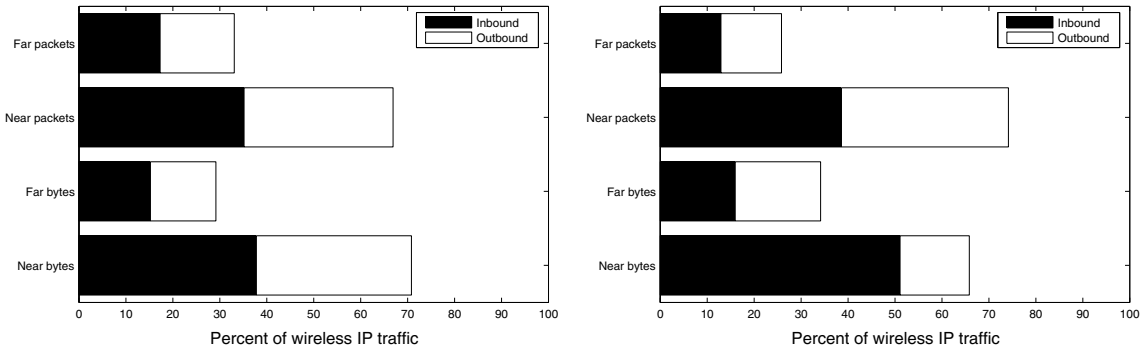
Residences and libraries continued to generate the most traffic. Fig. 14 shows the average daily traffic levels on each AP. It can be seen that the increase in traffic was not due to additional wireless coverage; as increased user population and traffic per user increased, the traffic per AP increased. We also see that residential and library buildings remained the most active. Libraries have become increasingly popular, whereas in 2001 residences saw more traffic. The ordering of the less popular categories changed, but the majority of wireless network traffic continued to occur in residential, academic and library buildings.

Fig. 15 shows the maximum amount of traffic seen at an AP over the course of our trace. Per AP traffic levels have risen considerably. The busiest AP in 2001 saw 17.7 GB in one day, whereas in 2003/2004 the busiest AP saw 120.4 GB. We again see that libraries have become more popular; while residences still make up most of the busiest



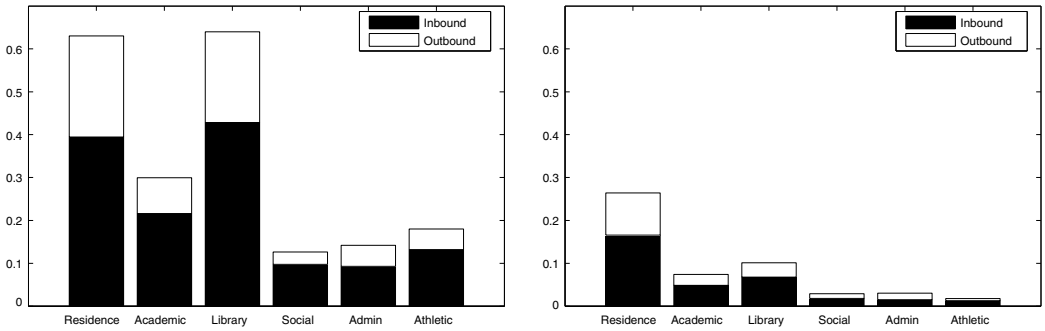
12.1: Fall/Winter 2003/4 12.2: Fall 2001

Fig. 12. [tcpdump] Total traffic (GB), by application. Note the different *x*-axis scales.



13.1: Fall/Winter 2003/4 13.2: Fall 2001

Fig. 13. [tcpdump] Proportion of near and far traffic. "Near" traffic is to or from dartmouth.edu, all else is "Far". Here, traffic was measured in bytes.



14.1: Fall/Winter 2003/4 14.2: Fall 2001

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

APs, the three busiest APs in 2003/2004 are located in a library. One of these library APs saw almost only outbound traffic on one of its busiest days; we are unclear as to the reasons for this result.

5. Specific applications

In Section 4, we 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 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

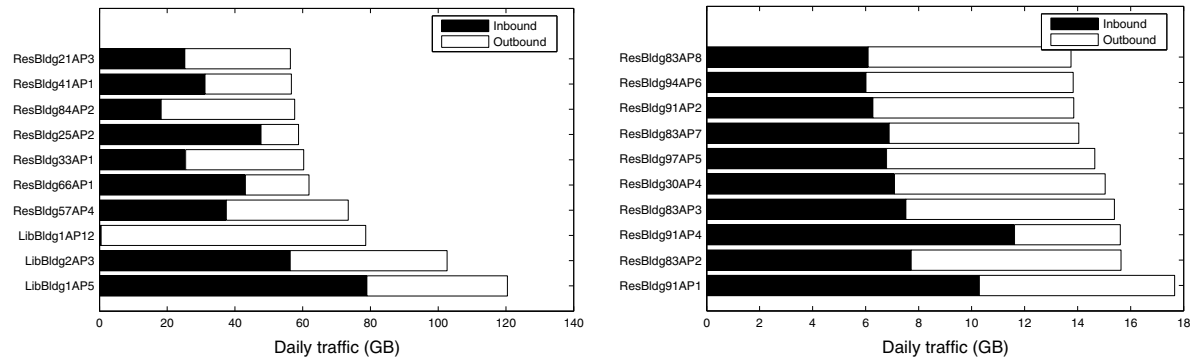


Fig. 15. [SNMP] Maximum daily traffic (GB), for the busiest APs. Ranked by their busiest day. Note the different *x*-axis scales.

Table 4
VoIP devices

Device	Count
Wired Cisco VoIP telephone	80
Wireless Cisco VoIP telephone	20
Cisco SoftPhone	86
Telesym PocketPC SoftPhone	6
Vocera VoIP badge	70
Total	262

our SNMP data to determine whether a given call was made on the WLAN. If either IP address in a record was seen in an SNMP poll during the period of the call, we consider the call to be wireless.

Table 4 lists the number of devices that made a call during our trace period. Some devices, while active on the network, made no calls at all during the trace period, and so there are discrepancies with Table 1.

VoIP usage mirrors general network usage. Fig. 16 shows the number of calls made each day over our monitoring period. We again see two dips for Thanksgiving and Christmas break. VoIP usage shows diurnal patterns (Fig. 17), and these are similar to those for overall WLAN usage (Fig. 4).

VoIP population was static. The number of regular VoIP users shows little growth over the course of our trace (figure omitted for brevity). We again had 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 s (Fig. 18). For calls from wired devices, the median duration was 42 s, whereas for wireless devices, the median duration was 31 s. A Kolmogorov–Smirnov 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 from our analog PBX. The median duration of the off-campus VoIP calls[6] was 63 s, whereas the median duration for off-campus non-VoIP calls was 103 s. A K–S test indicates that non-VoIP calls are significantly longer. It is not clear why VoIP calls, both wired and wireless, would be shorter than 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, wired devices tended to make more calls than wireless devices (Fig. 19). Many wireless devices were only used once or twice, or not at all. Unfortunately, we lack 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 (Table 5). Campus and local calls were the next most popular destinations. This skew may be due to a recent decision by our network administrators to make all domestic 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 (the corresponding figure for VoIP calls was 75.6%).

5.2. Peer-to-peer applications

Peer-to-peer (P2P) traffic increased from 4.1% of the total traffic in 2001 to 37.6% in 2003/2004. The absolute increase was from 5.2 GB to 1723.6 GB, 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. Fig. 20 shows that by far 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

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

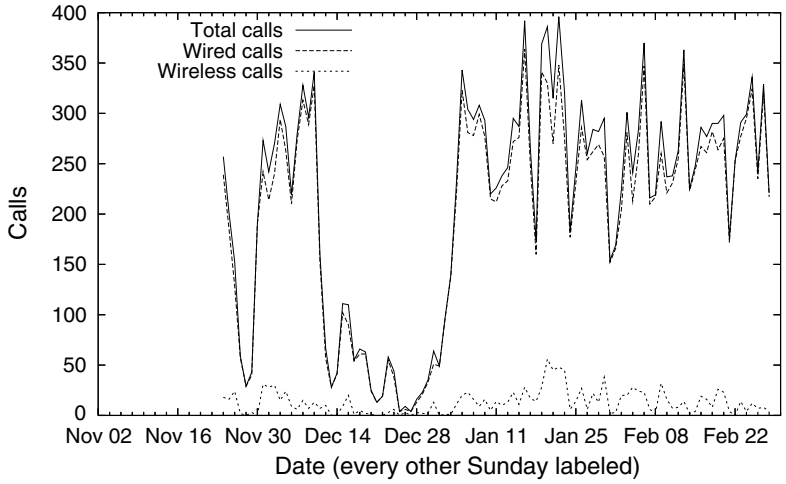


Fig. 16. [CDR] Calls made by wired and wireless devices over time. The wireless curve is much smaller than the wired curve.

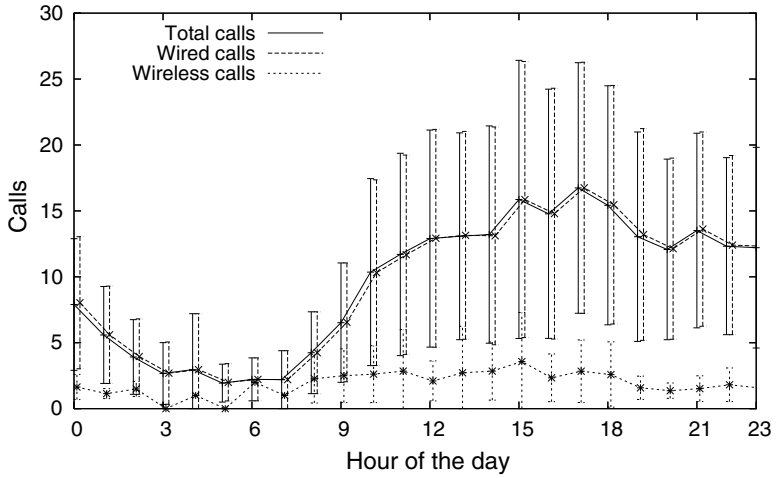


Fig. 17. [CDR] Number of calls made by hour. The line shows the mean, and the bars show standard deviation. The values are slightly offset so that the bars are distinguishable. The wireless curve is on the bottom.

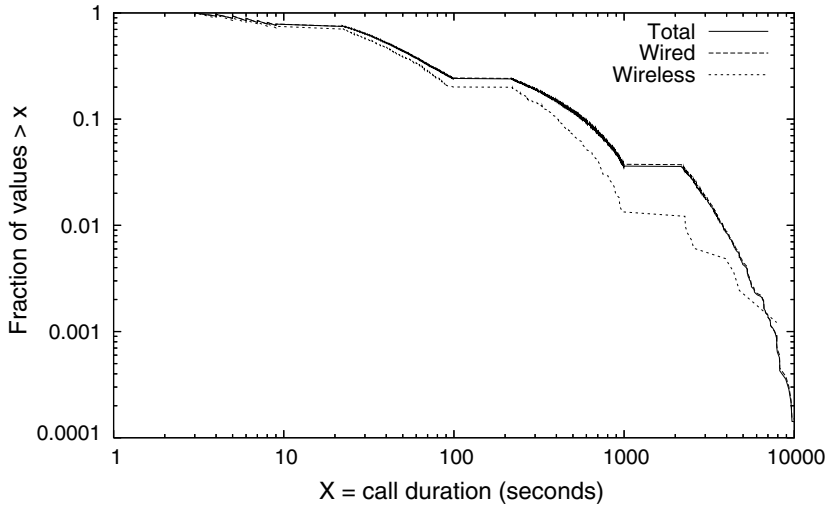


Fig. 18. [CDR] log–log plot of the distribution of call duration. We only consider calls of duration $\geqslant$1 s and $\leqslant$10,000 s.

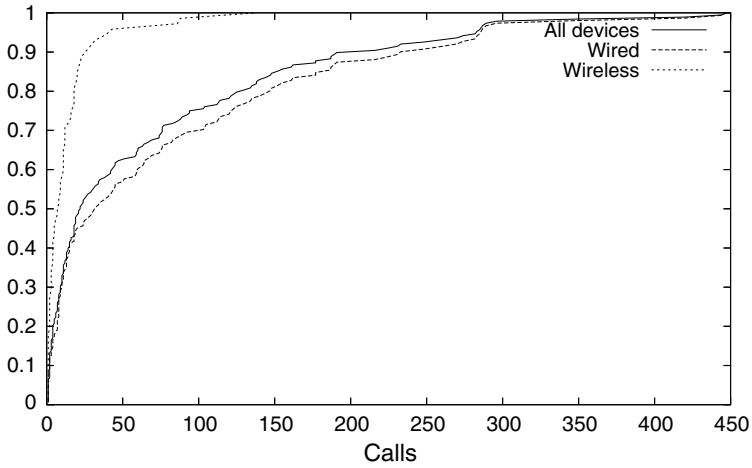


Fig. 19. [CDR] CDF of the number of calls made by a VoIP device.

Table 5
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%)
Non-geographic (411, 911)	7	(0.1%)	7	(0.1%)	0	(0.0%)
Voice-mail	1242	(9.2%)	1217	(9.7%)	25	(2.5%)
Total	13,567	(100.0%)	12,569	(100.0%)	998	(100.0%)

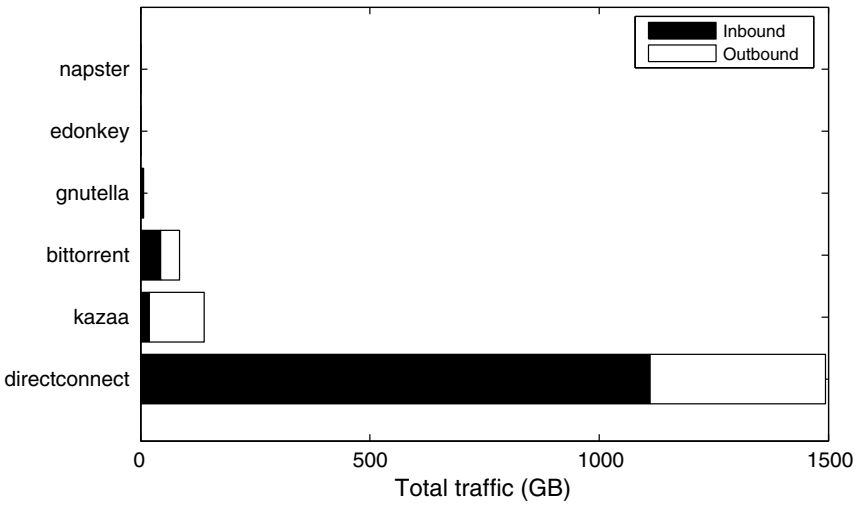


Fig. 20. [tcpdump] Total P2P traffic (GB), by application. Napster and eDonkey were non-zero but negligible.

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. 82.5% of the P2P traffic was between on-campus hosts. 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 [6] 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. The extremes of Fig. 21 show that a small number of cards send and receive a large amount of P2P data. Of the 945 cards that saw more than 1 MB of P2P traffic, a mere 30 cards (3.2% of the population) were responsible for over 50% (873.39 GB) of the traffic. This behavior has been observed on another campus, where 4% of the population was responsible for 50% of Kazaa traffic [24].

5.3. Streaming media

The proportion of wireless streaming audio/video traffic increased by 192% between 2001 and 2003/2004, and we saw 136.9 GB of streaming traffic in our 2003/2004 trace.

Most, but not all, streaming media was inbound. Fig. 22 shows that this traffic was made up mainly of two applications: RealAudio and iTunes. Most streaming traffic was inbound: applications such as RealAudio and Windows Media (ms-streaming) 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 55.9% of the iTunes traffic was outbound (see Fig. 22).

Most streaming traffic was within campus. Most (65.1%) of the streaming traffic was to or from hosts on campus. 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 were large (300–400 MB), and this content may account for much of the on-campus traffic. Another contributing factor is that by default, iTunes will

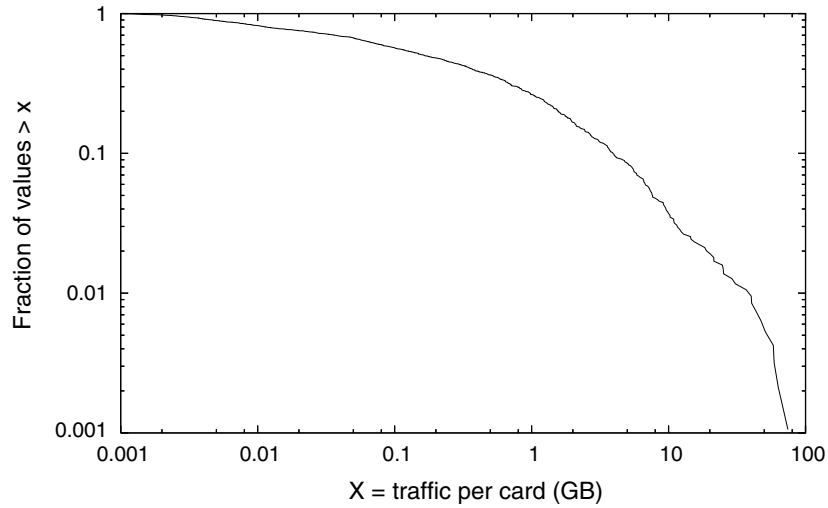


Fig. 21. [tcpdump] log–log distribution of traffic per card by P2P users. Cards that saw less than 1 MB are ignored.

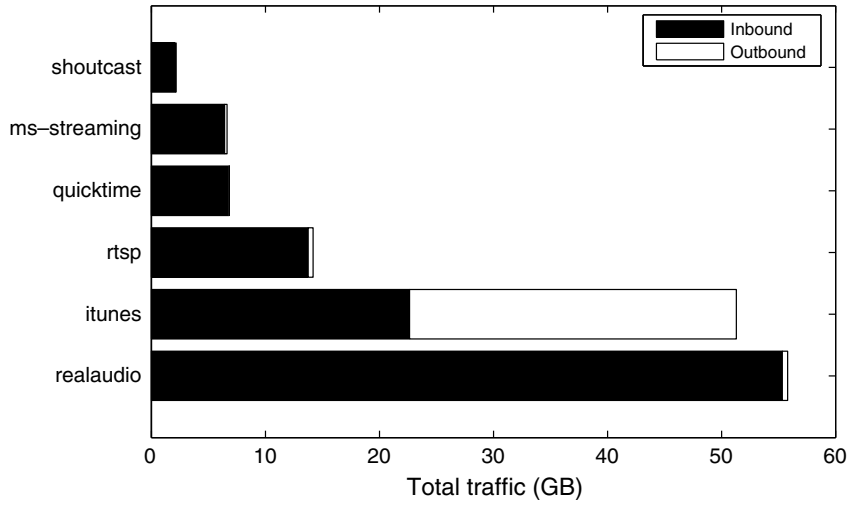


Fig. 22. [tcpdump] Total streaming traffic (GB), by application.

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 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. Fig. 23 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 [4], 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 50 m 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 a 50 m diameter of that AP to represent the card's home location. Like Balazinska and Castro, we do not consider users who spent less than 50% of their time at their home, due to the difficulty of accurately determining a "home" for such users. We note this fact on the left half of Fig. 23, simply projecting each curve to the *y*-axis. The *y*-intercept thus indicates the fraction of users, in each case, who had no well-defined home location.

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. Fig. 23 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. If we remove the overnight period (0000–0600) from our data, then we find that 50% of our users spend 69.2% of their time associated with a single AP, which is much closer to the 60% seen by Balazinska and Castro [4].

Another possible reason for the low observed mobility was the presence of "visitors"; users that visited our campus for one or two days, used the WLAN intermittently in one place, and then left, to be never seen again. We examined this hypothesis by removing any cards that were only ever observed for two days or less. Of the remaining cards, we found that 50% of the cards spent 71.6% of their time associated with a single AP, or 97.9% in a home location. Even regular users of the WLAN are immobile.

Our results may also differ from the corporate data because we use syslog records, with a 1-s timestamp resolution, whereas Balazinska and Castro use SNMP with a 5-min poll period. Their longer 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 [4]. Fig. 24 again shows that our users were less mobile (had lower prevalence) than corporate users: the dashed line represents the line of best fit for the corporate data [4]. (Although the figure informally compares a fit curve from one study with a histogram from another study, it is visually evident that the two are quite different.) Note that we cannot be sure about the absolute difference in prevalence between these two datasets as the SNMP-collected corporate data missed short visits to APs. If users tended to visit APs for short periods of time and then return (e.g., "ping-ponging"), then the SNMP data would overestimate prevalence. On the other hand, if users

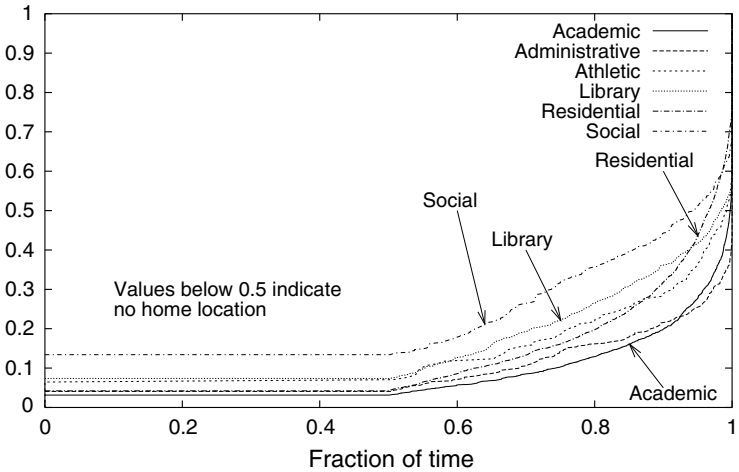


Fig. 23. [syslog] Fraction of time that users spend at their home location, by the building type of their home location.

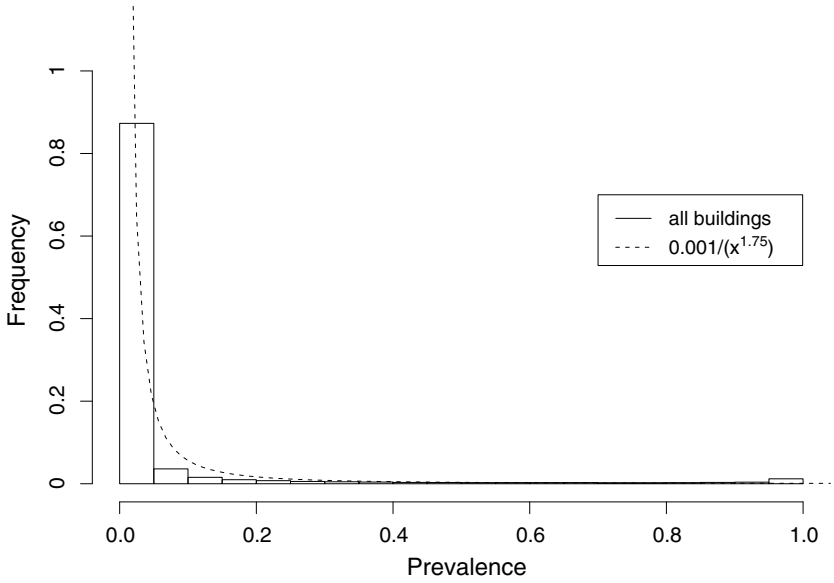


Fig. 24. [syslog] CDF of prevalence values for all buildings. Zero-values are discarded.

tended to visit many APs quickly in succession, the SNMP data would underestimate 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 [4]. We again consider persistence using our 50 m 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 50 m, we output a persistence value and clear the list.

The line in Fig. 25 marked $0.92/x$ is the line of best fit from Balazinska and Castro [4]. It is clear that our data are 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. Thus, in Fig. 25 we have also calculated persistence as originally defined (the line marked "All locations (by AP)"). These values are lower, as they include roams within a 50 m 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 the SNMP approach tends to overestimate persistence, the fact that we saw longer persistence in our data is not an artifact of the two measurement techniques; if anything, the difference is stronger than it appears.

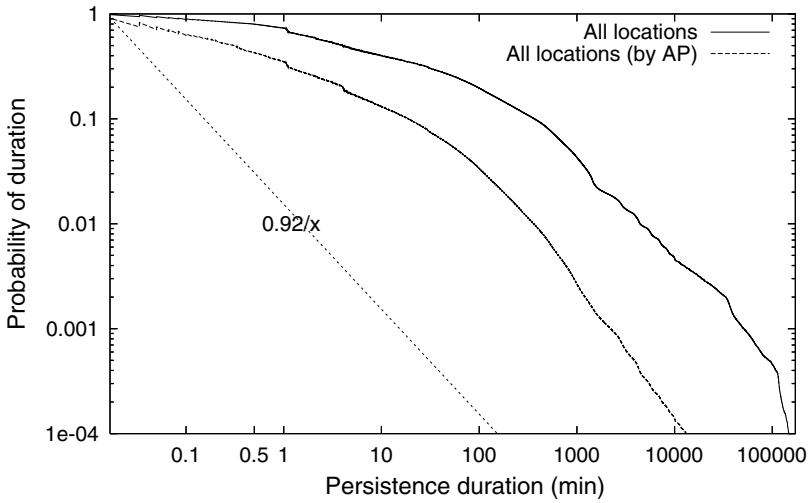


Fig. 25. [syslog] log–log distribution of user persistence values. We show values calculated using our session diameter metric and persistence on a per AP basis for comparison. All locations (by AP) is the leftmost curve.

Different devices traveled more widely. Fig. 26 shows the total number of APs visited by a device. The median number of visited APs has risen from 9 in 2001 to 12 in 2003/2004. 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 their laptop and connects. A similar effect can be seen in Fig. 27, which shows the session diameter for different device types. The always-on VoIP devices tend to travel further than laptops and PDAs.

If we consider the distance traveled by a device over a longer timescale, however, the differences between devices diminish. Fig. 28 shows the "daily session diameter", that is, the maximum inter-AP distance traveled in a day by a given device. We see that VoIP devices and laptops traveled similar amounts over the course of a day. This may be because laptops are used nomadically, and so a session-centric metric underestimates their mobility. A user traveling from point *A* to point *B* carrying both a VoIP device and a laptop might use the VoIP device en route to *B*, thus creating high session mobility, but only use the laptop at points *A* and *B*, in separate sessions. When considering the daily diameter, both devices have visited *A* and *B* in the same day, and so their daily diameters are equivalent.

Fig. 28 also shows that PDAs traveled far further in a day than other devices. This may be due to the small sample size of PDA users, who may tend to be early adopters and thus have different usage characteristics to the more-general laptop population. On the other hand, one might expect PDAs to exhibit high mobility as their small size means that they can be carried anywhere on campus, and accessed more quickly than a laptop. This might lead PDAs to be used in more locations over the course of a day.

Different devices had different session characteristics. Some of the mobility differences between devices can be attributed to the different session types for different devices. Fig. 29 shows the distribution of session durations for different device types. As many sessions lasted almost

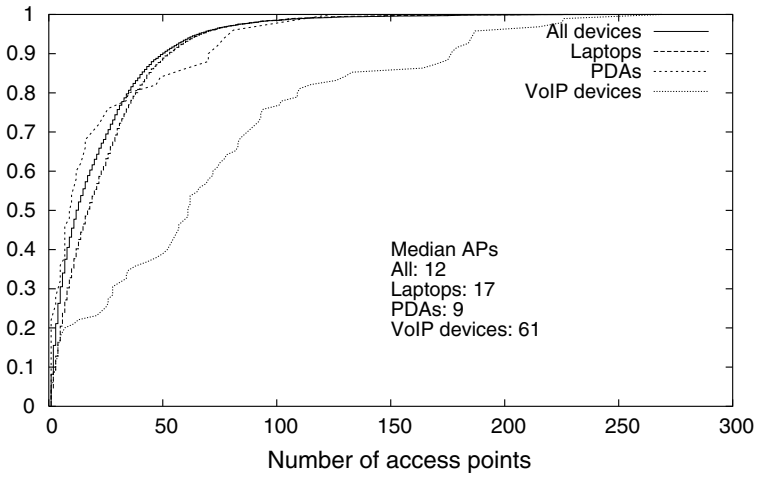


Fig. 26. [syslog] CDF of the number of APs visited by a user.

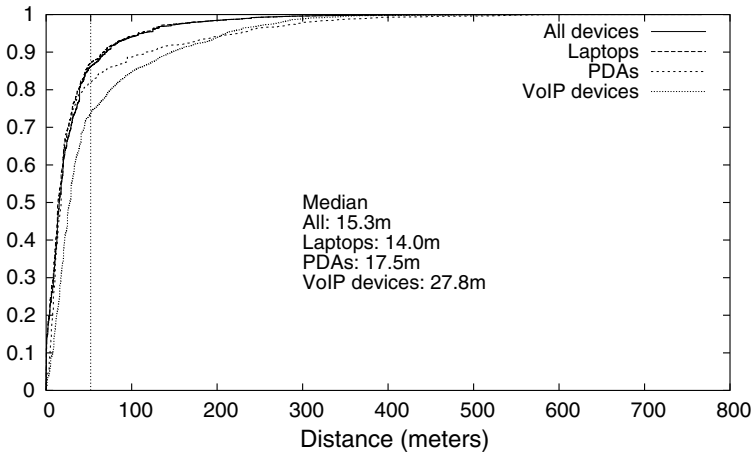


Fig. 27. [syslog] Session diameter, distribution across sessions, by device. The vertical dashed line indicates 50 m, our threshold for a mobile session.

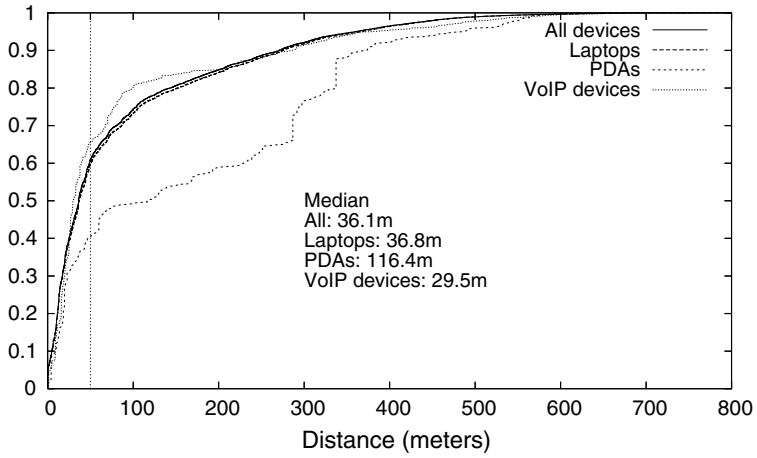


Fig. 28. [syslog] Session diameter, distribution across days, by device. The vertical dashed line indicates 50 m, our threshold for a mobile session.

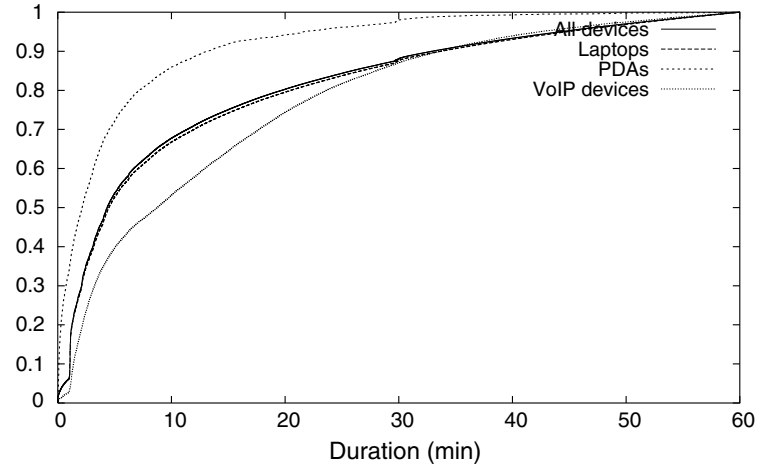


Fig. 29. [syslog] Session duration, by device. Data are limited to those sessions shorter than 1 h; the tails of these distributions are very long – several weeks, in some sessions.

the length of our trace period (stationary devices that were never switched off), the plot shows those durations of less than 1 h for clarity. All of the device types have a short median session duration, less than 10 min. The short median, consistent with our earlier results, is detectable in the syslog data but would be difficult to observe with a longer 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.

Different applications had different mobility characteristics. In Section 5, we focus on three of the newest wireless applications: VoIP, P2P, and streaming media. In Fig. 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 audio or video application tends to involve active user participation, and so mobility is impeded by the need to continuously look at or access 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.

Different applications were used in different locations. We consider application usage in different locations by examining the applications used by a device when the device was situated in its home location (as defined above), and when the device was situated elsewhere (the "non-home location").

As our sniffers only covered a subset of the campus, we could not examine the behavior of every device, since some devices' home locations lay outside our sniffer coverage area. Instead, we only consider those devices that have home locations covered by our sniffers. In addition, we

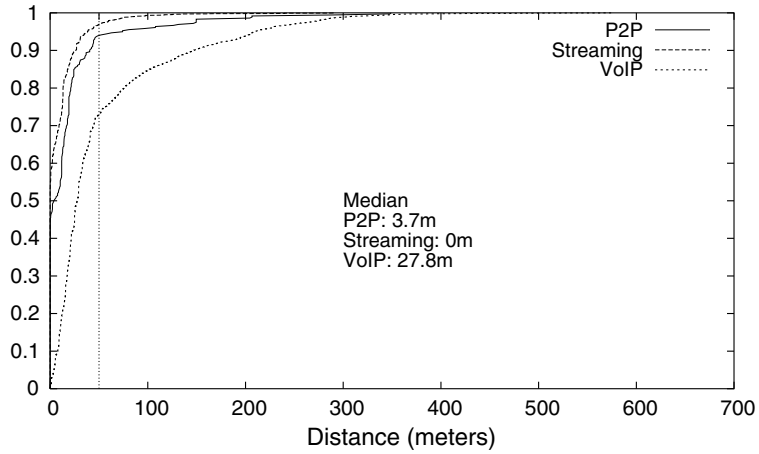


Fig. 30. [syslog] Session diameter, distribution across sessions, by application. The vertical dashed line indicates 50 m, our threshold for a mobile session.

only consider laptops, since single-purpose VoIP devices would have identical application behavior in both home and non-home locations, as they can only use one application. From those laptops with home locations covered by our sniffers, we selected the 100 most 'mobile' clients, in terms of the amount of time spent in a home location. We chose those clients who spent the largest amount of time away from home, to prevent non-home application usage being skewed by devices that were seldom used in non-home locations.

Figs. 31 and 32 show the application mix for home and non-home locations, using the same methodology as Section 4. We see that P2P applications are used both in home and non-home locations. This is surprising, since P2P file transfers might require a long amount of time on the network, which may not be convenient when a user is mobile or at a non-home location. We observe that there is more outbound P2P activity in non-home locations, which may indicate that users are unknowingly sharing files when away from home, for instance by leaving their P2P client active. Streaming media is used in a greater proportion in non-home locations than at home, and non-home streaming traffic exceeds web traffic. One explanation is that a user not at home might not have access to their usual music collection, and so choose to stream audio from other sources (e.g., iTunes users in their current subnet).

Compared to the overall application mix (Fig. 12.1), the amount of filesystem traffic is much lower in our 100 most mobile laptops. We speculate that highly-mobile users might be less likely to rely on remote filesystem mounts, since these do not cope well with high levels of mobility. It appears that less-mobile users do generate more filesystem traffic than these highly-mobile users. From the syslog data we extracted a list of cards that only ever appeared at a single location, but who also appeared on more than two days (to remove "visitors"). We then filtered this list to only include laptops that were covered by our sniffers. This resulted in a list of 717 cards. The application mix for these

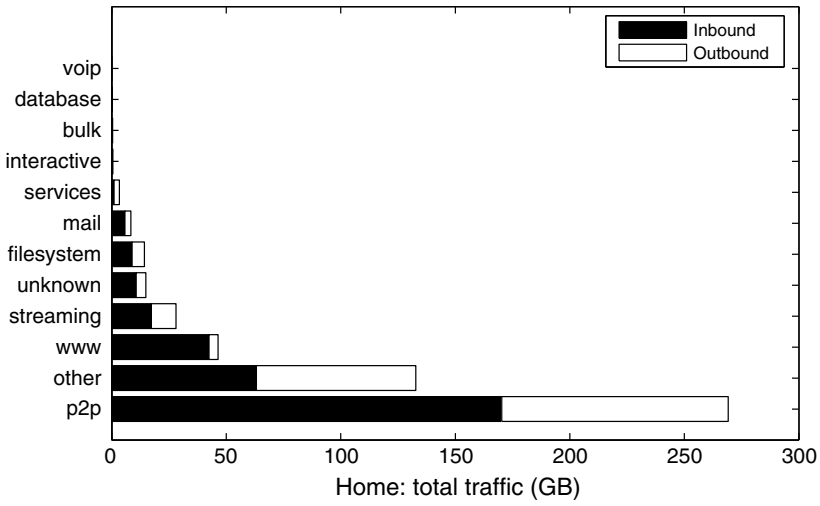


Fig. 31. [tcpdump] Traffic at home location (GB), by application.

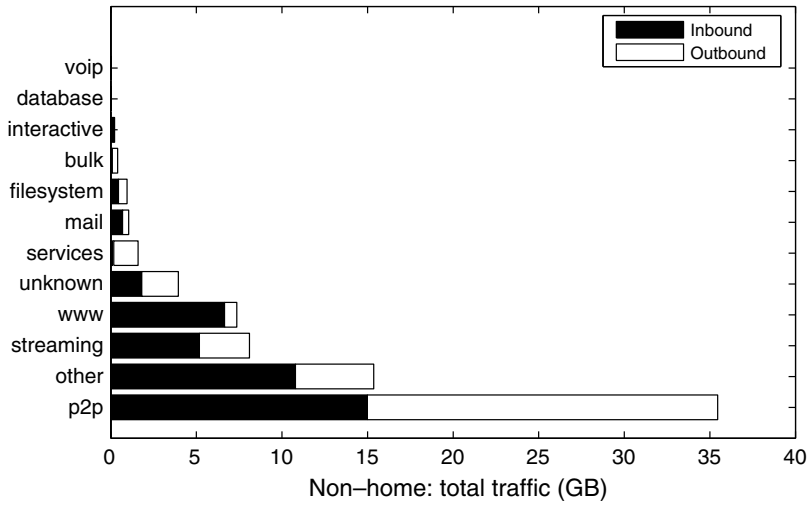


Fig. 32. [tcpdump] Traffic at non-home locations (GB), by application.

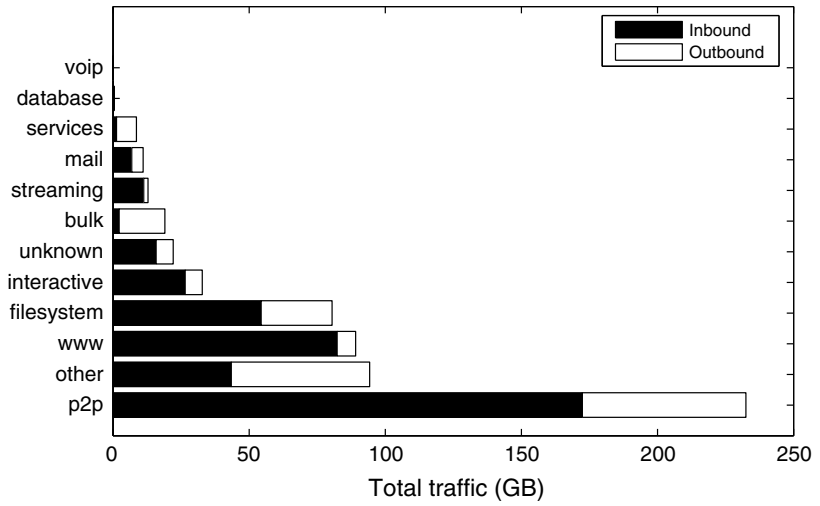


Fig. 33. [tcpdump] Traffic for regular non-mobile clients (GB), by application.

cards (Fig. 33) shows that the proportion of filesystem traffic is similar to the overall application mix.

7. Related work

Our study is the largest characterization of WLAN users to date. One of the earliest analyses of WLAN usage was by Tang and Baker, who used tcpdump and SNMP to trace 74 users in the Stanford CS Department over a 12-week period in 2000 [27]. 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. [3] 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 s to receive SNMP responses from all of our APs. Since they study a conference, user behavior is homogeneous, with clients following the conference schedule. Most sessions were short (<10 min). 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 [13]. 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 precise than ours due to a 15-min poll interval.

Saroiu et al. [24] traced 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 HTTP. This result is similar to our observations (38% P2P and 19% HTTP), even though we examined traffic within the campus and they examined the border.

McNett and Voelker [19] installed a tool on wireless PDAs, and used this tool to collect 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 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.

Schwab and Bunt study the University of Saskatchewan campus WLAN [25]. This WLAN uses a central RADIUS authentication server, allowing for accurate session determination. Their trace is significantly smaller than ours (136 users on 18 APs over one week), and their WLAN does not cover residential areas, so their diurnal usage patterns differ from ours. The largest identified protocol was HTTP, at 28% of packets. They were unable to identify 35% of TCP packets, 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 other major P2P protocols).

Chinchilla et al. analyzed WWW users on the University of North Carolina campus WLAN [8]. 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. In later work [23], Papadopouli et al. find that session durations on the UNC WLAN can be modeled by a BiPareto distribution.

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. Mahajan et al. look at wireless MAC-layer behavior at the 2004 SIGCOMM conference [18] and present techniques for merging traces from multiple sniffers. They find that the wireless medium is used inefficiently, with nodes often backing off unnecessarily. The Jigsaw project looks at wireless-side traffic in the UCSD CSE department [7], and also offers methods for merging traces. The CSE building is busy, with 1026 unique MAC addresses observed in a single day, and with similar diurnal patterns to our academic buildings.

All of the above studies look at academic users. We have already mentioned Balazinska and Castro [4], who traced 1366 corporate users over four weeks, and developed two metrics for mobility, *prevalence* and user *persistence*. Blinn et al. [5] look at users on a commercial hotspot network, and find diurnal usage patterns similar to our own.

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 had matured, and as the userbase grew 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 [26].

Although most users stayed predominantly in one location, different devices and applications had different mobility characteristics. In particular, always-on VoIP devices associated with more APs and had longer-lived and farther-ranging sessions. Always-on devices are becoming more popular, and as a result WLANs may see increased numbers 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 standards such as Mobile IP 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 reasons, such as security, since many embedded wireless devices lack support for newer security standards like 802.11i.

There was a large increase in the amount of P2P traffic on our WLAN, despite the presence of a high-speed wired Ethernet network throughout our campus. Evidently the convenience of a wireless solution outweighs the limited bandwidth of an 802.11b network. As 802.11 is a shared-medium, large P2P 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 s, significantly shorter than calls on the old non-VoIP phone network. If such short calls are representative of wireless VoIP usage, this may impact protocol design: 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; 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 [6], and business surveys have started to examine academic wireless usage in addition to public usage [14]. Indeed, a university campus contains elements of an enterprise, a residential community, public hotspots (libraries and restaurants), research labs, and educational workloads.

Another caveat to be considered is that our results only look at the wireless portion of our LAN. Some of the changes that we have observed, for example, the increase in P2P usage, 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. Since the time of this study, Dartmouth College has upgraded the WLAN to a 802.11/a/b/g network with over 1600 APs, added authentication, and migrated the campus CATV network to an IP-based streaming video platform. Hand-held Wi-Fi devices are now more commonly-available, including hybrid and UMA cellphones, VoIP phones, media players and games consoles. As a result, we expect to see higher mobility and more streaming media usage on the WLAN, in particular higher-bandwidth video on the 802.11a network that is difficult to provide over 802.11b.

Our existing measurement infrastructure only looks at the wired side of our wireless APs. We are extending our sniffing capability to include wireless sniffers, to monitor the 802.11 MAC layer. Whilst some researchers have taken 802.11 wireless measurements [10,20,29,18,7], these have typically taken place in laboratory or small-scale conditions, and there is little wireless monitoring of a large campus WLAN. 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 these data will provide further insights into WLAN usage, and the effects of new applications on the network.

We are also interested in understanding *why* we have seen the changes in network behavior that we have presented in this paper. We have collaborated with a sociologist to ask our wireless network users why they are using particular applications, or exhibiting particular behavior [2,11].

Due 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 further study would be valuable. We saw large amounts of P2P traffic, but limited to a small fraction of users; further analysis of this distribution could be useful in developing AP-load-balancing algorithms that balance users according to their bandwidth requirements.

We welcome other researchers to make use of our data, and sanitized versions of both our 2001 and 2003/2004 traces, and even newer data, are publicly available through CRAWDAD [16]. Moreover, if other researchers choose to collect similar data at other networks, such as corporate or public networks, this could aid in conducting cross-validation studies to determine the generality of our conclusions.

Acknowledgements

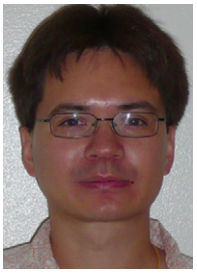
The authors are thankful for the help of Charles Clark and Udayan Deshpande in setting up the sniffers. We thank Kobby Essien for collecting the Fall 2001 trace. We are grateful for the assistance of the staff of Dartmouth Computing Services, particularly Jim Baker, Craig Bisson, David Bourque, Steve Campbell, Robert Johnson and Brad Noblet; of Computer Science, particularly Wayne Cripps and Tim Tregubov; and of Engineering, Ted Cooley and D.J. Merrill.

This project was funded by the Cisco Systems University Research Program, NSF Infrastructure Award EIA-9802068, and Award Number 2000-DT-CX-K001 from the Science and Technology Directorate of the US Department of Homeland Security (DHS). Points of view in this document are those of the authors and do not necessarily represent the official position of any sponsor.

References

[1] E. Adar, B.A. Huberman, Free riding on Gnutella, First Monday 5 (10) (2000). October.
[2] D. Anthony, T. Henderson, D. Kotz, Privacy in location aware computing environments (PLACE), IEEE Pervasive Computing 6 (4) (2007) 64–72. October–December.
[3] A. Balachandran, G.M. Voelker, P. Bahl, P.V. Rangan, Characterizing user behavior and network performance in a public wireless LAN, in: Proceedings of the 2002 ACM SIGMETRICS Conference, Marina Del Rey, CA, June 2002, pp. 195–205.
[4] M. Balazinska, P. Castro, Characterizing mobility and network usage in a corporate wireless local-area network, in: Proceedings of the First International Conference on Mobile Systems, Applications, and Services (MobiSys), San Francisco, CA, May 2003, pp. 303–316.
[5] D. Blinn, T. Henderson, D. Kotz, Analysis of a Wi-Fi hotspot network, in: Proceedings of the International Workshop on Wireless Traffic Measurements and Modeling, Seattle, WA, USA, June 2005, pp. 1–6.
[6] Campus Computing Project, The 2003 National Survey of Information Technology in US Higher Education, October 2003. <http://www.campuscomputing.net/pdf/2003-CCP.pdf>.

[7] Y.-C. Cheng, J. Bellaro, P. Benko, A.C. Snoeren, G.M. Voelker, S. Savage, Jigsaw: solving the puzzle of enterprise 802.11 analysis, in: Proceedings of the ACM Conference on Applications, Technologies, Architectures and Protocols for Computer Communications (SIGCOMM), Pisa, Italy, September 2006, pp. 39–50.
[8] F. Chinchilla, M. Lindsey, M. Papadopouli, Analysis of wireless information locality and association patterns in a campus, in: Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), Hong Kong, China, March 2004, pp. 906–917.
[9] Fyodor, Remote OS detection via TCP/IP stack fingerprinting, Phrack 54 (8) (1998). <http://insecure.org/nmap/osdetect/>.
[10] G. Gaertner, V. Cahill, Understanding link quality in 802.11 mobile ad hoc networks, IEEE Internet Computing 8 (1) (2004) 55–60. Jan/Feb 2004.
[11] T. Henderson, D. Anthony, D. Kotz, Measuring wireless network usage with the experience sampling method, in: Proceedings of the First Workshop on Wireless Network Measurements, Trentino, Italy, April 2005.
[12] T. Henderson, D. Kotz, I. Abyzov, The changing usage of a mature campus-wide wireless network, in: Proceedings of the 10th Annual International Conference on Mobile Computing and Networking (MobiCom), Philadelphia, PA, USA, September 2004, pp. 187–201.
[13] R. Hutchins, E.W. Zegura, Measurements from a campus wireless network, in: Proceedings of the IEEE International Conference on Communications (ICC), vol. 5, New York, NY, April 2002, pp. 3161–3167.
[14] Intel Corporation, Most unwired college campuses survey, April 2004. <http://web.archive.org/web/20040421034947/http://www.intel.com/products/mobiletechnology/unwiredcolleges.htm>.
[15] D. Kotz, K. Essien, Analysis of a campus-wide wireless network, Wireless Networks 11 (2005) 115–133. An earlier version appeared in ACM MobiCom 2002, and as Dartmouth College Technical Report TR2002-432.
[16] D. Kotz, T. Henderson, I. Abyzov, CRAWDAD data set dartmouth/campus (v. 2007-02-08), February 2007. <http://crawdad.org/dartmouth/campus>.
[17] C. Logg, Characterization of the traffic between SLAC and the Internet, Tech. rep., Stanford Linear Accelerator Center, Menlo Park, CA, July 2003. <http://www.slac.stanford.edu/comp/net/slac-netflow/html/SLAC-netflow.html>.
[18] R. Mahajan, M. Rodrig, D. Wetherall, J. Zahorjan, Analyzing the MAC-level behavior of wireless networks in the wild, in: Proceedings of the ACM Conference on Applications, Technologies, Architectures and Protocols for Computer Communications (SIGCOMM), Pisa, Italy, September 2006, pp. 75–86.
[19] M. McNett, G.M. Voelker, Access and mobility of wireless PDA users, ACM Mobile Computing and Communications Review 9 (2) (2005) 40–55. April.
[20] A. Mishra, M. Shin, W.A. Arbaugh, An empirical analysis of the IEEE 802.11 MAC layer handoff process, ACM SIGCOMM Computer Communication Review 33 (2) (2003) 93–102. April.
[21] p0f. <http://lcamtuf.coredump.cx/p0f.shtml>.
[22] J. Padhye, S. Floyd, On inferring TCP behavior, in: Proceedings of the ACM Conference on Applications, Technologies, Architectures and Protocols for Computer Communications (SIGCOMM), San Diego, CA, August 2001, pp. 287–298.
[23] M. Papadopouli, H. Shen, M. Spanakis, Characterizing the duration and association patterns of wireless access in a campus, in: Proceedings of the 11th European Wireless Conference, Nicosia, Cyprus, April 2005.
[24] S. Saroiu, K.P. Gummadi, R.J. Dunn, S.D. Gribble, H.M. Levy, An analysis of Internet content delivery systems, in: Proceedings of the 2002 Symposium on Operating Systems Design and Implementation (OSDI), Boston, MA, December 2002, pp. 315–328.
[25] D. Schwab, R. Bunt, Characterising the use of a campus wireless network, in: Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), Hong Kong, China, March 2004, pp. 862–870.
[26] L. Song, D. Kotz, R. Jain, X. He, Evaluating next cell predictors with extensive Wi-Fi mobility data, IEEE Transactions on Mobile Computing 5 (12) (2006) 1633–1649. December.
[27] D. Tang, M. Baker, Analysis of a local-area wireless network, in: Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking (MobiCom), Boston, MA, August 2000, pp. 1–10.
[28] Vocera. <http://www.vocera.com>.
[29] J. Yeo, M. Youssef, A. Agrawala, A framework for wireless LAN monitoring and its applications, in: Proceedings of the Third ACM Workshop on Wireless Security (WiSe'04), Philadelphia, PA, October 2004, pp. 70–79.



Tristan Henderson is a Lecturer in Computer Science at the University of St. Andrews in Scotland. His research interests include network measurement, wireless networks, security, network economics and multiplayer networked games. He serves on the JANET UK Wireless Advisory Group and the steering committee of the NetGames workshop. Dr. Henderson holds an M.A. in Economics from the University of Cambridge and an M.Sc. and Ph.D. in Computer Science from University College London. For more information, see http://www.cs.st-andrews.ac.uk/tristan.

David Kotz is a Professor of Computer Science at Dartmouth College in Hanover NH. He also serves as the Director of the Center for Mobile Computing, which focuses on wireless networks and mobile computing. He was the Executive Director of the Institute for Security Technology Studies from 2004 to 2007, which is dedicated to interdisciplinary research and education in cyber security and trust. His research interests include security and privacy, pervasive computing, and wireless networks. He has published over 100 refereed journal and conference papers.

After receiving his A.B. in Computer Science and Physics from Dartmouth in 1986, he completed his Ph.D. in Computer Science from Duke University in 1991 and returned to Dartmouth to join the faculty. He is a Senior Member of the ACM and of the IEEE Computer Society, and a member of the USENIX Association. For more information, see http://www.cs.dartmouth.edu/dfk/.

Ilya Abyzov is an Emerging Markets Fixed Income analyst at HBK Capital Management, a global multi-strategy hedge fund. He received his B.A. in Computer Science from Dartmouth College in 2005, where his research interests included wireless networking and voice over IP. Before joining HBK, he worked as an analyst in the Media and Telecommunications M&A group at Morgan Stanley.