Analysis of a Local-Area Wireless Network

Diane Tang
Stanford University
Gates 3A
Stanford, CA 94305-9030
+1 650 725 1776
dtang@cs.stanford.edu

Mary Baker
Stanford University
Gates 4A
Stanford, CA 94305-9040
+1 650 725 3711
mgbaker@cs.stanford.edu

ABSTRACT

To understand better how users take advantage of wireless networks, we examine a twelve-week trace of a building-wide local-area wireless network. We analyze the network for overall user behavior (when and how intensively people use the network and how much they move around), overall network traffic and load characteristics (observed throughput and symmetry of incoming and outgoing traffic), and traffic characteristics from a user point of view (observed mix of applications and number of hosts connected to by users).

Amongst other results, we find that users are divided into distinct location-based sub-communities, each with its own movement, activity, and usage characteristics. Most users exploit the network for web-surfing, session-oriented activities and chat-oriented activities. The high number of chat-oriented activities shows that many users take advantage of the mobile network for synchronous communication with others. In addition to these user-specific results, we find that peak throughput is usually caused by a single user and application. Also, while incoming traffic dominates outgoing traffic overall, the opposite tends to be true during periods of peak throughput, implying that significant asymmetry in network capacity could be undesirable for our users.

While these results are only valid for this local-area wireless network and user community, we believe that similar environments may exhibit similar behavior and trends. We hope that our observations will contribute to a growing understanding of mobile user behavior.

Keywords

Local-area wireless networks, network analysis.

1. INTRODUCTION

More companies and schools are installing wireless networks to support a growing population of mobile laptop and PDA users. Part of the motivation for these installations is to reduce the costs of running cable. Another important motivation is to meet the demands of users who wish to stay connected to the network, communicating with others and accessing on-line information no matter where they are.

In this paper, we analyze a 12-week trace of a local-area wireless network installed throughout the Gates Computer Science Building of Stanford University. Our goal is to answer questions such as how much users take advantage of mobility, how often we observe peak throughput rates, what causes the peaks, and what application mix is used. This study is similar in nature to a previous study [14], however the scale of this network is much smaller, and its characteristics in terms of delay and bandwidth are much more favorable. We thus find that not all questions asked previously make sense in this context; for instance, we do not analyze user mobility for frequently-used paths through the network. In contrast to the previous study, however, we explore information about network data traffic and can ask questions about application mixes, symmetry of outgoing and incoming traffic, and traffic throughput.

Amongst other results, we find that users fall into distinct location-based sub-communities, each with its own behavior regarding movement and periods of activity. We find that almost all users run some version of Windows at least some of the time and exploit the network for web-surfing activities. Besides other house-keeping activities (such as *dns*, *icmp,* and setting the time), many people also use their laptops for session-oriented activities (such as *ssh* and *telnet*) and chat-oriented activities (such as *talk*, *icq*, *irc,* and *zephyr*). The high number of chat-oriented activities shows that some users take advantage of the mobile network for synchronous communication with others. In addition to these user-specific results, we find that peak throughput is caused 80% of the time by a single user and application. Also, while incoming traffic dominates outgoing traffic overall (34 billion bytes compared to 12 billion bytes), the opposite tends to be true during periods of peak throughput, implying that significant asymmetry in network capacity could be undesirable for our users.

We hope that the results we present here will help researchers and developers determine how users take advantage of a local-area wireless network, helping to focus efforts on topics that will achieve the most improvement in user experience. While these results are only necessarily valid for this particular local-area wireless network and user community, we believe that similar environments may exhibit similar behavior and trends.

In this paper, we first present background information about the data we collected and then present the results of our analysis. We divide the analysis into three sections: overall user behavior, overall network traffic characteristics, and user traffic characteristics. We also comment on the network data visualization tools we used, describe related work, and list some possible directions for future work.

2. BACKGROUND

In this section we describe the network analyzed and our tracing methodology. In the Gates Computer Science Building at Stanford University, administrators have made a "public" subnet available for any user affiliated with the university [1]. Users desiring network access via this subnet must authenticate themselves to use their dynamically assigned IP address [5] to access the rest of the departmental and university networks and the Internet.

This subnet, as shown in Figure 1 and described in Table 1, is accessible both from a wireless network and from Ethernet ports in public places in the building, such as conference rooms, lounges, the library, and labs. The wireless network is a WaveLAN network with WavePoint II access points acting as bridges between the wireless and wired networks [15]. The access points each have two slots for wireless network interfaces; both slots are filled, one with older 2 Mbps cards to support the few users who have not updated their hardware yet, and the other with WaveLAN IEEE802.11-compatible 10 Mbps cards.

To help explain the results we present in the next sections, we briefly describe our building and its user community. The building is L-shaped (the longer edge is called the a-wing, and the shorter the b-wing). It has four main floors with offices and labs, a basement with classrooms and labs, and a fifth floor with a lounge

Table 1: Brief summary of the wireless network and community in the Gates Computer Science Building.

Total number of access points	12
Number of floors in building	6
Approximate area covered by an access point	75ft x 150ft
Number of wireless users	74

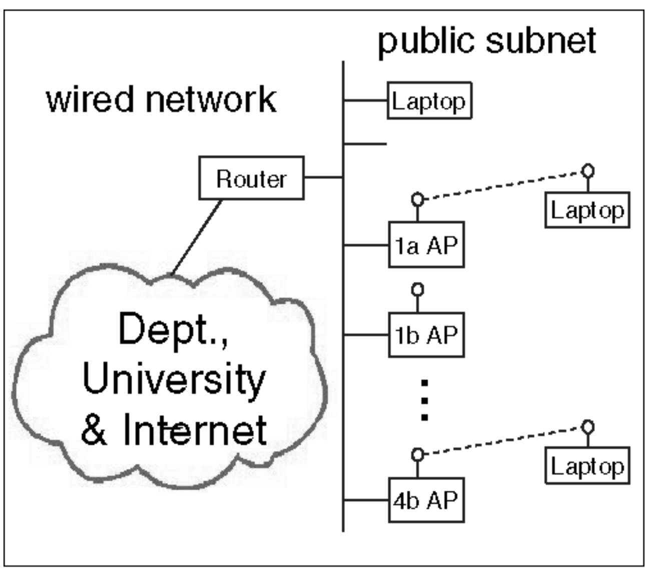

Figure 1: The public subnet and its connectivity to the rest of the departmental and university networks and the Internet. An AP is an access point for wireless connectivity.

and a few offices. Each of the main floors has two access points, one for each wing. Additionally, the first floor has an access point for a large conference room; the library, which spans both the second and third floors, also has an access point. The basement has two access points, one near the classrooms and one for the Interactive Room, a special research project in the department [7]. The smaller fifth floor only has one access point.

The wireless user community consists of 74 users who can be roughly divided into four groups:

- 35 first year PhD students, who were each given a laptop with a WaveLAN card upon arrival (which corresponds to the beginning of the trace). Their offices are primarily in the 2b wing.
- 22 graphics students and staff, the majority of whom received laptops with WaveLAN cards a week into the tracing period. Their offices are primarily in the 3b wing.
- Three robots, used by the robotics lab for research. The robots do not have to authenticate themselves to reach the outside network. While the robots are somewhat mobile, they stay in the 1a wing. Although these WaveLAN cards are intended to be used by the robots, students in the robotics lab also use the network cards for session connections and web-surfing.
- 14 other users (students, staff, and faculty) scattered throughout the building.

In addition to these 74 users, there were also four users who authenticated themselves but only connected to wired ports on the public subnet rather than the wireless network. We do not consider these users in the rest of this analysis of the wireless network.

We obtained permission to collect these traces from the Department Chair and informed all network users that this tracing was taking place. We additionally informed users we would record packet header information only (not the contents) and that we would anonymize the data. Knowledge of the tracing may have perturbed user behavior, but we have no way of quantifying the effect.

Because all of the wireless users are on a single subnet (which promotes roaming without the need for Mobile IP or other such support), we gathered traces on the router shown in Figure 1 that connects the public subnet to the rest of the departmental wired network. The router is a 90 MHz Pentium running RedHat Linux with two 10 Mbps network interfaces. One interface connects to the public subnet, and the other connects to the departmental network.

To gather all of the information we wanted, we collected three separate types of traces during a 12-week period encompassing the 1999 Fall quarter (from Monday, September 20 through Sunday, December 12). The first trace we gathered is a *tcpdump* trace of the link-level and network-level headers of all packets that went through the router [9]. We use this information in conjunction with the other two traces.

The second trace is an SNMP trace [4]. Approximately every two minutes, the router queries, via Ethernet, all twelve access points for the MAC addresses of the hosts currently using that access point as a bridge to the wired network. Once we know which access point a MAC address uses for network access, we know the approximate location (floor and wing) of the device with that MAC address. We pair these MAC addresses with the link-level addresses saved in the packet headers to determine the approximate locations of the hosts in the *tcpdump* trace.

The overhead from the SNMP tracing is low: 530 packets or 50 KBytes is the average overhead from querying all twelve access points every two minutes. The overhead for querying an individual access point is 3.2 KBytes if no MAC addresses are using that access point; otherwise, the base overhead is 14.5 KBytes for one user at an access point, plus 1 KByte for every additional user.

The last trace is the authentication log, which keeps track of which users request authentication to use the network. Each request has both the user's login name as well as the MAC address from which the user makes the request. We pair these MAC addresses with the link-level addresses saved in the *tcpdump* trace to determine which user sends out each packet.

We use the common timestamp and MAC address information to combine these three traces into a single trace with a total of 78,739,933 packets attributable to the 74 wireless users. An additional 37,893,656 packets are attributable to the SNMP queries and 1,551,167 packets are attributable to the four wired users. The number of packets attributable to the SNMP queries might seem high, but each access point is queried every two minutes even if no laptops are actively generating traffic.

Thus, for every packet sent over the course of this twelve-week period we record:

- a timestamp,
- the user's identity,
- the user's location (current access point),
- the application, if the port is recognized, otherwise, the source and destination ports,
- the remote host the user connects to,
- and the size of the packet.

Note that because we do not record any signal strength information, and since our access points generally cover a whole wing of a floor, we cannot necessarily detect movement within a wing but only movement between access points.

3. OVERALL USER BEHAVIOR

In this section we consider the network-related behavior of users, focusing on their activity and mobility. Specifically, we ask the following questions:

1. When and how often do people use the network?
2. How many users are active at a time?
3. How much do users move?

The answers to these questions help researchers understand whether and how users actually take advantage of a mobile environment. Also, by understanding user behavior, network planners can better plan and extend network infrastructure.

In general we find that most users do not move much within the building, but a few users are highly mobile, moving up to seven times within an hour. We also find that users fall into location-based sub-communities, each with its own movement and activity characteristics. For example, the sub-community in the 2b wing tends to move around a fair amount and use the network sporadically, whereas the 4b wing sub-community steadily uses the network but does not move around very much.

3.1 Active Users

We first look at average user activity by time of day. We consider a user to be active during a day in the trace if he sends or receives a packet sometime during that day. We see from Figure 2 that on weekdays more people use the network in the afternoon than at any other time (on average there are 12 to 16 users in the mid-afternoon, with a maximum of 34 users between 2 and 4 in the afternoon). We also see from the steady number of users throughout the night and weekend that four to five users, on

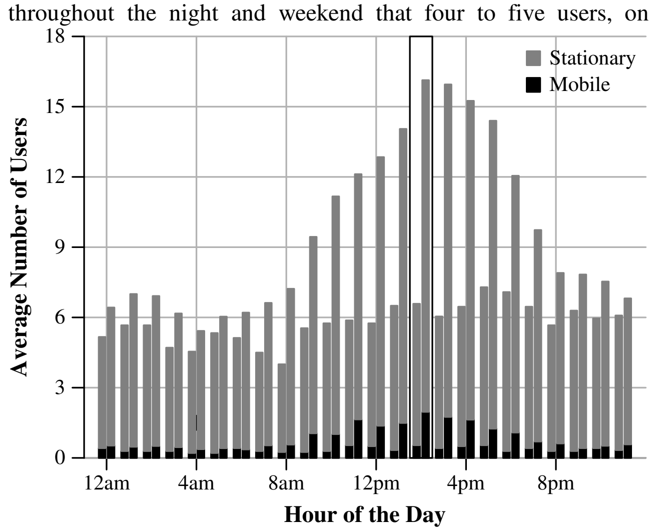

Figure 2: The average number of active users of the mobile network each hour of the day. Each hour has two bars, the left one for weekends and the right one for weekdays. The darkness of the bar indicates whether the users are stationary or mobile (active at two or more access points) during that hour. For example, the highlighted bars show that at 2pm, on average 16.2 users use the network on weekdays (2 of which, on average, visit at least 2 locations over the course of that hour), and on average, 6.6 users use the network on weekends (0.5 of which, on average, visit at least 2 locations over the course of that hour).

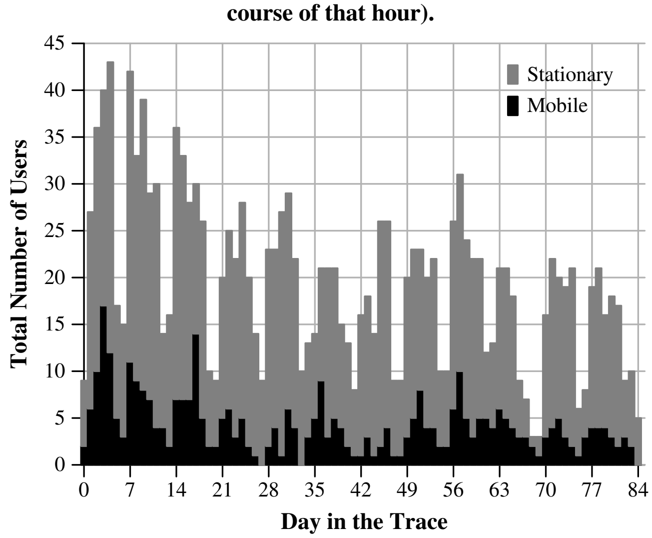

Figure 3: The total number of users of the mobile network per day. The 0th day is Monday, the 1st day is Tuesday, etc. The darkness of the bar indicates whether the users are stationary or mobile during that day.

average, leave their laptops turned on in their offices rather than take them home.

Figure 3 is a graph analogous to Figure 2, presenting the number of active users per day in the trace rather than per hour of the day. We observe a weekly pattern with more active users during the week than on weekends. We also note some trends across the course of the trace: the network supports the most users at the beginning of the trace (up to 43 on the first Friday of the trace), when many users first received their laptops, with a lull in the middle of the quarter corresponding to midterms and comprehensive exams, followed by an upswing corresponding to final project due dates, before a drop during finals week and an exodus for winter vacations. We also believe that the number of users falls off as new Ph.D. students received their permanent office assignments elsewhere in the building. It seems that many users still prefer stationary desktop machines over laptops when both are available to them.

Figure 4 presents overall activity from a user point of view: the total number of days users are active during the traced period.

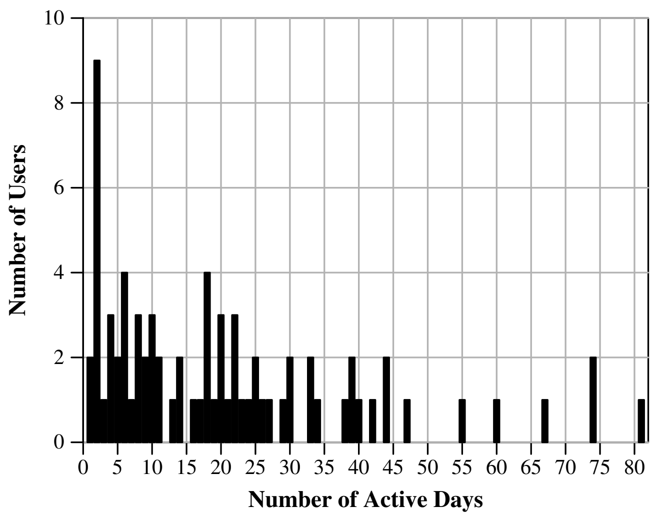

Figure 4: Number of active days for mobile users over the course of the entire trace.

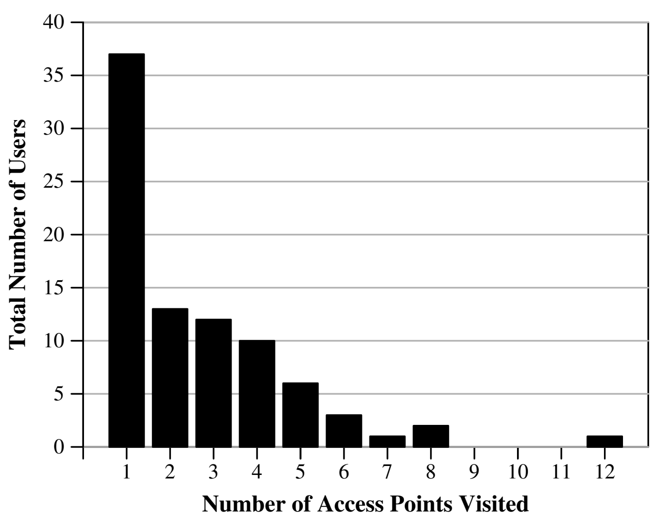

Figure 5: Number of users visiting some number of access points over the course of the entire trace.

While some users rarely connect their laptops to the network (17 users do so on 5 days or fewer), others connect their laptops frequently (14 users are active at least 37 days during the traced period).

3.2 User Mobility

We next explore user mobility. Turning back to Figure 2, we see information about average user mobility by time of day. Most users are stationary, meaning they do not move from one access point to another. Only a few users (1.3 on average) move between access points during any given hour. However, some users are highly mobile with a maximum of seven location changes for a user within an hour. We can now look at Figure 3 to see how many users are mobile on a daily basis rather than an hourly basis.

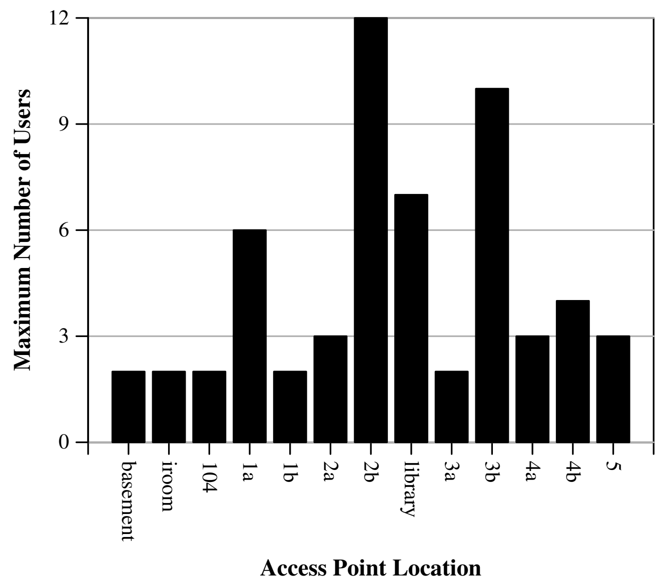

Figure 6: Maximum number of users each access point handles within a five-minute period. iroom = Interactive Room.

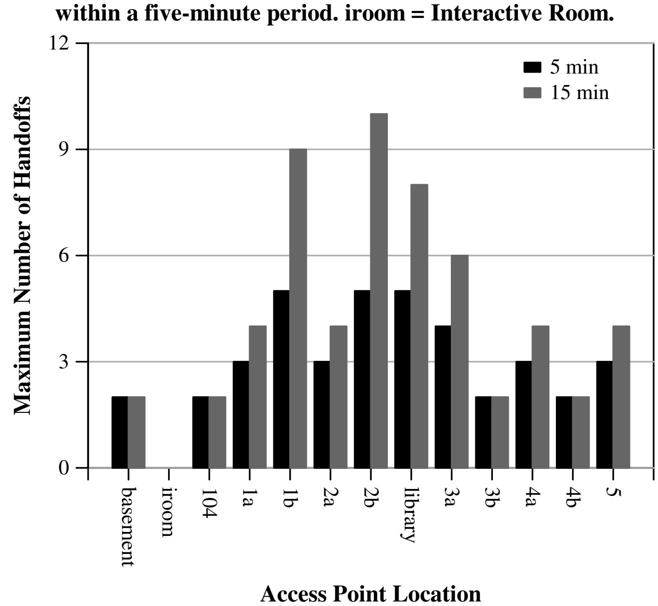

Figure 7: Maximum number of handoffs each access point handles within a 5-minute or 15-minute period. iroom = the Interactive Room in the basement.

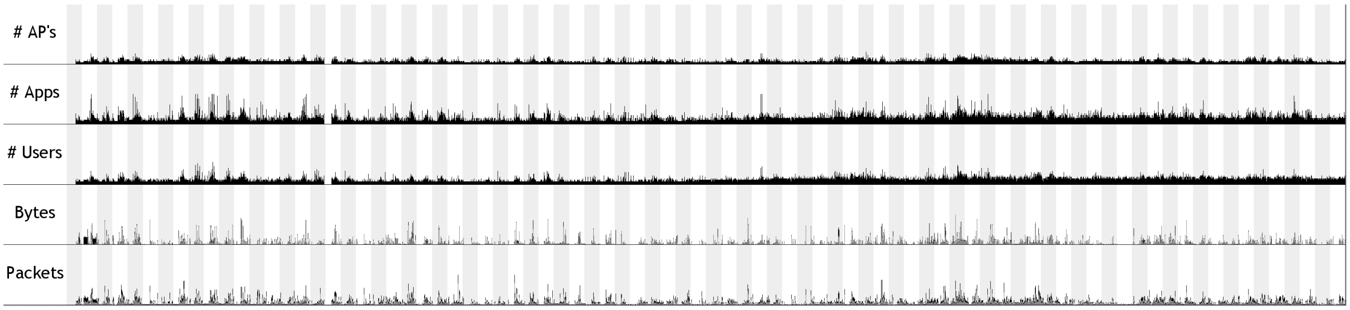

Figure 8: Overview of the throughput trends over the entire trace, both in bytes (maximum of 5.6 Mbps) and packets (maximum of 1,376 packets per second), as well as the number of access points (AP's, maximum of 9 simultaneous AP's), applications (maximum of 56 simultaneous applications), and users (maximum of 17 simultaneous users) responsible for generating the traffic.

The number of mobile users is high towards the beginning of the trace, with up to 13 mobile users during a day, and decreases towards the end of the trace, to only one to two mobile users during a day. As in Figure 2, however, we see that most users are stationary on any given day with only a few (3.2 on average) moving around.

Looking at total mobility across the trace in Figure 5, we see that while 37 users are stationary throughout the entire period, a few users exploit the mobile characteristic of the network: 13 users visit at least five distinct access points during the course of the trace and one user visits all twelve access points.

3.3 User Sub-communities

We now turn to location-based user behavior by associating user activity and mobility with access points. Figure 6 shows that the access points in the 2b and 3b wings handle the most users (up to 12 or 10 users, respectively, within a five-minute period), which is not surprising given the large number of mobile users with offices in those two wings. Figure 7 shows how many handoffs access points have to handle. Contrasting Figure 6 with Figure 7, we see that the number of users is not necessarily correlated with movement. The users on the 3b wing rarely move, while the users on the 2b wing move around more often. The few users in the 1b wing move even more frequently.

Table 2: Brief description of the activity at each access point throughout the course of the trace.

Access Point	Description
basement	occasional spikes corresponding to meetings
iroom	big peak in weeks 8, 9 (project deadline)
104	occasional spikes corresponding to meetings
1a	heavy usage weeks 1-3, occasional afterwards
1b	occasional usage corresponding to network testing
2a	occasional usage, small peak towards end
2b	closely follows overall pattern in Figure 2
library	meetings in weeks 1-3, slight peak weeks 6-7
3a	lower usage, follows overall pattern in Figure 2
3b	follows overall pattern in Figure 2
4a	1-2 users regularly
4b	1-3 users constantly
5	1-2 users in late afternoons, Monday-Friday

Table 2 summarizes user activity by access point location. The basement and the conference room in 104 are primarily used only when meetings occur, while the 5th floor lounge is used when people take a break in the late afternoon. The 4th floor users are steady users who rarely move, the 3rd floor users connect to the network more sporadically, and the 2nd floor users are also sporadic but more mobile. These results reveal that while each access point covers approximately the same amount of space, the load on each access point depends on the behavior of the community it serves.

3.4 Access Point Handoffs

One side effect of user mobility is the need for access points to perform handoffs. We thus take a closer look at how many handoffs access points handle. A handoff is defined as a user appearing at one access point and then moving to a different access point within a given period of time. Looking at Figure 7, we see that handoffs are not a major burden on access points: an access point handles at most five handoffs within a five-minute period, or ten within a 15-minute period. Note that 95% of all user location changes occur within 15 minutes.

4. OVERALL NETWORK TRAFFIC CHARACTERISTICS

In this section we consider overall characteristics of the network, such as throughput, peak throughput, and incoming and outgoing traffic symmetry. Specifically, we ask the following questions:

1. What is the throughput through the router? Through the access points?
2. What is the peak throughput?
3. How often is peak throughput reached?
4. What causes the peaks? Several users or only one or two? Multiple applications or only a few?
5. How symmetric is the traffic? (How similar is incoming traffic to outgoing traffic?)
6. How much traffic is attributable to small versus large packets?

The answers to these questions help determine how wireless hardware and software should be optimized to handle the amounts of traffic wireless networks generate. Such optimizations may include using asymmetric links or optimizing for a few large packets versus many smaller packets.

While we believe that latency is critical to users, the latency of the WaveLAN network is equivalent to wired ethernet, and we thus choose not to analyze our trace for this metric. The latency users see on our network is attributable almost entirely to the outside network, especially the Internet [12].

In general, we find that router throughput reaches peaks of 5.6Mbps and that peak throughput is caused 80% of the time by a single user and application, usually a large file transfer. On average, the incoming traffic is heavier than outgoing traffic, but the periods of peak throughput are actually skewed more towards outgoing bytes. From this result, we conclude that significant asymmetry in network capacity would not be desirable for our users. We also find that in our network's application mix, low per-packet processing overhead to handle many small packets is just as important as high overall attainable byte throughput.

4.1 Network Throughput

Figure 8 gives an overview of throughput over the traced period, as well as how many access points, users, and applications are responsible for generating the traffic. Throughput through the router is typically around one to three Mbps. Usually, the throughput as a whole increases as the number of users increases. The throughput through the router reaches peaks of 5.6 Mbps. Table 3 shows the maximum throughput attained through the router and each access point. In no case is the peak throughput maintained for more than three seconds, indicating that the network is not overwhelmed, but rather that traffic is heavy enough to hit the peak rate on occasion.

For the majority of peaks, the maximum throughput is achieved by a single user and application, rather than distributed across several users, as we might expect since the access points with the largest peaks are also the access points with the most users. Specifically, of the 1,492 peaks of magnitude 3.6 Mbps or greater, 80% of those peaks have 94% of their traffic generated by a single user and application, and 97% of the their traffic generated by a single user. The application responsible for 53% of those peaks is ftp, with web traffic responsible for 15%, and the remainder caused by applications such as X, session traffic (e.g., *ssh* and *telnet*), and mail downloads (e.g., *eudora*, *imap*, and *pop*).

Table 3: Maximum throughput attained through the router, the public Ethernet ports, and each access point.

Location	Max Packets	Max Bits	% of Peaks > 3 Mbps
router	1,376 pps	5.6 Mbps	100.00%
ethernet	1,096 pps	5.1 Mbps	5.80%
basement	530 pps	3.2 Mbps	0.10%
iroom	446 pps	3.6 Mbps	3.30%
1a	521 pps	3.4 Mbps	0.60%
1b	455 pps	3.6 Mbps	2.00%
104	429 pps	3.4 Mbps	0.70%
2a	783 pps	3.1 Mbps	0.01%
2b	824 pps	4.5 Mbps	8.90%
library	745 pps	4.5 Mbps	2.40%
3a	737 pps	3.6 Mbps	6.30%
3b	883 pps	4.6 Mbps	69.70%
4a	804 pps	1.7 Mbps	0.00%
4b	675 pps	3.9 Mbps	0.07%
5	703 pps	3.4 Mbps	0.10%

From this data, we also observe that evidence of user subcommunities with different behaviors carries over to traffic throughput characteristics. While the wings with the most users (2b and 3b) also have the highest peak throughput, the users on the 3b wing attain that throughput more often (69% of peaks of magnitude greater than 3 Mbps are attributable to the 3b wing), indicating that although these users may not be very mobile, their traffic causes more load on the network.

4.2 Network Symmetry

Another network characteristic we investigate is the symmetry of incoming and outgoing traffic. We might expect that because the most common application is web-surfing (see Section 5) that incoming packets and bytes would overwhelm outgoing packets and bytes. Instead, we find that while the total incoming traffic (34 billion bytes and 62 million packets) is larger than the total outgoing traffic (12 billion bytes and 56 million packets), the peaks are actually skewed more towards outgoing traffic. Of the peaks of magnitude greater than 3.6 Mbps, 60% are dominated by outgoing rather than incoming traffic. From this data, we conclude that significantly asymmetric capacity in wireless networks would be undesirable to users in environments similar to ours.

4.3 Packet versus Byte Throughput

The last overall network characteristic we explore is how packet throughput differs from byte throughput. Figure 9 presents a closer look into the distribution of packet sizes in the network, showing that over 70% of packets are smaller than 200 bytes. However, this same number of packets represents only about 30% of all bytes transmitted. We thus conclude that low per-packet processing overhead is just as important to users in this environment as high overall attainable throughput. Note that fragmented packets are not reassembled for this graph. However, of the 78,738,933 total packets, only 206,895 (0.26%) are fragments and should therefore not impact the distribution much.

We further look at the packet size distribution across several commonly used applications, shown in Figure 10, to determine how these applications can be categorized in terms of packet size. We see that http and database applications should be optimized to handle large incoming and small outgoing packets. In contrast, session, chat, mail, and X applications should be optimized to handle many small outgoing and incoming packets. While the optimizations for mail, an application for asynchronous personal communication, may be independent of latency, the optimizations for session, chat, and X applications must not only optimize for the many small packets but also minimize delay to facilitate user interactivity.

5. USER TRAFFIC CHARACTERISTICS

In this section we consider traffic characteristics from a user perspective. The specific questions we ask in this section are:

1. Which applications are most common?
2. How much does application mix vary by user?
3. How many hosts do users connect to?
4. How long are users active?

Answering these questions helps determine which applications and application domains to optimize for mobile usage. Knowing the traffic mix and how it varies by time can also help researchers model user traffic better, which is important when simulation is used to evaluate mobile protocols. Finally, knowing which and how many remote hosts users connect to also helps when modeling network connectivity.

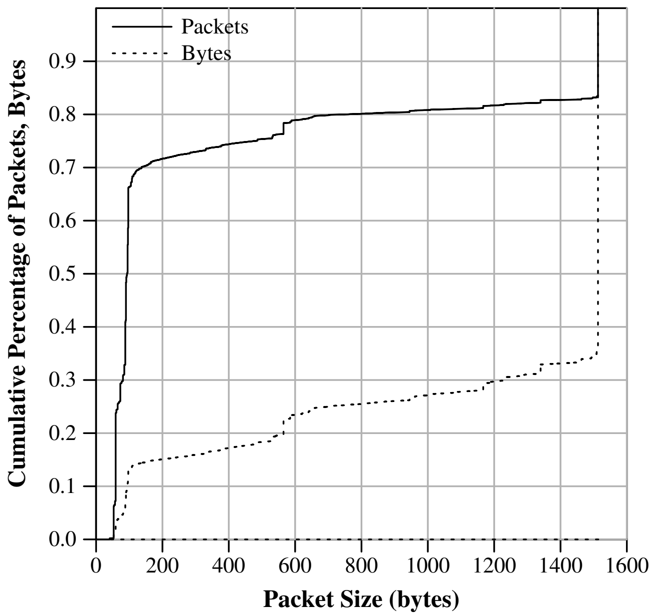

Figure 9: Cumulative histogram showing the percentage of packets that are a certain number of bytes long and the percentage of bytes transferred by packets of that length.

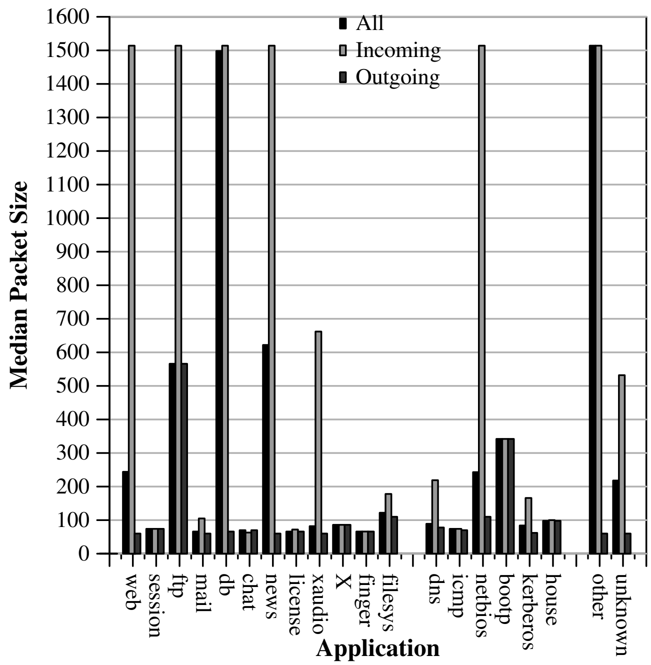

Figure 10: Median incoming and outgoing packet sizes for some commonly used applications. Session applications include ssh and telnet; mail includes pop, imap, and Eudora; filesys includes nfs and afs; chat includes talk, icq, zephyr, and irc; house includes housekeeping applications such as ntp.

We find that the most popular applications are web-browsing and session applications such as ssh and telnet. These two classes of applications are frequently run together, so some optimization of their interaction might be useful. About half our users frequently execute chat-oriented applications (such as talk, icq, irc, and zephyr), showing that some users exploit the mobile network for synchronous communication with others. We also find that user application mixes can be classified into several patterns, such as the terminal pattern, wherein people primarily use their laptops to keep sessions to external machines open, or

Table 4: The most common applications by user, incoming packets and bytes, and outgoing packets and bytes (all in millions). The first group of applications contains basic services, the second group contains client applications, and the last group contains other (the aggregation of all other recognized applications) and unknown (the aggregation of the unrecognized packets). Session applications include ssh and telnet; mail includes pop, imap, and Eudora; filesys includes nfs and afs; chat includes talk, icq, zephyr, and icq; house includes housekeeping applications such as ntp.

App Class	Num Users	Num Inc. Pkts	Num. Inc. Bytes	Num Out. Pkts	Num Out. Bytes
dns	74	0.2	45	0.21	17
icmp	74	0.74	69	0.73	64
netbios	71	7.3	6200	7.2	1200
bootp	56	0.007	2.3	0.007	2.2
kerberos	34	0.004	1.8	0.002	6.8
house	73	0.046	10.8	0.054	6.8
web	73	14.4	15700	11	1100
session	63	6.7	1400	6.8	789
ftp	62	2.8	2600	4	2400
mail	47	0.7	418	0.5	55
db	44	0.005	6.4	0.002	0.2
chat	38	0.03	14.7	0.03	2.2
news	21	0.24	266	0.15	9.6
license	21	0.001	0.5	0.002	0.2
xaudio	21	0.002	1.2	0.001	0.2
X	20	1.5	854	2.2	218
finger	18	0.001	0.2	0.001	0.08
filesys	16	0.07	37	0.06	36
other	33	0.1	141	0.1	51
unknown	68	6	3500	3.7	1000

the web-surfing pattern, wherein most network traffic is web traffic to many different hosts.

5.1 Application Popularity and Mixes

Table 4 lists the most common classes of applications by number of users and total number of packets and bytes. Unsurprisingly, basic service applications such as dns, icmp, and house-keeping applications such as ntp are used by everyone. The high amount of netbios traffic indicates that almost all the users (71) run some version of Windows on their laptops. This reflects our system administrators' choice to install Windows as the default system on laptops.

Of the end-user applications, http, session applications, and file transfer applications are the most popular (with 73, 63 and 62 users, respectively). There are several interesting points of comparison in this data. First, 62 people use some file transfer

Table 5: Description of the eleven application mixes and number of users per application mix. The only applications considered are web, session, X, mail, ftp, and chat. The starred entries are shown in more detail in Figure 11.

Name	Num Users	Description
home*	13	upload in the morning, download for lunch or before going home. Mostly web and ftp, some session and mail traffic. Usually weekday only, occasional weekend traffic.
web-surfer*	12	big web-surfer visiting lots of sites, session to one or two sites, plus a bit of the other applications.
rare	10	one or two single peaks, always web, sometimes session.
dabbler*	9	fairly evenly distributed among the applications, three users active on weekdays only, six users active on both weekdays and weekends.
talkies	8	fairly normal hours, weekday and weekend, mostly web and session, but significant chat traffic too.
terminal	7	weekday only, mostly session traffic, some web, occasional ftp.
mail-client	6	three users active weekdays only, three users active weekdays and weekends, leave their laptops overnight as mail-clients, plus some web surfing and other applications during the day.
late-night	5	lots of chat, web, and ftp late at night. More “normal” traffic (session, web) during the day.
X-term	3	lots of X, session, and web traffic. A little traffic from the other applications.
day-user	3	web and mail during the day, a little session, ftp, and X traffic.
ftp	2	lots of ftp with some session and web traffic, a little bit of the other applications.

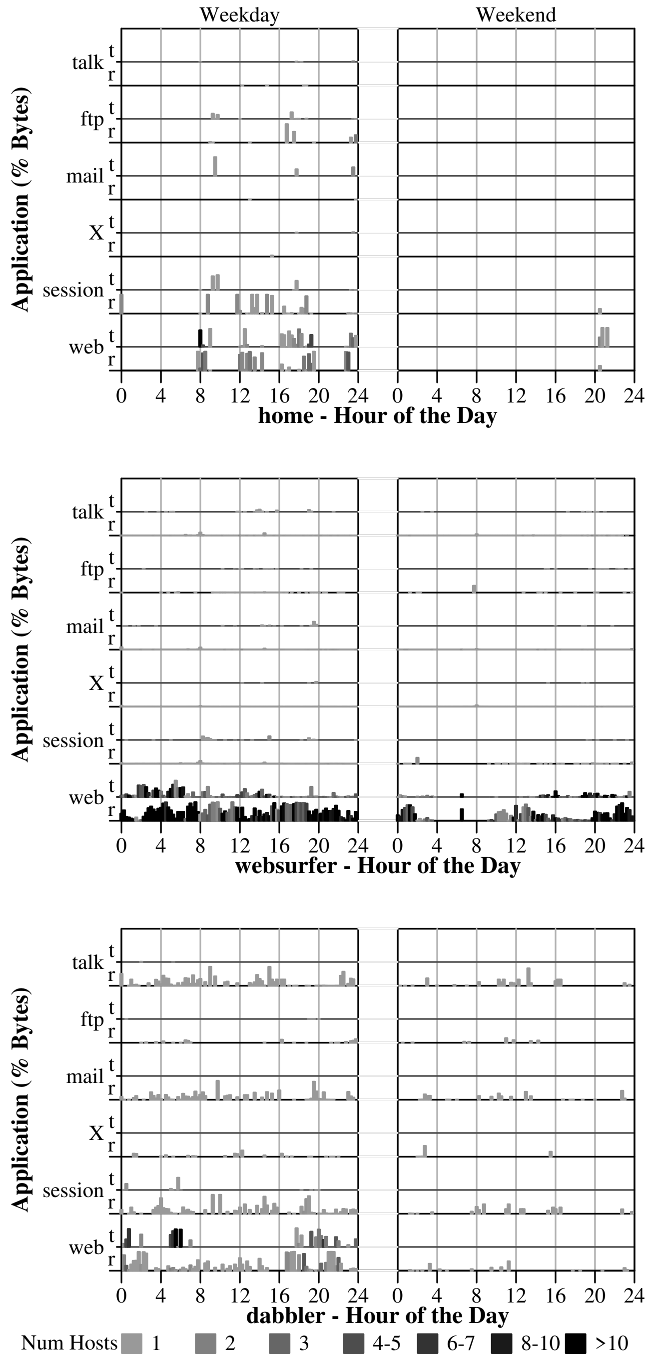

Figure 11: Three application mixes (home, web-surfer, and dabbler). Each graph shows the percentage of bytes sent at that time of day per application, for both weekdays and weekends. Each application is split into two graphs, one for traffic to “repeat” hosts (r), and one for traffic to “throwaway” hosts (t). A repeat host is one the user connects to on at least two different days. The darkness of the bar indicates how many different hosts the user connects to. The darker the bar, the more hosts.

protocol compared to 16 people who use some remote filesystem such as NFS [13] or AFS [8]. This disparity indicates that many users find it necessary to transfer files to and from their laptops, but only a few users are either willing or find it necessary to use a distributed filesystem, perhaps due to the lack of support in distributed filesystem servers for dynamically assigned addresses.

Also interesting is the number of people who use their laptops as a mere terminal compared to the number of people who run applications directly on their laptops. Specifically, only 20 users run X. In comparison, 47 users run some direct mail client (pop, imap, eudora, smtp, etc.); 21 connect to some license server (Matlab, Mentor Graphics, etc.), presumably to run the application directly on their laptops; 21 connect directly to a news server; and 38 users run some sort of chat software (talk, icq, zephyr, irc, etc.). These numbers reveal a tendency to use laptops as stand-alone machines with connectivity, rather than mere terminals. However, most users do still use session applications such as ssh or telnet, showing that users still need to connect to some other machines. Finally, over half of the users execute chat software, indicating that some users treat their laptops in part as personal synchronous communication devices.

In addition to overall application usage, we also look at eleven characteristic user application mixes, shown in Table 5. The main characteristics we consider in this categorization are the percentage of traffic in a given period of time that can be attributed to each application, at what time each application dominates the user's traffic, and the number of hosts to which a user connects. We only use a coarse-grained time characterization: weekday versus weekend and during the day versus late at night. We also confine the categorization to six common applications: web-surfing, session applications, X, mail applications, file transfer, and chat applications.

Figure 11 provides more detail for three of the application mixes. The first mix is the home mix, wherein the user is active in the morning uploading information or work from the laptop, at lunch downloading information or work, and in the evening downloading materials before heading home. These users typically connect to only one or two sites which are frequently repeated. The next mix is the web-surfer mix, wherein users contact many different web sites (up to 3,029 distinct web sites for one user). Many of these sites (up to 1,982 for one user) are visited more than once by the same user. The last application mix we focus on is the dabbler mix, in which users run all of the application types at least once.

We derive several conclusions from this application mix characterization. First, while at some point every possible combination of applications is run together, the applications most commonly run together are web and session applications. Second, while http and ssh are the most popular applications across all users, different users do run different mixes of applications, and they do so at different times of the day. There is no single application mix that fits all mobile users. Finally, not only do the mixes vary by application and time, but also by the number of hosts to which users connect. Some users connect to as few as six hosts, while others connect to as many as 3,054 distinct hosts. (The router connects to a total of 15,878 distinct hosts over the course of the entire trace; 13,178, or 83%, of those hosts are accessed via the web.)

5.2 Web Proxies

Given these access patterns, we can ask whether a web proxy for caching web pages might be an effective technique in our environment. For a rough evaluation, we looked for web sites visited multiple times, either on different days or by different users. Of the 13,178 hosts connected to via the web, 3,894 (30%) are visited multiple times by more than one user, 5,318 (40%) are visited on more than one day during the trace, and 5,359 (41%) are visited either by more than one user or on more than one day. These results indicate that web proxies would be at least a partially effective technique in an environment such as ours.

5.3 Network Sessions and Lease Times

The final question we ask is how long people use the network at a sitting. Since the wireless network is part of the "public" subnet, users must authenticate themselves when they want to access any host outside the subnet. The current policy is to require users to authenticate themselves every 12 hours [1]. Twelve hours was believed to be a good balance between security concerns and user convenience. Of the 1,243 leases handed out over the course of the trace, 23% (272 leases) are renewed within one second of the previous lease's expiration, 27% (310 leases) are renewed within 15 minutes of expiration, 30% (339 leases) are renewed within one hour of expiration, and 33% (379 leases) are renewed within three hours of the previous lease's expiration. Of the 69 users who authenticate themselves, 48 users authenticate themselves again within an hour of a previous lease expiration at least once during the traced period. Given the high percentage of users who re-authenticate themselves very quickly, we conclude that 12 hours is not very convenient for our users and that 24 hours might be a better balance between security and ease of use.

6. VISUALIZATION TOOLS

During the course of this analysis, we use Rivet [3] to create interactive visualizations quickly for exploring the data. A screen shot from one such visualization is shown in Figure 8. Visualizing this large amount of data (78,739,933 packets) is especially useful for gaining an overall understanding of the data and for exploring the dataset, leveraging the human perceptual system to spot unexpected trends. While traditional analysis tools (such as perl scripts, gnuplot, and Excel) are useful, they require the user to formulate questions *a priori*. By using an interactive visualization to explore the data, we are able to spot unexpected trends, such as the division of users into sub-communities and the lease times being too short.

7. RELATED WORK

Other studies of local-area networks exist, but they tend to have a less user-oriented focus. For example, researchers at CMU examined their large WaveLAN installation [6]. This study focuses on characterizing how the WaveLAN radio itself behaves, in terms of the error model and signal characteristics given various physical obstacles, rather than on analyzing user behavior in the network. Other researchers also studied the campus-wide WaveLAN installation at CMU [2]. However, this study focuses on installing and managing a wireless network rather than on user behavior.

Another related effort is joint work from Berkeley and CMU [11]. The researchers outline a method for mobile system measurement and evaluation, based on trace modulation rather than network simulation. This work differs from our own in several ways. First, the parameters they concentrate on deal with latency, bandwidth, and signal strength rather than with when users are active and which applications they run. Second, their emphasis is on using these traces to analyze new mobile systems, rather than on understanding the current system. In this paper, our goal is to understand how people use an existing mobile system.

We previously studied a metropolitan-area network [14], but focused more on user movement than on user traffic in that analysis. Also, that network had very different characteristics, including number of users, geographical size, network delay and bandwidth, than the network analyzed in this paper.

Also at Stanford University, our research group performed an earlier study of a combined wireless and wired network [10]. However, this study was limited in that only eight users participated and the trace only lasted eight days.

8. FUTURE WORK

The greatest weakness in our work is its possible specificity: our results only necessarily apply to our network and user community. While we believe many of our observations would hold true in other similar environments, we have not verified this. We would thus like to study other local-area networks, including a much larger building-wide or even campus-wide WaveLAN network to explore whether our conclusions are affected by scale. With a larger network, it might also make sense to look for geographical patterns of user mobility as people go to classes, offices, lunch, and so forth. We also wonder whether our results are specific to an academic environment and would like to perform a similar study in a corporate or commercial setting. Only through the collection of several different studies can we detect important trends that hold for many wireless environments.

9. CONCLUSION

Although these results are specific to this WaveLAN wireless network and this university user community, we hope our analysis is a start on understanding how people exploit a mobile network. We find that the community we analyze can be broken down into subcommunities, each with its own unique behavior regarding how much users move, when users are active (daily, weekly, and over the course of the trace), and how much traffic the users generate. We also find that although web-surfing and session applications such as ssh and telnet are the most popular applications overall, different users do use different sets of applications at different times and connect to different numbers of hosts. In addition to this user behavior, we also find that asymmetric links would likely be unacceptable in this type of wireless network, and that optimizing packet processing is just as important as optimizing overall throughput.

The trace data we have collected is publicly available on our web site:

`http://mosquitonet.stanford.edu/`

10. ACKNOWLEDGMENTS

This research has been supported by a gift from NTT Mobile Communications Network, Inc. (NTT DoCoMo). Additionally, Diane Tang is supported by a National Physical Science Consortium Fellowship.

11. REFERENCES

[1] Appenzeller, G., Roussopoulos, M., and Baker, M. User-Friendly Access Control for Public Network Ports. Proceedings of IEEE Infocom 1999, March, 1999, 699-707.

[2] Bennington, B.J. and Bartel, C.R. Wireless Andrew: Experience Building a High Speed, Campus-Wide Wireless Data Network. Proceedings of the Third Annual ACM/IEEE International Conference on Mobile Computing and Networking. August, 1997. p. 55-65.

[3] Bosch, R., Stolte, C., Tang, D., Gerth, J., Rosenblum, M., and Hanrahan, P. Rivet: A Flexible Environment for Computer Systems Visualization. To appear in Computer Graphics 34(1), February 2000.

[4] Case, J.D., Fedor, M., Schoffstall, M.L., and Davin, C. Simple Network Management Protocol (SNMP). RFC 1098. April, 1989.

[5] Droms, R. Dynamic Host Configuration Protocol (DHCP). RFC 2131. March, 1997.

[6] Eckardt, D. and Steenkiste, P. Measurement and Analysis of the Error Characteristics of an In-Building Wireless Network. Computer Communication Review 26, 4 (October 1996), 243-254.

[7] Fox, A., Johanson, B., Hanrahan, P., and Winograd, T. Integrating Information Appliances into an Interactive Workspace. IEEE Computer Graphics and Applications, Vol. 20, No. 3, May, 2000, 54-65.

[8] Howard, J.H., Kazar, M.L., Menees, S.G., Nichols, D.A., Satyanarayanan, M., Sidebotham, R.N., and West, M.J. Scale and Performance in a Distributed File System. ACM Transactions on Computer Systems, Vol. 6, No. 1, February 1988, p. 51-81.

[9] Jacobson, V., Leres, C., and McCanne, S. tcpdump. Available via anonymous ftp to ftp.ee.lbl.gov, June 1989.

[10] Lai, K., Roussopoulos, M., Tang, D., Zhao, X., and Baker, M. Experiences with a Mobile Testbed. Worldwide Computing and Its Applications, Lectures notes in Computer Science (1368). Berlin: Springer, 1998, 222-237.

[11] Noble, B., Satyanarayanan, M., Nguyen, G., and Katz, R. Trace-Based Mobile Network Emulation. Computer Communication Review 27, 4 (October 1997), 51-61.

[12] Paxson, V. End-to-End Internet Packet Dynamics. Computer Communications Review 27, 4 (October 1997), 139-152.

[13] Sandberg, R., Goldberg, D., Kleiman, S., Walsh, D., and Lyon, B. Design and Implementation of the Sun Network Filesystem. Proceedings of the Summer 1985 USENIX Conference, June, 1985, p. 119-130.

[14] Tang, D. and Baker, M. Analysis of a Metropolitan-Area Wireless Network. Proceedings of Fifth Annual ACM/IEEE International Conference on Mobile Computing and Networking. August, 1999. p. 13-23. Revised version to appear in Wireless Networks.

[15] WaveLAN. http://www.wavelan.com.