Characterizing Usage of a Campus-wide Wireless Network

David Kotz and Kobby Essien
Department of Computer Science, Dartmouth College
Hanover, NH, USA 03755
dfk@cs.dartmouth.edu

Dartmouth Computer Science Technical Report TR2002-423
March 12, 2002

Abstract

Wireless local-area networks (WLANs) are increasingly common, but little is known about how they are used. A clear understanding of usage patterns in real WLANs is critical information to those who develop, deploy, and manage WLAN technology, as well as those who develop systems and application software for wireless networks. This paper presents results from the largest and most comprehensive trace of network activity in a large, production wireless LAN. For eleven weeks we traced the activity of nearly two thousand users drawn from a general campus population, using a campus-wide network of 476 access points spread over 161 buildings. Our study expands on those done by Tang and Baker, with a significantly larger and broader population.

We found that residential traffic dominated all other traffic, particularly in residences populated by newer students; students are increasingly choosing a wireless laptop as their primary computer. Although web protocols were the single largest component of traffic volume, network backup and file sharing contributed an unexpectedly large amount to the traffic. Although there was some roaming within a network session, we were surprised by the number of situations in which cards roamed excessively, unable to settle on one access point. Cross-subnet roams were an especial problem, because they broke IP connections, indicating the need for solutions that avoid or accommodate such roams.

1 Introduction

Wireless local-area networks (WLANs) are increasingly common on university and corporate campuses, in public spaces such as airports and hotels, and even in many personal residences. Although technology such as IEEE 802.11b ("Wi-Fi") is broadly deployed and usage is increasing dramatically, little is known about how these networks are used. A clear understanding of usage patterns in real WLANs is critical information to those who develop, deploy, and manage WLAN technology, as well as those who develop systems and application software for wireless networks.

This paper presents results from the largest and most comprehensive trace of network activity in a large, production wireless LAN. Dartmouth College has 11 Mbps 802.11b coverage for nearly every building on campus, including all administrative, academic, and residential buildings, and most athletic facilities. We collected extensive trace information from the entire network throughout the Fall term of 2001.

Our work significantly expands upon the WaveLAN study by Tang and Baker [TB00], which traced 74 computer-science users in one building for 12 weeks. Our study traces nearly two thousand users drawn from a general campus population, across 161 buildings for one academic term (11 weeks). It also expands upon the Metricom study by Tang and Baker [TB99, TB02] which traced a metropolitan-area network for seven weeks. Although that trace covers a wide geographical area and almost 25,000 users, our trace includes much more detailed information about amount and nature of the network traffic. The size, population diversity, and detail of our data collection offers extensive insight into wireless network usage. Although every environment is different, our study has characteristics common to both residential and enterprise deployments.

We next describe the environment of our study, the campus of Dartmouth College, and then detail our tracing methodology in Section 3. In Section 4 we present and discuss the most interesting characteristics of the resulting trace data. Section 5 compares our results with those of earlier studies, and Section 6 concludes.

This research was supported by Cisco Systems and by the Dartmouth Center for Mobile Computing.

2 The test environment

Our study examines usage of an 802.11b "Wi-Fi" network on the campus of Dartmouth College. The campus is relatively compact, with over 161 buildings on 200 acres, including administrative, academic, residential, and athletic buildings. 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 installed 476 access points from Cisco Systems, each an Aironet model 350[1], to provide 11 Mbps coverage to nearly the entire campus. Each access point (AP) has a range of about 130–350 feet indoors, so there are several APs in all but the smallest buildings. 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 161 covered buildings span 81 subnets, so in many cases a wireless client roaming from one building to another will be forced to obtain a new IP address. (Dartmouth chose not to construct a separate campus-wide subnet for the wireless network, unlike the Wireless Andrew project [BB97].)

Dartmouth College has about 5,500 students and 1,215 full-time professors. Most of the approximately 4,200 undergraduate students live on campus. Each is required to own a computer. Each year, approximately 1000 undergraduate students enter Dartmouth College, and most purchase a computer through the campus computer store. Of those purchases, laptops have become increasingly dominant in recent years: 27% in 1999, 45% in 2000, and 70% in 2001. Assuming that that students obtaining computers elsewhere choose laptops in the same fraction, and that in 1998 (for which no data is available) about 15% purchased laptops, about 40% of current undergraduates own laptops. All laptops purchased in 2001 had Wi-Fi built in, and over 1000 Wi-Fi cards have been sold over the past year to other users. In addition, all business-school students, and most engineering-school graduate students, own laptops. Clearly there is a large and growing population of mobile and wireless users.

3 Trace collection

We began collecting data in April 2001, when the first access points were installed. After preliminary study of the data in May 2001 [Ste01], and further tuning of the data-collection scripts in Summer 2001, we began full-scale data collection when students returned to campus in September 2001. In this paper we focus on the data collected during the eleven-week Fall 2001 term, Tuesday September 25 through Monday December 10, inclusive. Although we have data for about a week prior and about a month after, there was significantly less usage during vacation periods and so we limit our analysis to the active period.

At the beginning of the trace period there were 465 access points (APs). Eleven more APs were installed in the first month to bring the total to 476 by October 21. As we discuss below, it appears that some of the "installed" APs were not completely or correctly configured during the tracing period, however, which resulted in fewer APs represented in our data.

We used three techniques to collect data about wireless-network usage: syslog events, SNMP polling, and tcpdump sniffers.

3.1 Syslog

We configured the Cisco Aironet 350 access points used on the Dartmouth campus to transmit a syslog message for various events of interest. The APs published a syslog message every time a client (specifically, an 802.11b network interface card) authenticated, associated, reassociated, disassociated, or deauthenticated with the access point (see definitions below). The syslog messages arrived via UDP at a server in our lab, which recorded all 3,533,352 of them for later analysis.

Most APs contributed to the syslog trace as soon as they were configured and installed. Of the 476 APs, only 430 were represented in our trace. Although some appear never to have been used, many were misconfigured and did not send syslog messages. Furthermore, we have incomplete data for a few dates when the campus experienced a power failure, or when a central syslog daemon apparently hung up. Finally, since syslog uses UDP it is possible that some messages were lost or misordered. As a result of these spatial and temporal holes in the trace, some of our statistics will undercount actual activity.

Our syslog-recording server added a timestamp to each message as it arrives. Each message contained the AP name, the MAC address of the card, and the type of message:

Authenticated. Before a card may use the network, it must authenticate. Our network is currently configured to authenticate any card. We ignore this message.

Associated. After authentication, a card chooses one of the in-range access points and associates with that AP; 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 supports roaming. It appears that the firmware in many cards aggressively reassociates with new APs whenever

[1] www.cisco.com/warp/public/cc/pd/witc/ao350ap/prodlit/a350a_ds.htm

conditions are not good at the current AP. We saw many situations in which a card jumps back and forth among a small set of APs, as often as once per second. In some cases, where the APs were from multiple subnets, it is doubtful the user had much luck using the network!

Unfortunately, cards from some vendors apparently never use the "reassociate" protocol, and always use "associate" [Chr].

Roamed. Sent by the old AP when a card reassociates with a new AP. We ignore this message; because it depends on an inter-AP protocol below the IP layer, it only occurs when a card roams to another AP within the same subnet.

Disassociated. When the card no longer needs the network, it disassociates with its current AP. We found, however, that the syslog contained almost no such messages. More common were messages indicating that the AP chose to disassociate a card due to an error, such as a card's attempt to use an AP with which it was not currently associated.

Deauthenticated. The card's authentication status ends. While it is possible for the card to request deauthentication, this almost never happened in our log. In the normal case, the associated AP deauthenticates the card after 30 minutes of inactivity. In our log it is common to see several deauthentication messages from a widely roaming card, one message from each subnet visited in the session. In some cases an AP deauthenticates a card as a result of an error, such as a card attempting to use an AP with which it is not authenticated.

Note that although the APs emit an "authentication" message for each card, there is no user authentication; our network does not use MAC-layer authentication in the APs, or IP-layer authentication in the DHCP server. Any card may associate with any access point, and obtain a dynamic IP address. We thus do not know the identity of users, and the IP address given to a user varies from time to time and building to building. We make the approximating assumption to equate cards with users, although some users may have multiple cards, or some cards may be shared by multiple users. Throughout this paper we use the term "card" for precision, although with the intention that cards approximate users.

3.2 SNMP

Two Linux hosts in our lab used the Simple Network Management Protocol (SNMP) to periodically poll the APs; 451 of the 476 APs responded to our polls. We chose to poll every 5 minutes to obtain information reasonably frequently, within the limits of the computation and bandwidth available on our two polling workstations. Our trace period includes approximately 193,111,734 of these SNMP records. Unfortunately, we have incomplete data for the following dates: October 7, 9, and 12 (maintenance of our server), November 19 (unknown causes), and December 5 (a campus-wide power failure). We chose to entirely exclude those dates from our analysis, because most of our SNMP-based plots examine traffic per day, a number that would be polluted by "short" days.

Each poll returned the MAC addresses of recently associated client stations, and the current value of two counters, one for inbound bytes and one for outbound bytes. The AP does not reset the counters when polled, so we compute the difference between the values retrieved by one poll and the values retrieved by the next poll. The counters are 32-bit unsigned integers, and our computation properly handles counter roll-over. We ignore the result, however, in two 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. In the former 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 latter case, the AP (and its counters) were likely reset due to maintenance or a power failure.

Although each SNMP record contains a list of cards associated with the AP, we chose to use the syslog data for tracking cards because the syslog data provides the exact series of events for each card, whereas the SNMP polling data was less precise. We do use the list to compute per-card traffic statistics, and to help in our analysis of the sniffer data (below).

3.3 Sniffers

The syslog and SNMP traces allowed us to compute basic statistics about traffic, users, and mobility. To get a better picture of what the users were doing with the network, we used tcpdump to capture all of the packet headers on a selection of the APs around campus. Because of the volume of data, and privacy concerns, we recorded only packet headers. Because of the number and geographic distribution of APs, the structure of our network (many subnets, and switched Ethernet), and the volume of traffic, it was not possible to capture all of the wireless traffic. We placed a tiny Linux computer in each of four wiring closets around campus. In each case we attached this "sniffer" and the building's APs to a common hub, and attached the hub's uplink to a switch port on the campus network. With the sniffer in promiscuous mode, we used tcpdump to record the header of every packet passing by; in our later analysis, we focus only on the wireless packets (those with a wireless MAC as source or destination; the

list of such MACs was obtained from the SNMP data).

We chose four representative locations:

Sudikoff: the Department of Computer Science (6 APs). There is a hole in the data on December 5.

Brown: a dormitory with many first-year students (2 APs). There are holes in the data on September 28 through October 3, October 8, 20–24, 26, 28, 29, and 31, November 1–2, 4–5, and December 5.

Berry: the main campus library. Due to the size of the building and the switched nature of its network, were only able to sniff 5 of the 13 APs. There were holes in the data on October 26–31 and December 5.

Collis/Thayer: two buildings, the student center and dining hall, containing five cafes, several lounge areas, several meeting rooms, and some offices (total 9 APs). There were holes in the data on October 26, November 8–12, and November 26 through December 5.

Many of the holes were caused by power outages, in which case the sniffer lost power, but so did the the access point and nearby networking hardware. Thus there was no traffic to sniff during the power failure. Since, after power was restored, the sniffer no doubt took more time to boot than the access point and network hardware, we probably missed a small amount of data. Thus our statistics will slightly undercount the traffic on that date. The Collis sniffer, unfortunately, was more seriously affected by the power failures and required several days to repair.

Again, for any day in which there was a gap in the data, we discarded all data for that day before analysis.

3.4 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 must work with the data available, however, and we need precise definitions:

Card: a wireless network interface card, identified by MAC address.

Active Card: a card involved in a session (see below), during the hour, during the day, or at the place, in question.

Mobile Card: an active card that roams (see below) during the hour, during the day, or at the place, in question.

Session: The period between a session's begin time and session's end time (below).

Session length: the length of a session, in seconds: session end time minus session start time.

Session start time: when a card associates with an access point. Exception 1: any Associate messages that arrive less than *SessionThreshold* after the the preceding Associate or Reassociate message are treated as if they were a Reassociate message rather than starting a new session. Thus they may indicate a Roam (see below). Exception 2: for any card that never used Reassociate during our trace, we assumed that card is of the variety that uses Associate (within a session) to mean Reassociate, so we counted as roams any Associate arriving within an existing session.

Session end time: determined by one of three cases:

1. If a Deassociate or Deauthenticate message is received from the last access point used by the card (other such messages are ignored), the session is clearly over. If the reason is "Inactivity," and this message arrived more than 30 minutes after the session start time, we compute the session end time to be 30 minutes prior to this message's time. Otherwise, the session end time is this message's time.

2. As mentioned above, we treat some Associate messages arriving during an existing session as marking a new session. The time of this Associate message defines the end time of the current session and the start time of the new session. This rule was necessary because it appeared that many sessions did not end with a Disassociate or Deauthenticate message, either because the AP did not send the message or we did not receive it.

3. The end of the trace is reached. When this occurs, all ongoing sessions end at the last AP being used by the client and the session is assumed to end at the time of the last log in the entire trace.

Roam: a card switches access points 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 ignore roams that occur sooner than *RoamThreshold* after the session start, or the previous roam. Specifically, a rapid sequence of roams is reduced to one roam, with the time of the first reassociate but the AP of the last reassociate.

Extra-subnet roam: a roam to an AP in another subnet.

Intra-subnet roam: a roam to an AP in the same subnet.

Stationary session: a session containing no roams.

Mobile session: a session containing roams.

Extra-subnet session: a session containing an extra-subnet roam.

Intra-subnet session: a mobile session containing no extra-subnet roams.

Inbound: traffic sent by the card to the access point.

Outbound: traffic sent by the access point to the card.
These network-centric definitions of "in" and "out" are the reverse of Tang and Baker [TB00].

A note about the *SessionThreshold* mentioned above. On occasion, a card would Associate rather than Reassociate, apparently because the state machine on the card was out of sync with that on the AP [Chr]. 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 set *SessionThreshold* to 30 seconds, under the assumption that anything shorter is certainly not a new "session" in the eyes of the user.

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 that any reordering that affects causality is rare.

4 Results

We collected an enormous amount of data, and can present only a subset of the interesting characteristics in this paper. We begin with a few fundamental statistics:

- Days in trace: 77
- Cards: 1706
- APs: 476 (installed), 430 (syslog), 451 (SNMP), or 22 (tcpdump).
- Buildings: 161, which we divide into five categories: 82 Residence, 32 Academic, 6 Library, 19 Social, and 22 Administrative.

The residential buildings are mostly undergraduate dormitories and fraternities, but also include some Dartmouth-owned housing for faculty and staff, and a residential facility for the business school. All business-school students have laptops and (as the data shows) many are busy wireless users. The social buildings include dining facilities, the arts center, and athletic facilities (including a lodge at the ski area and a boathouse on the river).

In the rest of this section we present a series of questions about the network's usage, and our analysis based on the data. For each figure or table, we identify the data source as [syslog], [SNMP], or [tcpdump].[2]

4.1 Traffic

Perhaps the most fundamental questions about a new network involve how much it is used, and when:

[2]Readers using Acrobat with the PDF file can click on a figure or bibliographic reference to jump directly to that location.

Figure 1: [SNMP] **Daily traffic**. A date's bar appears to the right of its ticmark. Gaps in the plot represent holes in our data. Note that there is typically more outbound than inbound traffic.

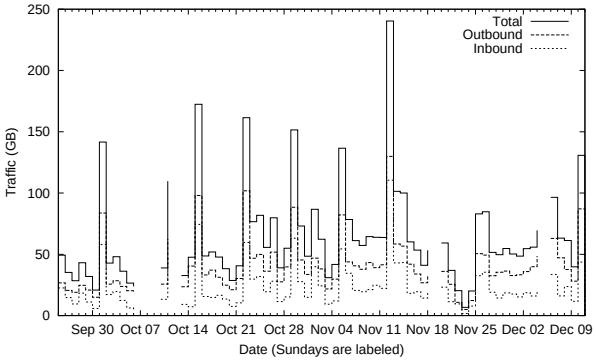
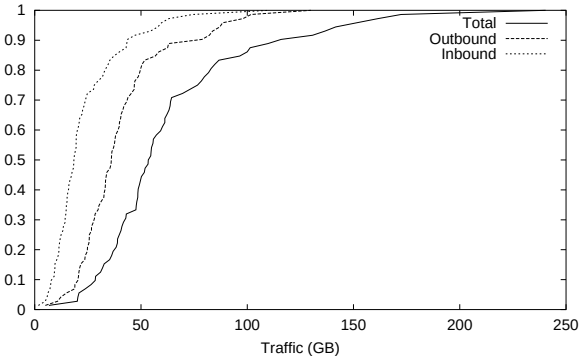


Figure 2: [SNMP] **Daily traffic, distribution across days.**



- How much traffic does the network handle?
- How much traffic per card?
- How does traffic vary across hours, days, weekdays?

Over the course of our study period we measured 4 terabytes of total traffic, although the daily traffic varied considerably. Figure 1 is a time series, and Figure 2 is a *cumulative distribution function*; we use the CDF format in all of our distribution graphs. On the busiest day the network moved over 240 GB, whereas the median daily traffic was 53 MB. There is a clear dip around the Thanksgiving holiday. Although there was usually less inbound traffic than outbound traffic, given the nature of the protocols used (Section 4.6), there were days where inbound data dominated: the proportion of inbound data varied daily between 18 and 89%.

In Figure 3 we normalize the data by the number of cards active in that day. This presentation flattens the curve somewhat, although there is still a wide variation in daily activity.

These figures show a reasonably strong weekly pattern with some surprising peaks on Mondays. In Figure 4 we see the weekly patterns more clearly. Friday and Saturday are the quietest days, as students relax, but Sunday picks

Figure 3: [SNMP] **Daily traffic *per card*.**

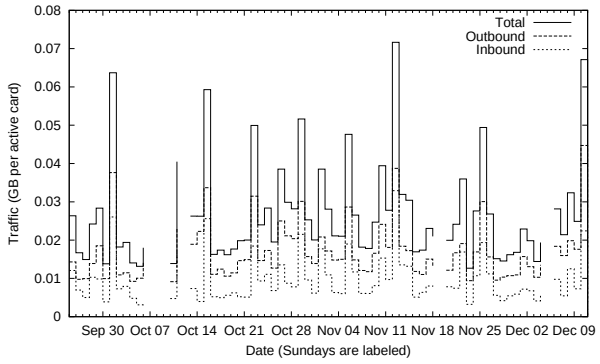


Figure 4: [SNMP] **Average daily traffic, by weekday.**

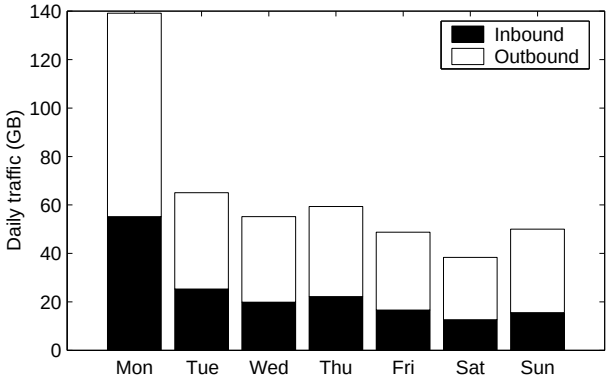


Figure 5: [SNMP] **Average hourly traffic, by hour.**

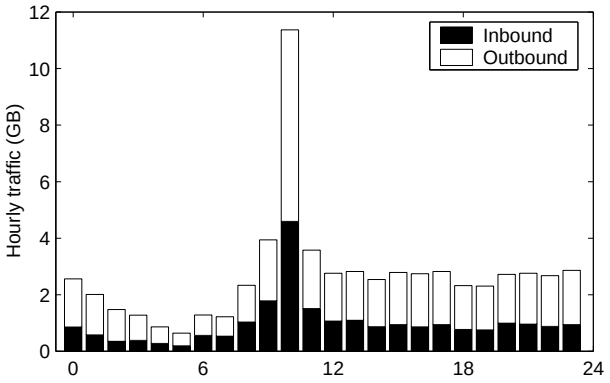


Table 1: [syslog] **Common vendors** of network interface cards, as identified by the vendor component of the MAC addresses.

Number	Vendor
624	Lucent/Agere
536	Apple Computer
489	Cisco/Aironet
57	*Other (15 brands)*
1706	**Total**

up as students begin their homework. Monday's average is skewed by some some especially busy Mondays; these peaks are likely a result of weekly actions, such as backups (inbound) or software updates (outbound).

We further our speculation in Figure 5 that this regular Monday activity occurs between 9–11am, when we see high maximum traffic. Otherwise Figure 5 shows relatively constant traffic throughout the afternoon and evening, tailing off through the night when students finally go to sleep, and rising again as employees return to work. Other than the 10am spike, we do not see the classic diurnal bell curve that one might see in a typical workplace, because our environment is a mixture of residential and academic uses.

4.2 Users and user mobility

As mentioned in Section 3, we did not (and could not) track *users*, but since for the most part each card is associated with one user, and most users have just one card, we examined cards as if they represent users. We ask:

- How many cards are there? From which vendors?
- How many days is each card active?
- How many APs does a card visit?
- How many buildings does a card visit?

There were 1706 unique MAC addresses seen in our syslog trace, most from a few common vendors (Table 1). Dartmouth's campus computing store resells exclusively Apple and Dell computers, and as of 2001 all laptops sold to first-year students have wireless cards built-in: Agere (part of Lucent) cards in the Dell laptops, and Apple Airport cards in the Apple laptops. The store also sells Cisco (Aironet) wireless PC cards, an option for those with older laptops.

Users varied in the number of days that they used their cards, from only once to every day in the 77-day trace (Figure 6). Many users are students, living on campus, and it is not surprising to see some with wireless laptops on their dorm-room desk, always on-line. Interestingly, the distribution is roughly uniform between one and 77 days, with a median of 28 days.

The graph also shows that few cards move around much, with a median of five buildings and nine APs, and no card visiting even half of the entire network. Indeed, 18% of the cards spent all their time in one building. Clearly, most users limit their activity to a few key sites in their daily routine. We expect to see this pattern change as more small devices, such as PDAs with Wi-Fi installed on a CompactFlash card, ease mobility.

Figure 6: [syslog] **Activity per card, distribution across cards.** Maximums: 77 days, 62 buildings, and 147 APs. Medians: 28 days, 5 buildings, and 9 APs.

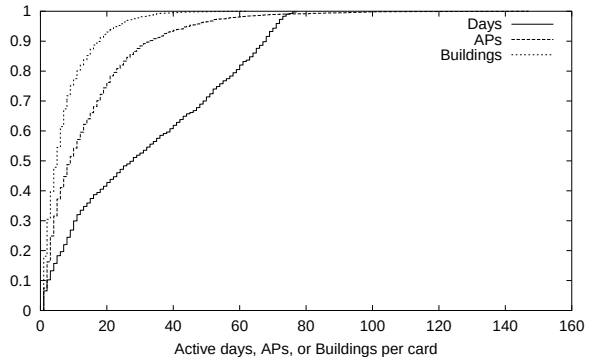


Figure 7: [syslog] **Number of active and mobile cards per day.** A date's data appears to the right of its tick-mark.

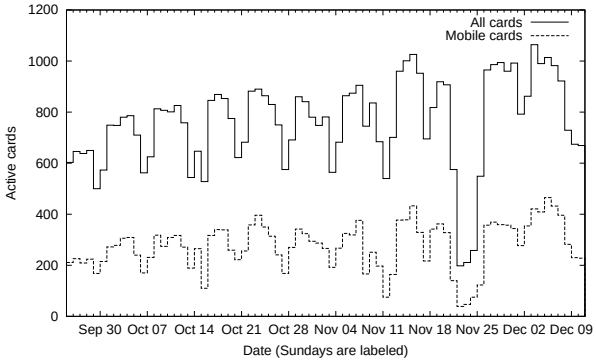
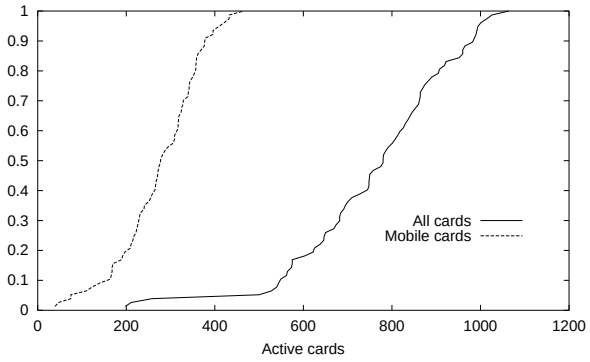


Figure 8: [syslog] **Number of active or mobile cards per day, distribution across days.** Medians are 780 (all) and 278 (mobile).



4.3 Card activity

Now that we have seen the network from the card's perspective, we examine the cards from the network's perspective:

- How many cards are active?
- When are cards active?
- How long are sessions?
- How many sessions are started each day?
- How are sessions distributed among buildings?
- How many sessions are mobile? extra-subnet?
- How often do cards roam per session?

Although there were 1706 cards seen in our traces, not all were active every day. Figure 7 shows the number of cards active in each day of our trace period. Clearly visible are the Thanksgiving holiday, weekly cycles, and a tail-off at the end of the term. Also visible is a slow trend toward more active cards per day, as more users obtain wireless capability and choose to use it more often. Here we define "active" to mean any card that is associated with an access point, regardless of whether the user is actually using the computer or network. The plot also shows mobile cards: an active card is "mobile" in a day if it roams during any session that day. (Note that a card may visit several APs during a day, in separate sessions, but not be "mobile" unless it roams *during* one of those sessions.)

In another view, Figure 8 shows the distribution of the number of active cards and mobile cards in any given day. Almost half of our card population was active on a typical day, and about a third of those were mobile.

The visible weekly cycle of Figure 7 is reinforced in Figure 9, which we believe reflects a typical student pattern of activity, hustling to complete their work early in the week, relaxing on Friday and Saturday, and picking up again on Sunday.

Figure 10 shows diurnal patterns. Again, this pattern matches a mixture of workplace and residential patterns, with the bulk of the activity during the weekday, particularly the afternoon, substantial activity during the evening, and a slow decline in activity through the wee hours of the morning. With most office workers away on weekends, the weekend mid-day activity is lower, but due to the residential population the evening and overnight hours remain about the same on weekends and weekdays. We reach similar conclusions about mobile cards, not shown.

Figure 11 demonstrates the different patterns, and relative activity, of different categories of buildings on campus. Residential activity dominates. Residences and social spaces tend to be used more in the evening hours, whereas academic and administrative buildings are active during the work-day, and libraries are somewhat in-between. Figure 12 shows far fewer mobile cards, particularly during the overnight hours.

Sessions. We are interested in when, and for how long, users choose to use the wireless network. In the preceding section we define a "session," intuitively, to be the period of activity with the network, although it is difficult to ac-

Figure 9: [syslog] **Number of active or mobile cards per weekday.** The curve shows the mean, while the bars show minimum and maximum. The two curves are slightly offset so the bars are distinguishable.

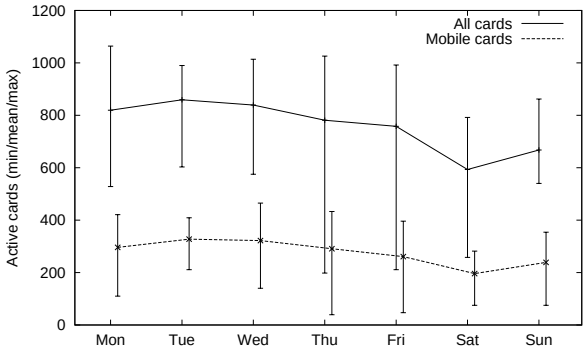


Figure 10: [syslog] **Number of active cards per hour.** The number of active cards for each hour of the day, separately for weekdays and weekends. The curve shows the mean, while the bars show minimum and maximum. The two curves are slightly offset so the bars are distinguishable.

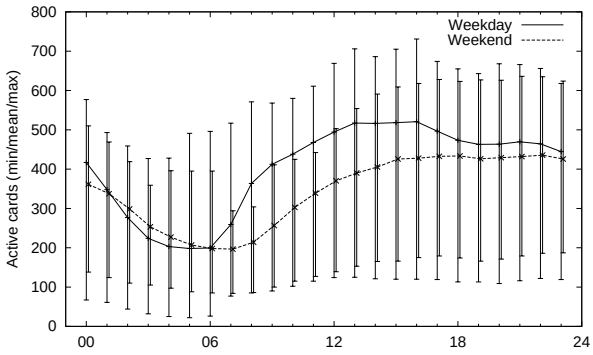


Figure 11: [syslog] **Mean active cards per hour, by category.** A card visiting multiple building categories within an hour was counted once for each category it visits.

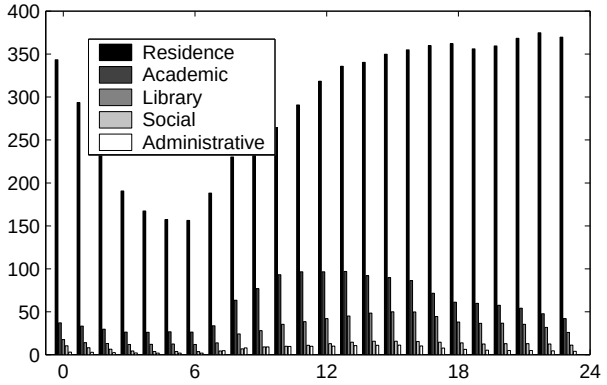
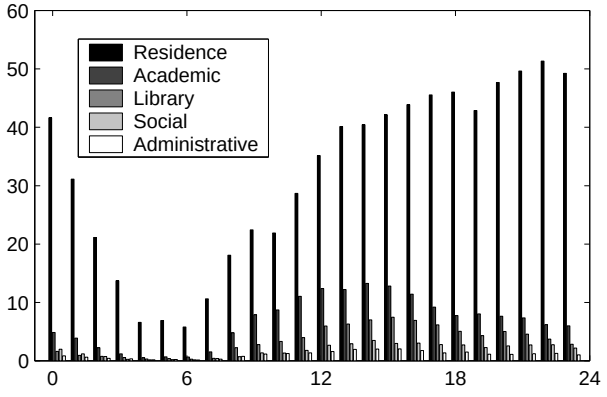


Figure 12: [syslog] **Mean mobile cards per hour, by category.** A card visiting multiple building categories within an hour was counted once for each category it visits.



curately detect the beginning and end of all sessions given the syslog data. We believe that our results are a reasonable approximation of the notion of a user session.

Our data (Figure 13) shows that most sessions are short. The median session length was 16.6 minutes, and 71% of sessions finish in less than one hour. Given that students move frequently from class to class to dining to dorm, and like to check email in between, these numbers are reasonable.

On the other hand, there were a few sessions that were very long (69 days in one case). These extremely long sessions are likely artifacts of holes in the syslog data, in which we lost the session-ending message. There are many short sessions: 27% of sessions last less than a minute. Despite our 30-second SessionThreshold, our session-begin definition was apparently too liberal. Nonetheless, this data begs the question about why the cards associate so quickly and frequently. Examination of sample sessions show many instances in which a card Associates with an AP despite (from our reading) the fact that it is already associated, an indication that the state machine in the card and in the AP are out of sync [Chr]. Although further study is necessary, it appears that there is substantial room for improvement in the card firmware and possibly in the card-AP protocols.

Although most (83%) sessions are non-mobile, mobile sessions do include one or more roams. Figure 14 shows the distribution of the number of roams during mobile sessions. Most mobile sessions were short and roamed infrequently (the median is two roams). Nearly 60% of mobile sessions roamed only within one subnet. Unfortunately that means that over 40% roamed across a subnet boundary, which breaks connections and forces the user to obtain a new IP address. Some sessions roamed extremely frequently: one session roamed over 9,000 times!

So, why do cards reassociate so frequently? The cards aggressively search for a strong signal, and in an environment with many APs and overlapping cells, cards will roam frequently [Chr]. Either card firmware needs to be

Figure 13: [syslog] **CDF of session duration (truncated to 1 day).** The longest session measured 69 days, although that is probably an error due to holes in our data. The median is 16.6 minutes.

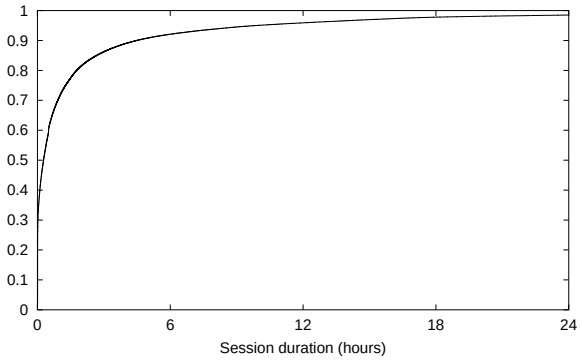


Figure 14: [syslog] **Roams per mobile session, distribution across sessions.** This graph is truncated. The maximum is 9,343 roams. About 17% of sessions were mobile.

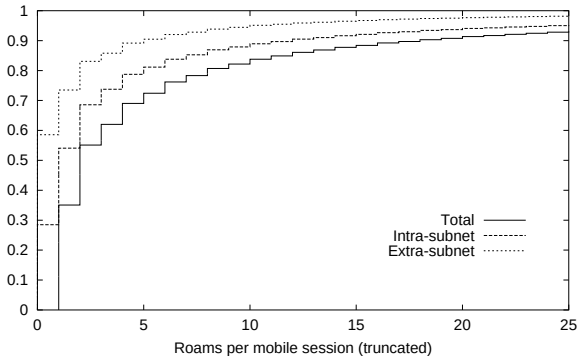


Figure 15: [syslog] **Number of roams, by RoamThreshold.** Over the entire trace period. Notice the y scale.

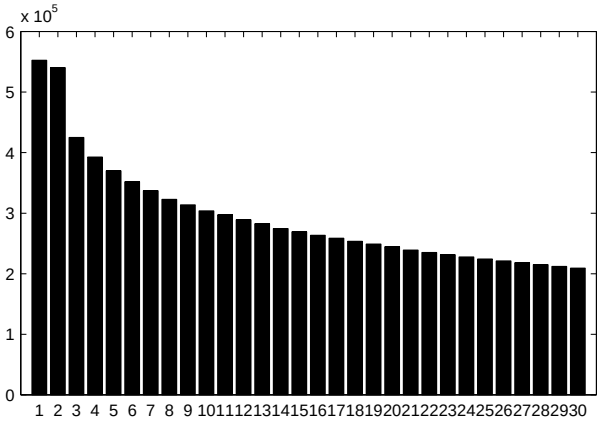
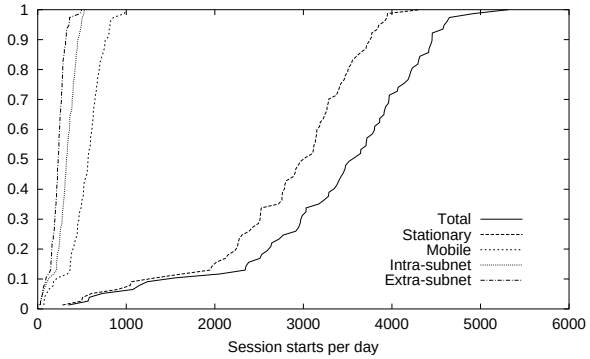


Figure 16: [syslog] **Number of session starts per day, distribution over days.**



less aggressive, or our environment needs to reduce cell overlap, to reduce the roaming, reduce the resulting load on the network, and give better service to the user. Furthermore, since it is expensive to deploy a single campus-wide subnet for the wireless network [Hil99, HJ96], Mobile IP [Per99] or similar services are required to support seamless roaming.

We experimented with the *RoamThreshold* parameter, which ignores any roam if it occurs less than *RoamThreshold* seconds after the session start or a previous roam, and the results are in Figure 15. Clearly, this parameter filters out many roams. It is not clear, however, what *RoamThreshold* would be appropriate for general use in our analysis. If we set the threshold too high, we may mask roams caused by real user movement. If we set it too low, our data represents the "jumpy" nature of the real cards. For the purposes of this paper, we chose threshold 3, because of the step at that point in the figure. As it turns out, within the range 0–30 seconds, the choice has little effect on the graphs presented in this paper. Perhaps most significant is the measure of roams per session (Figure 14). This truncated plot does not change, but the maximum value drops to 1574 for threshold=30 rather than 19,902(!) for threshold=0. In short, the threshold removes the short-term jumpiness but does not affect the conclusions drawn from Figure 14.

Figure 16 is another view of daily network activity, in which we count the number of sessions started in each day, and here present the count as a distribution across days. The median is 3582 sessions, or 570 mobile sessions. Given the number of short sessions, these numbers are not surprising. Although most session starts are in the dominant category (residence), it is more interesting to examine the relative mobility of users in different building categories. Figure 17 shows that sessions in academic or administrative buildings tend to be more stationary, and that those in libraries tend to have slightly more extra-subnet roams. The latter may have more to do with the configuration of the libraries and subnets than any real physical mobility.

Figure 17: [syslog] **Number of session starts (normalized), by category.**

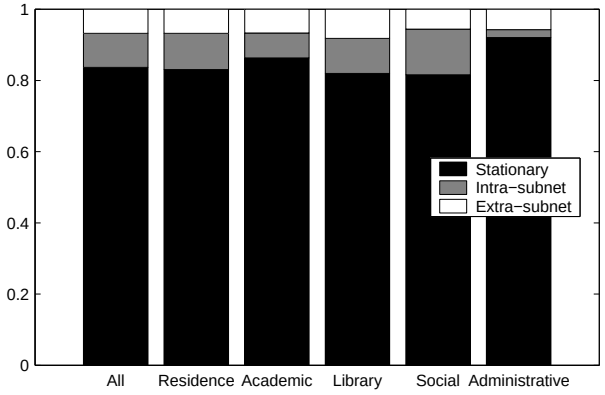


Figure 18: [syslog] **Maximum cards per hour, for the busiest APs.** Ranked by their busiest hour, in terms of active cards.

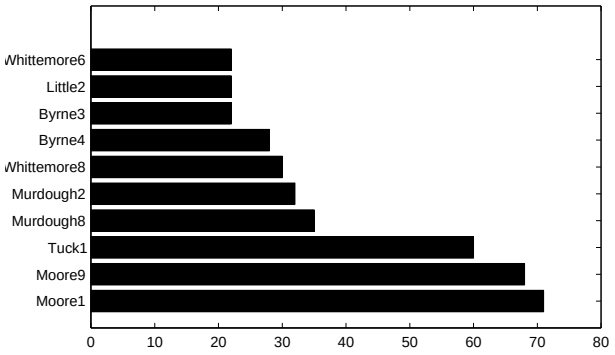
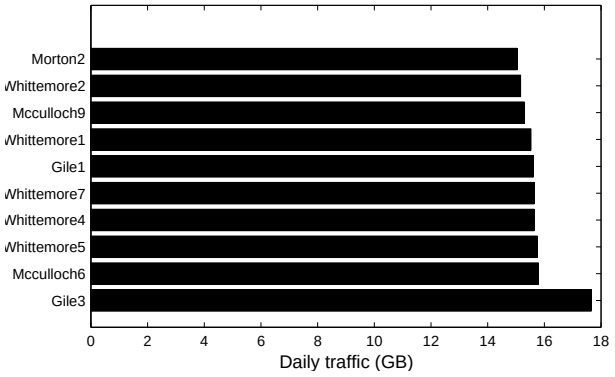


Figure 19: [SNMP] **Maximum daily traffic, for the busiest APs.** Ranked by their busiest hour, in GB.



4.4 AP activity

We now examine network activity in terms of the APs:

- How many APs are there?
- When are APs active?
- How does activity vary across APs, and which are most active?
- How does traffic vary across APs, and which have most traffic?

There were 476 APs installed by the end of the study. The data in this section are based on the 430 APs in the syslog trace and the 451 responding to our SNMP polls.

A detailed identification of the busiest APs is perhaps only of internal interest at Dartmouth College, and in any case we examine the related question about the busiest buildings in the next subsection. The APs with the most active cards in their busiest hour, were those located near large lecture halls; in its busiest hour, the busiest AP had 71 active cards. The traffic was elsewhere, however: the APs with the largest maximum and average daily traffic were from residences.

Figures 18–20 show the ten "busiest" access points, for three different definitions of busy. Figure 18: Moore1 is the access point covering three large lecture halls in the basement of Moore. Figure 19: Gile is a dorm, and apparently one or more of its users occasionally caused a lot of traffic. Figure 20: Brown is a dorm with many first-year students (recall that 70% of first-year students own wireless laptops), and Whittemore is the residential facility in the business school (where students are required to own laptops).

Figure 21 shows the variation in the number of APs active each day. Clearly visible are the weekly cycle, the Thanksgiving holiday, and a general trend to use more APs, as the number of cards increased and as people used the network more. In another view (Figure 22), we see that of the 430 access points never were fewer than 168, or more than 350, active in any one day. A typical day saw 291 active access points.

In Figures 23 and 24 we see that the number of active APs follows a pattern similar to the number of active cards.

Over the life of the trace, the APs varied widely in the amount of traffic they handled (Figure 25), with the median AP handling an average of only 39 MB per day, while the busiest AP handled an average of over 2 GB per day.

4.5 Building activity

An examination of buildings allows us to classify the most active locations on campus.

- How many buildings are there?
- When are buildings active?
- How does activity vary across buildings, and which are most active?
- How does traffic vary across buildings, and which have most traffic?

Figure 20: [SNMP] **Average daily traffic, for the busiest APs.** Ranked by their busiest day, in GB.

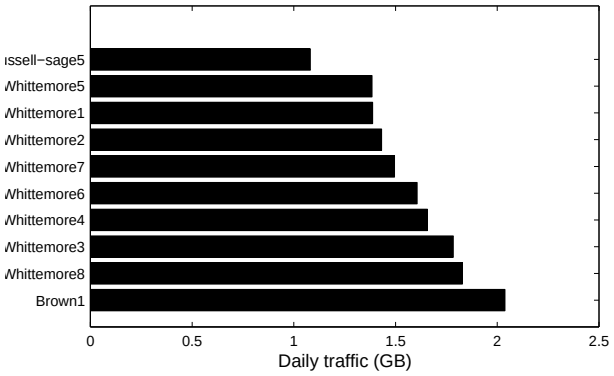


Figure 21: [syslog] **Number of active APs per day.** A date's data appears just to the right of its tick-mark.

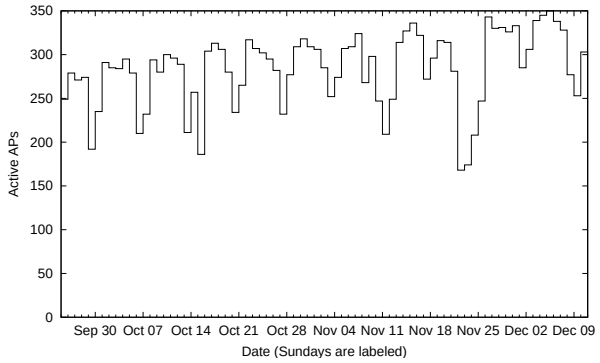


Figure 22: [syslog] **Number of active APs, distribution across days.** Note the x-axis is not based at zero.

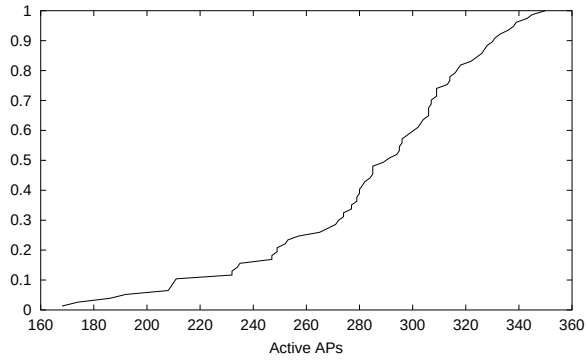
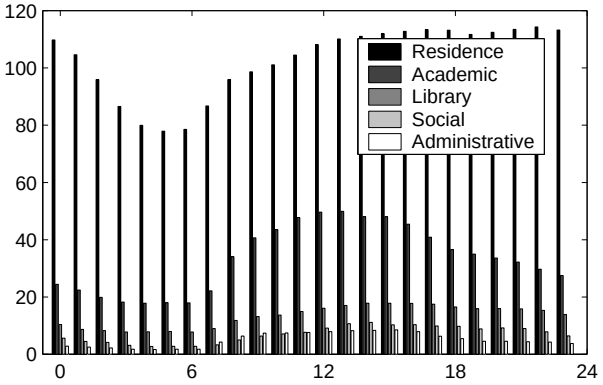


Figure 23: [syslog] **Mean active APs by hour, by category.**



- How does activity vary across building categories?
- How does traffic vary across building categories?

There were 161 buildings with installed APs, ranging widely from huge central libraries to tiny houses, and even a shed at the tennis courts. Although Figure 26 shows that the bulk of the traffic was seen in the residential buildings (averaging 48 GB per day), when normalized by population size (active cards, in Figure 27) or by building size (number of APs, in Figure 28) we see somewhat more balanced traffic. Residential users spend more hours in residences than most people spend in other buildings, accounting for some of this difference.

The building with the largest average daily traffic (Figure 29) was, by far, the business-school residence Whittemore. About a third of the residents have a wireless laptop, and there is clearly a culture that encourages wireless usage. Cummings is the engineering school, McCullouch is another business-school building, and Murdough is the library between the two. The other buildings are dormitories with large populations of first-year undergraduates.

Figure 30 normalizes by the number of APs, to reduce the importance of larger buildings, but Whittemore still dominates. The others are all undergraduate dormitories, particularly those full of first-year students. On the other hand, when we normalize traffic by the number of active cards (Figure 31), a sorority and several academic and administrative buildings (Strasenburgh, Sls, Silsby, Gilman, Kiewit) top the list. Gile and Smith are dorms.

Examining the busiest day for each building (Figure 32), the business school (Whittemore and McCullough) clearly dominates, but some academic buildings (Rockefeller and Silsby) and administrative buildings (Sls) appear. The others are dorms.

In Figure 33, the buildings with the busiest hour, in terms of the number of active cards, are mostly buildings with large lecture halls (Moore, Murdough, Tuck, Byrne, and Cummings), the main campus library (Berry), and some residences (Whittemore, Hinman, McLane, and Buchanan). Clearly network designers need to plan carefully for such large concentrations of usage.

Finally, in Figure 34, we see the buildings with the largest number of cards visiting over the entire trace. These are all large buildings where you expect a diverse population: libraries (Baker, Berry, Murdough, and Sanborn), social and dining spaces (Hop, Collis, and Thayer), and three buildings with large lecture halls (Gerry, Dartmouth, and Bradley) frequented by students in introduc-

Figure 24: [syslog] **Number of active APs per hour, distinguishing weekdays from weekends.** The curve shows the mean, and the bars show the minimum and maximum. The two curves are slightly offset so the bars are distinguishable.

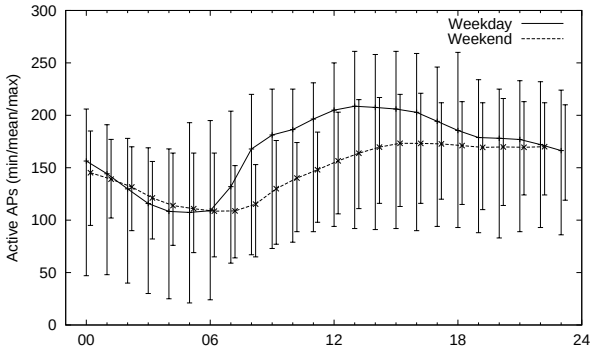
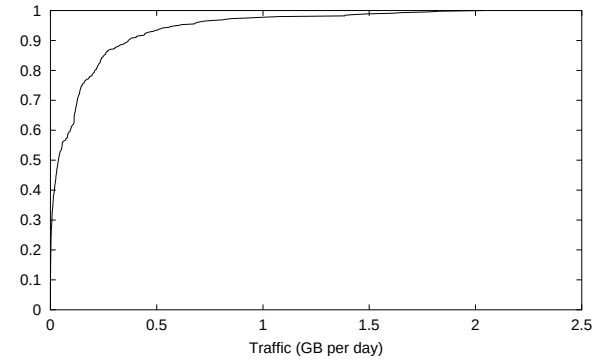


Figure 25: [SNMP] **Average daily traffic, distribution across APs.** Median is 39 MB.



tory courses. Figure 35 shows, though, that half of the buildings saw fewer than 64 users over the life of the trace, less than Moore saw in a single hour.

The number of active buildings followed a pattern similar to the number of active APs and number of active cards, as seen in Figure 37, although the variation is dampened somewhat as we consolidate the activity into buildings. Figure 36 demonstrates a similar effect. Interestingly, of the 82 residences only about half are active in any given hour.

4.6 Protocols

Although the sniffer data covers only four buildings and 22 APs, it covers a variety of populations (library, dormitory, student center, and academic computer science). Above, we examine questions about where, when, and how much people use the wireless network; now, we ask about *how* they used the network:

- Which protocols are the most commonly used?
- Which protocols consume the most traffic?

Figure 26: [SNMP] **Average daily traffic, by category.**

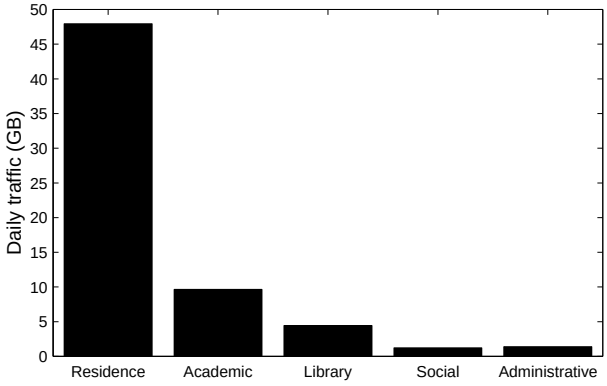
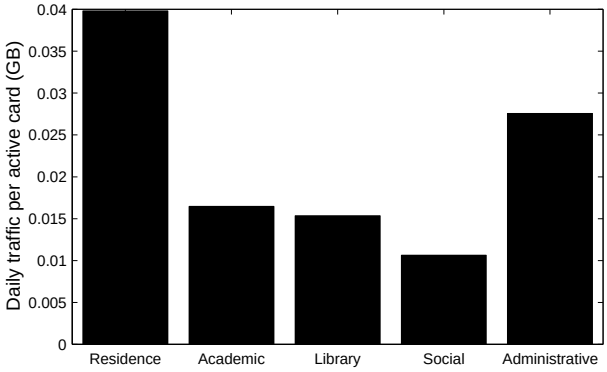


Figure 27: [SNMP] **Average daily traffic *per card*, by category.**



- For each protocol, how much data flows each way?

We collected data about only IP-based protocols.[3] We monitored nearly 183 GB of traffic in all four sniffers, 100 GB of which was from the dormitory (recall from Figure 29 that Brown is one of the busiest buildings on campus, and was a fortuitous choice for sniffing). Although we saw a tiny amount of PIM, RSVP, IGMP, and ICMP, the bulk of the traffic was UDP (3.3%) and TCP (94.8%). Table 2 shows the details. Although Brown saw only 55 cards, and Sudikoff (Computer Science) saw only 110 cards, the Collis student center saw 434 and Berry Library 556, as they are larger, public spaces with a diverse population.

We were able to identify 82 different TCP or UDP protocols in the IP packets we sniffed, by recognizing well-known port numbers. This technique is an approximation, of course, since it is possible that some applications use a "well-known" port for other purposes, but it provides a good overall estimate. Ten protocols account for 97.8% 179 GB of the total 183 GB traffic, as shown in Figure 38.

[3] Although our campus does include many Appletalk users, we expect nonetheless that IP dominates the wireless traffic. We hope to trace non-IP protocols in a future study.

Table 2: [tcpdump] **Common IP protocols** seen in our four sniffers. RSVP appears to be zero due to rounding.

Protocol	Total GB		Berry		Brown		Collis		Sudikoff	
PIM	0	0.0%	0.000	0.0%	0.001	0.0%	0.000	0.0%	0.000	0.0%
RSVP	0	0.0%	0.000	0.0%	0.000	0.0%	0.000	0.0%	0.000	0.0%
IGMP	0	0.0%	0.001	0.0%	0.000	0.0%	0.000	0.0%	0.000	0.0%
ICMP	0	0.0%	0.014	0.1%	0.006	0.0%	0.025	0.1%	0.012	0.0%
UDP	6	3.3%	2.819	12.0%	0.857	0.9%	0.275	1.3%	2.042	5.2%
TCP	178	96.7%	20.650	87.9%	98.322	99.1%	21.167	98.6%	37.530	94.8%
Total GB	183.720	100%	23.484	100%	99.185	100%	21.467	100%	39.584	100%

Figure 28: [SNMP] **Average daily traffic *per AP*, by category.**

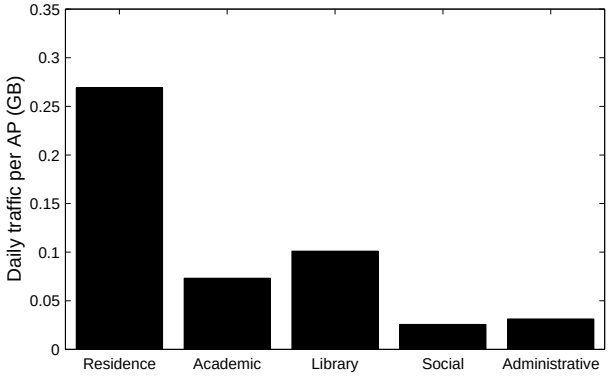


Figure 29: [SNMP] **Average daily traffic, for the busiest buildings.** Ranked by daily traffic, in GB.

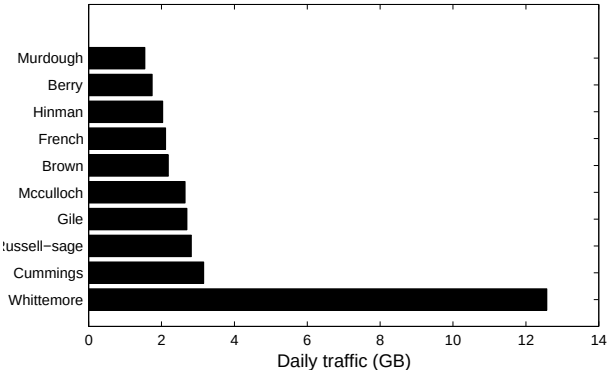


Figure 30: [SNMP] **Average daily traffic *per AP*, for the busiest buildings.** Ranked by daily traffic, per AP, in GB.

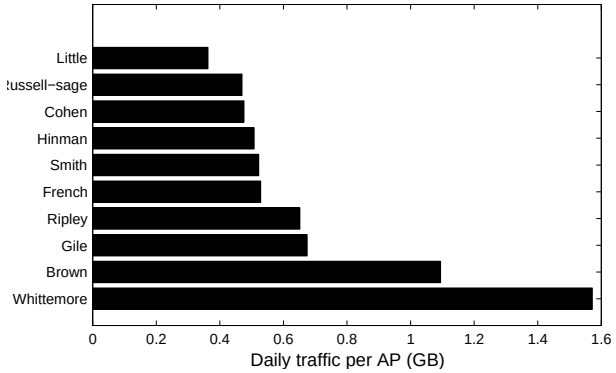
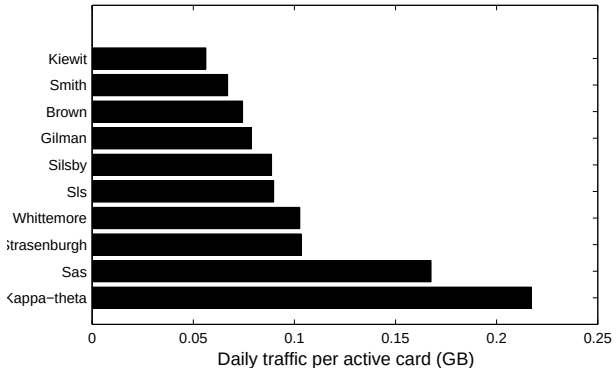


Figure 31: [SNMP] **Average daily traffic *per card*, for the busiest buildings.** Ranked by daily traffic per card (GB).



The symmetry of this traffic is explored in Figures 39 and 40. These protocols were:

http (90 GB), including both http and https, and some other common http ports (such as 8000). Clearly, web browsing is a significant fraction of any network traffic today. It is not dominant everywhere, however: outside Brown, there was less http traffic than "dantz" or all unidentified protocols (see below). Although most http traffic is outbound, there is substantial inbound traffic in Brown, indicating that some residents may be running a web server on a wireless computer.

unidentified (45 GB): all packets involving port numbers not identified in `/etc/services` are lumped into this category. We were surprised by the volume of traffic not clearly attributable to well-known ports. We intend to investigate the distribution of port numbers in this category, to determine whether there may be a few common (but unknown to us) protocols.

dantz (30 GB), a protocol for the Retrospect backup product from Dantz corp., in common use here for office Macintosh computers. Collis and Sudikoff have several such offices, and the "dantz" protocol dominates the traffic seen

Figure 32: [SNMP] **Maximum daily traffic, for the busiest buildings.**

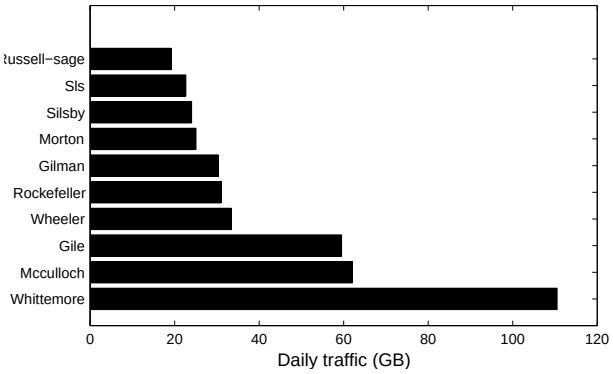


Figure 33: [syslog] **Maximum cards per hour, for the busiest buildings.** Ranked by their busiest hour (in number of active cards).

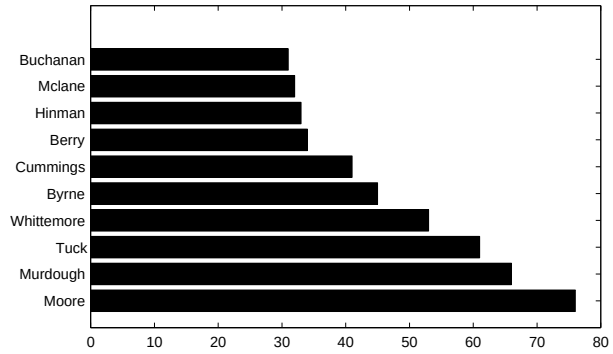


Figure 34: [syslog] **Number of active cards per building, for the ten most popular buildings.** Ranked by the number of unique cards visiting that building, over the whole trace.

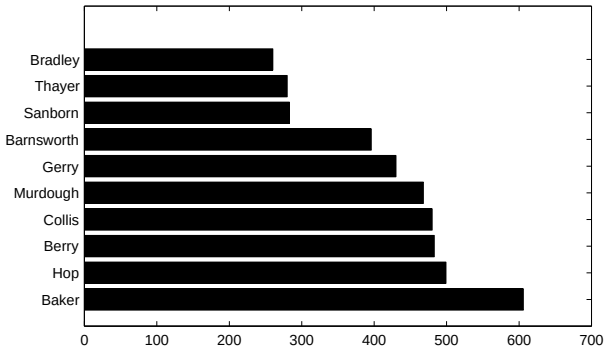
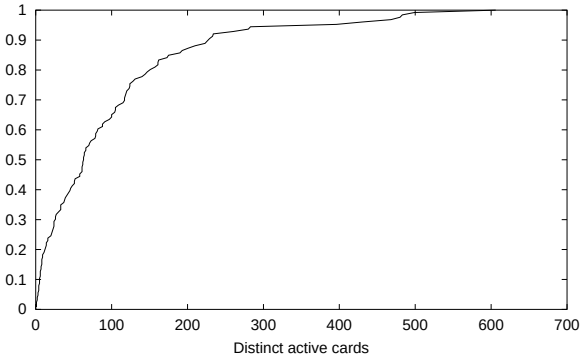


Figure 35: [syslog] **Number of active cards per building, distribution over buildings.**



by those sniffers. The traffic is dominantly inbound, of course, as wireless clients are backed-up to a wired server. While it was an unexpected frontrunner, a few backups conducted periodically can easily account for its volume. Indeed, we saw weekly spikes in the sniffed data, likely caused by these backups.

blitzmail (3.9 GB): BlitzMail is a locally developed email client, with a custom protocol, in ubiquitous use outside Sudikoff (computer scientists tend to use more traditional mail clients). The high volume is no doubt the result of large enclosures.

ftp (3.7 GB), including all variants of the common file-transfer protocol, including ftp, ftp-data, ftplog, bftp, tftp, and sftp. Curiously, there was nearly an even split between inbound and outbound data, although in each sniffer (not shown) it is more skewed toward either inbound (Collis and Berry) or outbound (Brown and Sudikoff).

netbios-ssn (3.0 GB), a Windows session protocol that supports Windows print and file sharing, including Samba.

ssh (1.4 GB), a secure-shell protocol that supports interactive logins, secure remote file copy, secure X-windows sessions, and other secure tunnels. Most of this traffic occurred in the computer-science building where these activities are more common than in the general population. The dominance of inbound traffic suggests that scp is often used for uploading files.

snmp (0.89 GB), the Simple Network Management Protocol. We are uncertain why there would be significant use of this protocol among personal laptops. Most of the outbound traffic was in Berry library, and most of the inbound traffic was in Sudikoff.

afpovertcp (0.87 GB), a tunnel for AppleTalk Filing Protocol over TCP, allowing Macintosh computers to mount other Macintosh disks, over TCP/IP. Mostly used for downloading files, it appears.

instsrv (0.62 GB), seen mostly in Brown (with 34 users). Although instsrv is listed as a protocol used by the "network install service," its port number (1234) is also used by several Trojan-horse programs. It is possible that this traffic represents clients that have been hacked.

Figure 36: [syslog] **Number of active buildings per day.** A date's data appears just to the right of its tick-mark.

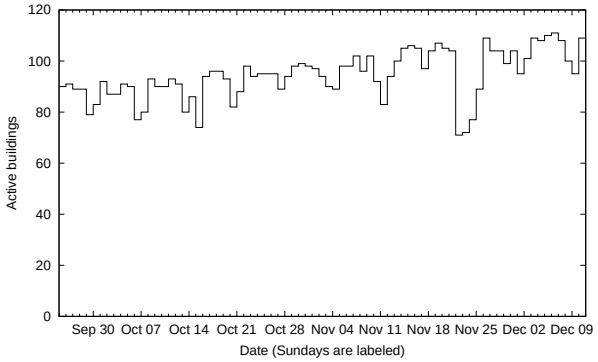


Figure 37: [syslog] **Mean active buildings per hour, by category.**

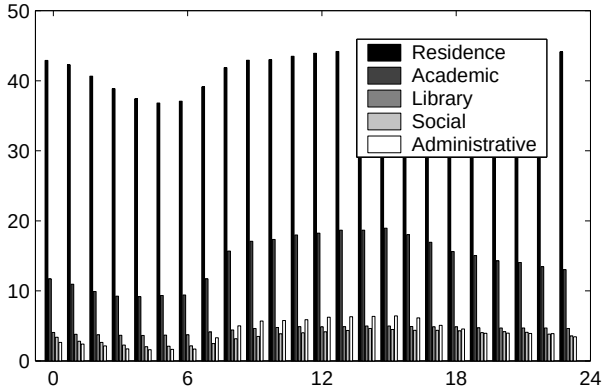


Figure 38: [tcpdump] **Total traffic, by TCP or UDP protocol.**

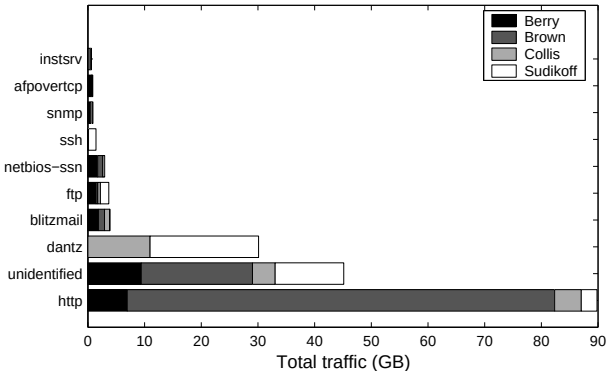
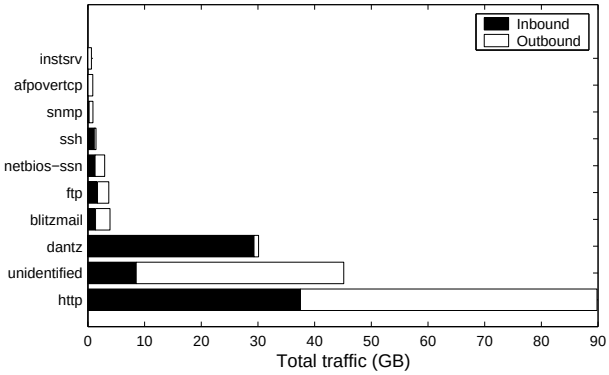


Figure 39: [tcpdump] **Total traffic, by TCP or UDP protocol.**



With the exception of SNMP, all of the above protocols are commonly used for file transfer, which accounts for their dominance in this ranking based on volume. We are preparing an analysis of the number of connections made for each protocol, regardless of the amount of data transferred, which should give another interesting perspective on how people use this network.

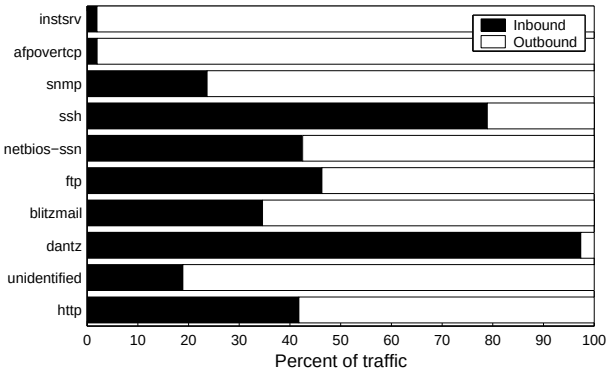
While the details of our protocol distribution may be specific to Dartmouth, we expect that others in academic environments will see approximately the same set of activities dominating: web, email, backup, file transfer, and file sharing.

5 Related work

Our study is the largest and most comprehensive characterization of wireless LAN users to date. In three separate studies, Tang and Baker have previously characterized wireless-network usage. In 1998 they used tcpdump in a limited study of eight laptops over eight days [LRT+98], focusing on the number of times the laptops switched between wired and wireless, and on the latency encountered by packets. They note that users did tend to behave differently on the wireless network than on the wired network, due to extremely high latencies. In 1999 they characterized the users of the Metricom Ricochet network, a wireless metropolitan-area network (MAN) service [TB99, TB02] This study is notable for its size (24,773 clients and 14,053 access points) and duration (about seven weeks). Due to the limitations of the data available, their analysis focuses on network activity and client mobility. Finally, in 2000 they use tcpdump and SNMP records to characterize the activity of 74 wireless users in the Stanford Computer Science Department, over a 12-week period [TB00]. While this study is similar to our own, our population is much larger and more diverse, the roaming patterns are more complex than a single subnet in a single building. We have syslog data that allows more precise measurements of roaming, but we do not have DHCP data that allow us to associated MAC addresses with users. Although we do not have sniffer data for the entire population, we do have it for four buildings rather than one. We do not have any way to measure latency, and so far have not analyzed geographic mobility analysis or clustered users.

The Wireless Andrew project at Carnegie-Mellon University created the first large WaveLAN installation, and

Figure 40: [tcpdump] **Total traffic, by TCP or UDP protocol, normalized.**



their papers discuss the design and deployment of that network [BB97, Hil99, HJ96]. Although they hint of plans for a usage study [BB97], there are as yet no published results.

Another recent study used data from Bell Mobility's Personal Communications Services (PCS) cellular network to study the characteristics of customers using WAP web browsers on their cell phones [KBZ+00]. They traced the network during seven months, using tcpdump to capture packets at the WAP gateway. Unfortunately, they were unable to identify unique users or phones, but the number of IP addresses assigned in any given day increased to about 400 by the end of the trace. The PCS network reassigned an IP address whenever the browser was idle for 90 seconds, so the session lengths were quite short (average 3.38 minutes). Otherwise, the usage followed the expected weekly and daily patterns in amount of traffic and number of users.

6 Conclusion

We conducted the largest trace-based study of wireless LAN users to date, in an effort to understand patterns of activity in the network. The activity and traffic varied widely from hour to hour, day to day, and week to week. While we do see clear daily and weekly patterns, they reflect a mixture of a residential campus and an academic workplace, including more overnight usage than might be common in enterprise WLANs. We found that many wireless cards are extremely aggressive when associating with access points, leading to a large number of short "sessions" and a high degree of roaming within sessions. About 17% of sessions involved roaming, and of these "mobile sessions" about 40% involved roaming to a different subnet. From anecdotal evidence, these extra-subnet roams often occur when when the user is stationary, leading to failures of IP traffic.

Network designers should note the high variance in the activity of buildings, access points, and cards, over both time and space. We need new solutions to prevent cards from roaming too frequently, without sacrificing coverage. We need network-layer [Per99] and application-layer solutions [MTK02] to support multi-subnet roaming. Finally, note that the traffic is not definitively dominated by outbound or inbound traffic. The ratio varied significantly from day to day, building to building, and protocol to protocol. This conclusion argues against any design with asymmetric bandwidth.

In the early stages of the wireless project, the staff at Dartmouth College debated whether it would be important to provide wireless coverage in the dormitories, which were already wired with at least one port per resident. Our data shows that the bulk of wireless activity occurs in the residences. Furthermore, for wireless network connectivity to be useful to a mobile user, it needs to be pervasive, allowing the user to grab their laptop on the way out the door, confident that there will be network access wherever they may go. Nonetheless, we saw that most users visited few APs and buildings over the life of the trace, and most users were stationary within a session.

Future work. Our study, and nearly all of the studies before it, characterized only the wireless network. It would be useful (but nearly impossible, on switched networks) to collect simultaneous information about usage on the wired and wireless networks, to determine what characteristics are unique to the wireless environment.

We would like to study the geographic patterns of mobility. Presumably most users have regular habits as they move from dorm to class to dining hall.

We were unable to distinguish users or types of users (students, faculty, staff). It may be possible to infer the type of users from their behavior (for example, students are seen frequently in dorms), or to use clustering techniques [TB00].

We plan to repeat the study in Spring 2002, with refined data-collection and -analysis scripts.

Acknowledgements

The authors graciously acknowledge the contribution of Pablo Stern, a Dartmouth undergraduate. Pablo constructed many of the scripts for collecting and analyzing the SNMP and sniffer data, and used them in an early analysis of the network as it came on-line in Spring 2001 [Ste01].

We are also indebted to the staff of Dartmouth Computing Services, particularly Steve Campbell, Punch Taylor, Jim Baker, and Charles Clark, for their assistance in installing our sniffers and configuring the syslog collection.

In Computer Science, Wayne Cripps, Arne Grimstrup, Ron Peterson, and Tim Tregubov have been very helpful with many matters of system administration.

Finally, we appreciate the funding, equipment, and technical assistance provided by Cisco Systems, which helped make this study possible.

References

[BB97] Bernard J. Bennington and Charles R. Bartel. Wireless Andrew: Experience building a high speed, campus-wide wireless data network. In *Proceedings of the Third Annual International Conference on Mobile Computing and Networking*, pages 55–65. ACM Press, September 1997.

[Chr] Jim Christy. Cisco Systems engineer. Personal communication, March 4, 2002.

[Hil99] Alex Hills. Wireless Andrew. *IEEE Spectrum*, 36(6):49–53, June 1999.

[HJ96] Alex Hills and David B. Johnson. Seamless access to multiple wireless data networks: A wireless data network infrastructure at Carnegie Mellon University. *IEEE Personal Communications*, 3(1):56–63, February 1996.

[KBZ+00] Thomas Kunz, Thomas Barry, Xinan Zhou, James P. Black, and Hugh M. Mahoney. WAP traffic: Description and comparison to WWW traffic. In *Proceedings of the Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM 2000)*, pages 11–19, Boston, MA, August 2000. ACM Press.

[LRT+98] Kevin Lai, Mema Roussopoulos, Diane Tang, Xinhua Zhao, and Mary Baker. Experiences with a mobile testbed. In *Worldwide Computing and its Applications*, number 1368 in LNCS, pages 222–237. Springer-Verlag, 1998.

[MTK02] G. Ayorkor Mills-Tettey and David Kotz. Mobile voice over IP (MVOIP): An application-level protocol for call hand-off in real time applications. In *Proceedings of the Twenty-first IEEE International Performance, Computing, and Communications Conference*. IEEE Computer Society Press, April 2002. Accepted for publication.

[Per99] Charles E. Perkins. Mobile networking in the Internet. *Mobile Networks and Applications*, 3(4):319–334, 1999.

[Ste01] Pablo Stern. Measuring early usage of Dartmouth's wireless network. Technical Report TR2001-393, Dept. of Computer Science, Dartmouth College, June 2001. Senior Honors Thesis.

[TB99] Diane Tang and Mary Baker. Analysis of a metropolitan-area wireless network. In *Proceedings of the Fifth Annual International Conference on Mobile Computing and Networking*, pages 13–23. ACM Press, August 1999.

[TB00] Diane Tang and Mary 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.

[TB02] Diane Tang and Mary Baker. Analysis of a metropolitan-area wireless network. *Wireless Networks*, 2002. To appear.