Mobile, Wireless and Sensor Networks, Edited by Rajeev Shorey.
ISBN 0-471-XXXXX-X

Chapter 2

Understanding the Use of a Campus Wireless Network

David Schwab and Rick Bunt

Introduction

The University of Saskatchewan (U of S) campus covers a large physical area, with more than 40 buildings distributed over 147 hectares of land on the banks of the South Saskatchewan River. Our geography has a significant impact on our approach to delivery of IT. The campus wireless network is one of several new projects we have introduced over the past three years to enhance the computing environment for our 18,000 students. Our approach is to provide mobile users with access to our wireline network through high-speed wireless access points located in very public areas.

Our initial deployment began in the 2001/02 academic year with a pilot project, consisting of a small number of access points (18) placed strategically in a number of locations. This trial deployment proved that wireless technology would be an effective and secure way to give students greater access to network resources and the internet. The demand for wireless networking has grown steadily since then. After the trial rollout was completed, the wireless access points were fully integrated into the campus network, and are now regularly used by a growing number of wireless users. We continue to expand

the network to meet that demand, and now have close to 80 access points. Further wireless installations are being planned, both for new buildings and as part of ongoing expansion.

In order for us to plan for any expansion, it is important that we understand current usage patterns—that we understand where, when, how much, and for what our wireless network is being used. It is also important to understand how usage patterns are changing and what future usage can be projected from current trends. This paper describes the methodology we are employing to collect data on usage, and what we are learning. Authentication logs were collected in co-operation with our Information Technology Services Division (ITS) over the 2003/04 academic year. In our analysis, this usage data is supplemented with short-term wireless packet traces gathered at specific campus locations.

The paper is organized as follows. Section II reviews related work in wireless network measurement. Section III describes the wireless network at the U of S, including its initial deployment, results from early user measurement research conducted on it and the production network configuration in use during the current study. In Section IV we describe the methodology followed when gathering and analysing the data, and compare it to the methodology employed during our earlier research. Section V contains the results of our analysis, and offers comparisons between current and past results. We conclude in Section VI with a summary of our key findings.

Related Work

The design of this study was motivated by work done by Balachandran *et al.* [1]. Their analysis and characterization of the traffic generated by attendees at a popular ACM conference in the summer of 2001 provided many useful insights. They employed two mechanisms to gather wireless traffic traces during the conference. One trace was gathered by periodically polling each of four access points positioned in the conference hall with SNMP requests. This trace revealed usage statistics at the access-point level, including the number of users currently connected and the number of transmission errors. The second trace was gathered at a router that connected the access points to the campus network. This tracing was done using *tcpdump* [2] to gather anonymised TCP packet headers. The analysis of those headers revealed access-point independent statistics, such as the total amount of traffic on the wireless network and the application mix of that traffic.

Although the conference trace was gathered successfully and analysed thoroughly, the findings from its analysis have limited applicability to a full campus setting. The conference had a set schedule, which caused readily apparent traffic patterns as all attendees moved from event to event. Furthermore, the access points were all placed in the same conference hall area, which resulted in almost identical usage patterns being observed at each access point.

The analysis of the Dartmouth College wireless network by Kotz and Essien [3] is more relevant to campus-wide networks. Dartmouth's wireless network is made up of 476 access points providing coverage in 161 buildings for almost 2000 users. The Dartmouth study used a combination of three forms of trace-gathering: event-triggered

log messages, SNMP polling and packet header recording. Because of the de-centralised structure of the Dartmouth network, however, packet headers could be gathered from only a small number of locations, and because the SNMP and log messages were sent by each access point individually via UDP packets, some of the data was lost or mis-ordered. Also, some of the access points experienced power failures or mis-configuration problems which resulted in gaps in the trace.

Both these studies were based on previous research done at the Stanford University Computer Science Department. Tang and Baker [4] used *tcpdump* and SNMP polling to gather statistics on 74 wireless users over a 12 week period. While their study did establish the methodology used by subsequent wireless network traces, the scope of their work was limited to a single department in a single building and does not fully reflect the activities of the broad spectrum of campus wireless users.

More recently, Papadoupouli *et al.* [5] studied wireless usage on the University of North Carolina campus at Chapel Hill. Their investigation focused on user mobility patterns, specifically the predictability of roaming behaviour and the correlation between association patterns and web access. Papadoupouli *et al.* believe that wireless users would benefit from localised, peer-to-peer and predictive caching systems, especially with regards to location-specific information and services. Although the web is not a location-based service, their study suggests that a significant percentage of all web requests – a larger percentage of requests from highly mobile users – could be considered location-dependent.

Network Environment

Our campus wireless network currently consists of close to 80 access points which provide service to over 700 clients, including students, faculty and staff. New buildings, such as our new Kinesiology building, are being constructed with wireless access points from day one. Older buildings are rapidly being added to the wireless network as new access points go online every month.

From the beginning, we have used a mix of Cisco AP350 and AP1200 access points. Both models support 802.11b (or Wi-Fi) connections, and the AP1200 is upgradeable to support 802.11g and/or 802.11a connections. Using the proprietary Cisco LEAP (Lightweight Extensible Authentication Protocol) authentication system, each connection is authenticated by verifying the username and password specified with a Cisco Secure Access Control Server (ACS). This allows users to connect to the wireless network using the same username and password as they use to login to laboratory machines and internet services. The ACS records every authentication and deauthentication that occurs on the wireless network. [6]

Clients can connect to the wireless network using any wireless network adaptor with drivers that support LEAP — such as the Cisco Aironet 350 or Apple Airport. To encourage early adoption of wireless technology, Aironet 350 wireless adaptors were made available at a subsidized price to students, faculty and staff through our Campus Computer Store during the first year.

Our initial wireless network was deployed as a pilot project during the summer of 2001 on a virtual subnet of our extensive wireline network. The use of a subnet enabled

us to distinguish wireless traffic from non-wireless traffic and helped ensure that unauthorized wireless users would not have access to campus services.

In early 2003, we conducted an initial study of the wireless network, and the results of this study are reported in [7]. Traffic on the entire wireless pilot project subnet was mirrored at the central campus router from January 22, 2003 to January 29, 2003. The mirrored traffic was recorded using the network analysis package *EtherPeek* [8]. Anonymised ACS log data from the period was also made available for this preliminary study. Although the week-long trace could not be seen as representative of average wireless user behaviour, our work established a useful methodology and our analysis provided a statistical snapshot of the status of the early campus wireless network. These preliminary observations are compared to our current results in Section V. During the summer of 2003 the campus shifted from a switched network to a routed network, and wireless access points were integrated into the common campus subnet.

We are currently pursuing several new applications for 802.11 wireless networking technology on our campus. In the spring of 2004, ITS began installing long-range point-to-point wireless links between the campus and remote research facilities located some distance outside the city. Trials are underway to allow low-end devices (such as handheld computers using the Palm OS and wireless inventory tracking devices) which do not support LEAP authentication to connect to the campus wireless network. These devices will be authenticated by comparing their machine (MAC) address to a list of allowed devices. We are also deploying Voice over IP (VoIP) phone service in some newly constructed buildings, and this may also soon be usable over the campus wireless network as an alternative to cellular service.

Methodology

Authentication Logging

The Cisco Secure ACS keeps track of every wireless user currently connected to the network and this information is logged for security monitoring purposes. The ACS log includes a record of each authentication and deauthentication that occurs on the wireless network. The information recorded includes the date and time, username, client card address, session id and access point address associated with each event. In addition, each deauthentication record includes the number of packets transmitted and received and the amount of data those packets contained. ITS has been saving anonymised copies of these log files for use in our research since late August of 2003.

Trace Collection

While the ACS logs reveal overall usage patterns over the entire campus, they do not give the specific information needed to characterize the applications and traffic patterns associated with wireless users. To gain this more detailed information, a trace of wireless traffic is needed.

In our earlier study of the wireless network [7], the network topology made it possible for us to mirror and trace the wireless traffic over the entire campus. Once we converted to a routed network, however, such mirroring was no longer possible. To capture only the wireless traffic now, it is necessary to mirror the traffic at each individual access point.

ITS agreed to use a trace gathering system we developed for this project to capture packet headers from a number of wireless access points on campus. Our trace gathering system was developed using a customized NetBSD kernel [9] and standard trace-gathering utilities (see below). It is configured to begin a new trace gathering session automatically upon each startup, operate continuously for long periods without direct monitoring and safely terminate the trace gathering process when shut down. This minimised the time and resources ITS needed to allocate to this project during the trace gathering period. Since our custom trace gathering system was specifically tailored for wireless user measurement research, the results recorded were far more detailed than those gathered using commercial network analysis tools [8] in our earlier study.

Our trace gathering system was deployed by ITS staff in three high-traffic campus locations between March 5 and May 3, 2004. The data we collected forms the basis of the Results section, below.

Anonymisation

ACS logs were sanitized by ITS using a custom-built anonymisation tool. These anonymised log files contain the same information as the actual ACS log, with two exceptions. First, the usernames contained in the anonymised logs were replaced with unique identifiers generated by the SHA1 one-way hashing algorithm. Second, events in the log that are not related to activity at wireless access points were removed. During our earlier study, the ACS logs were anonymised by simply stripping them of all private data fields, a process that greatly reduced the usefulness of the anonymised logs in our research. By hashing private identifiers we can now maintain user privacy without sacrificing the log's value to our research.

The trace gathering system we developed was also designed with user anonymity in mind. *Tcpdpriv* [10] anonymises *tcpdump*-formatted traces by stripping them of all packet payload information, leaving only the header fields for later analysis. Instead of gathering packet traces using *tcpdump* and then anonymising them afterwards, we opted to gather our traces using *tcpdpriv* directly. Although this restricts the depth of analysis we can perform on the trace data, it ensures that no private information contained in the wireless packets is ever recorded.

Analysis

We analysed the ACS log and packet header data using a combination of pre-existing and custom-written data analysis tools.

Storing the ACS log entries in a relational database allows for far more flexible and efficient analysis than was possible using the analysis scripts from our earlier study [7]. We used a custom-written data input tool to parse each of the nearly 400,000 log entries into individual fields. These fields were then inserted into database tables designed to store and analyse the log information efficiently. The database can then be used to select those table entries that match specific criteria quickly and return them to our analysis tools. The database can also perform more advanced queries, which join, group and summarize the data according to the values of particular fields. Entries in the log table can also be joined to other tables, such as a building-to-access-point relation, to analyse the log data according to other criteria.

Analysis of the packet header trace data was performed using the Coral Reef analysis package [11]. This software was used in previous studies [1] to analyse *tcpdump*-formatted traces.

Results

ACS Log Results

The ACS log data for this study consisted of almost 400,000 events collected over a nine-month period in the 2003/04 academic year (see Table 1). These events came from 710 users connecting at 78 different access points, installed in over 20 buildings. This represents substantial growth from the time of our earlier study when only 134 users were logged connecting at 18 access points. Authentication and deauthentication events occur with equal frequency throughout the log.

Table 1: ACS Log Summary

Attribute	*Value*
Total Events	399,103
Authentications	199,327
De-Authentications	199,776
Unique Usernames	710
Unique Machine Addresses	651
Access Points	78

Figure 1 shows usage by access point, expressed as both the number of users and the number of machine addresses seen. The discrepancy between the number of users and the number of machine addresses (Table 1) is due largely to the availability of a number of wireless cards which are loaned to students working in our main Library (13 cards were used by 5 or more distinct usernames over the course of the year). Three Library access points had the highest ratio of usernames to machine addresses. Of the 710 logged users, 155 connected to the wireless network through more than one computer. The

access point with the highest average of machine addresses per username was located at a help desk, where student laptops are configured by ITS staff who often use a single username to test numerous machines.

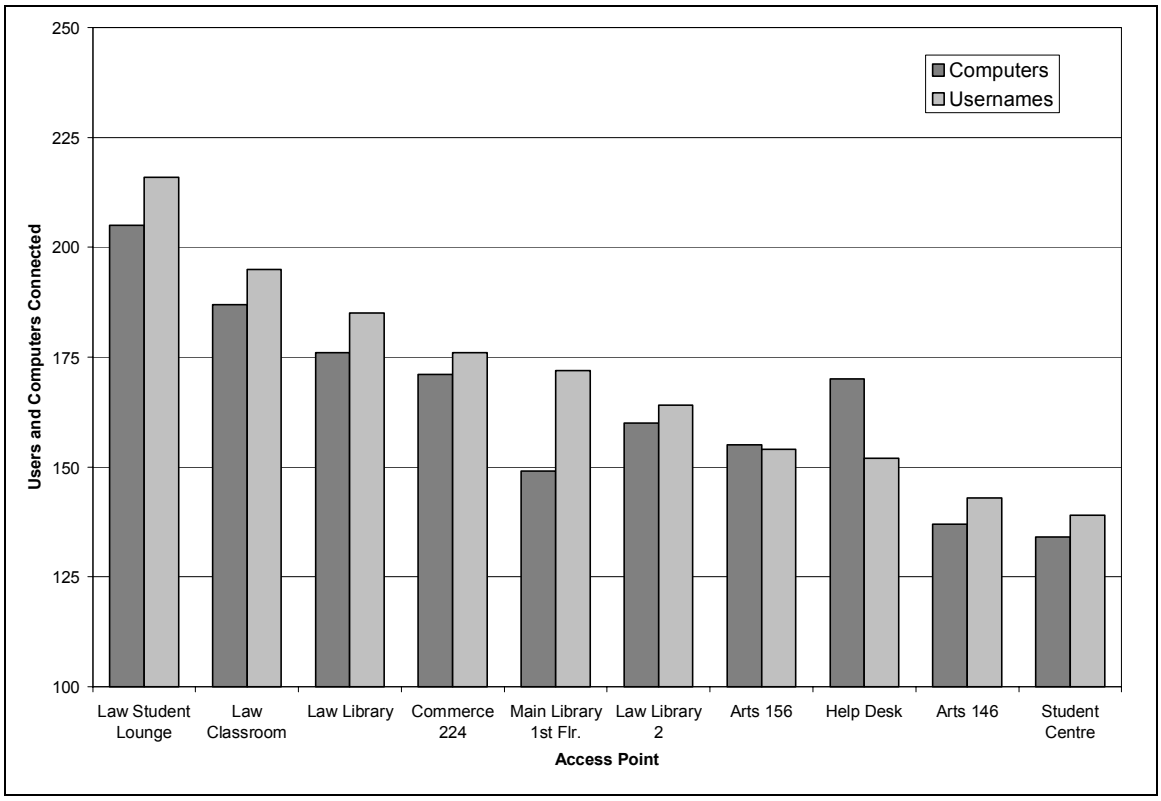


Figure 1. Unique Usernames vs. Machine Addresses at 10 Popular Access Points

Figure 2 shows the number of users per access point and the total number of users for each month in the study[11]. The usage levels were low during the summer, and climbed to a peak number of users per access point in November (which was also the month with the highest total number of authentication events). Usage dropped significantly (by approximately a factor of 2) in December because of exams and holiday closures.

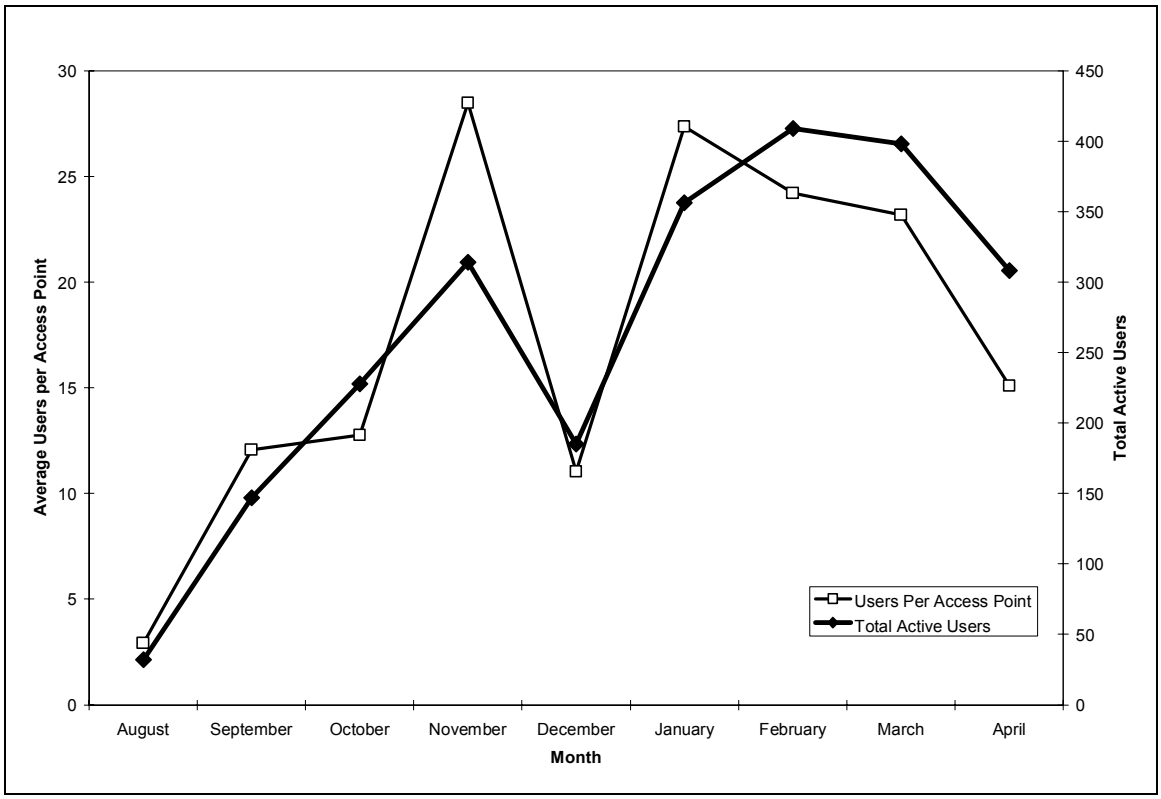


Figure 2. Average Users per Access Point vs. Total Users for Each Month

The number of total users continued to climb in the second term, but the average number of users per access point fell. We attribute this to both a decrease in roaming usage and an increase in the number of available access points as the wireless network expanded. Both usage measures fell late in April as exams and summer approached. Comparing the two terms, it is clear that the number of active wireless users grew significantly. Of the 710 total wireless users in the study, 447 were active from August to December 2003, and 609 were active in early 2004.

[1] ACS data covers dates between August 20, 2003 and April 17, 2004. August and April averages are based on available data.

Table 2: ACS Log Comparison

Statistic	*January 22-29, 2004*	*January 22-29, 2003*
Active Users	265	134
Active Access Points	48	18
Mean APs per User	3.12	2.99
Median APs per User	3	3
Mean Users per AP	17.25	22.28
Median Users per AP	5	14

By selecting only those authentication events which occurred between January 22 and January 29, 2004, we get a picture of how wireless network usage changed over the one year period following our earlier study in January 2003. Table 2 shows some basic statistics from this particular week-long January period. The number of active users has doubled and the average number of access points visited per user has risen slightly, while the median number of access points used remains constant.

The number of users connecting at each access point has changed more significantly. The mean number of users has dropped – but since the number of access points has more than doubled, a 22.5% drop in users per access point actually indicates an increase in overall network usage. The severe drop in the median number of users is due to the change in the distribution of users across the active access points, as shown in Figure 3. While in the earlier study, nearly 40% of the access points experienced above average usage, current usage is far more skewed, with only 23% of access points being accessed by a greater than average number of wireless users. In particular, the four most popular access points each experienced extremely high usage levels (100 users or greater) in January 2004.

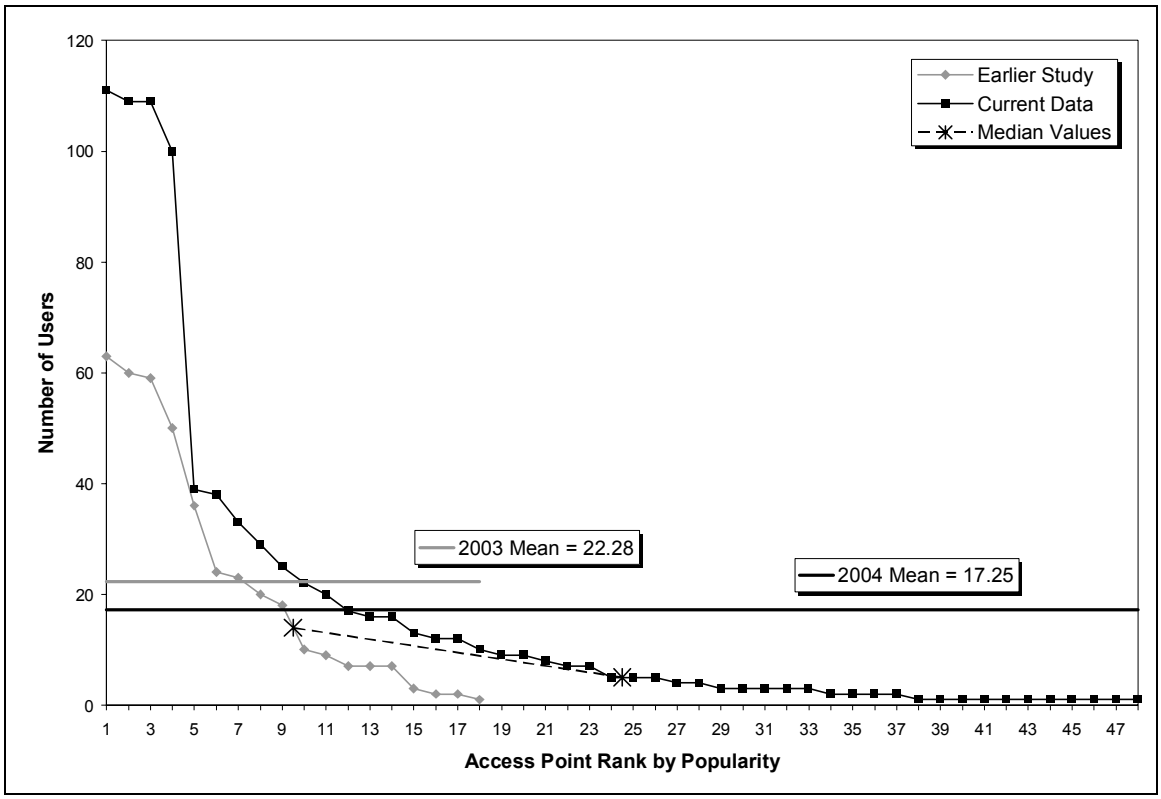


Figure 3. Change in Users per Access Point from 2003 to 2004 (Jan. 22-29)

Roaming Patterns

The roaming patterns of our users is something in which we are particularly interested — we want to determine the extent to which our users take advantage of the roaming opportunities wireless access affords them. Figure 4 shows the distribution of access points and buildings visited, for both the current data and the data from our earlier study. The highest point of both distributions occurs at one building/access point in the current data. This indicates that many users are connecting to the wireless network either as a wired connection replacement or as a local area network replacement, since they do not connect from other locations on campus.

The second highest point, which occurs at four access points visited, indicates that those users who do roam to multiple access points tend to connect at only a same small number of locations, even over long periods of time. In our earlier study, disproportionately heavy usage in our College of Law (an early adopter of wireless technology) resulted in a mode of 4 access points visited. Although usage in Law remains high, usage elsewhere on campus has risen significantly. Over 13% of our users visited more than ten access points (or more than five buildings) over the course of the 2003/04 term. The most actively roaming users on campus connected to almost half of the 78 access points currently installed.

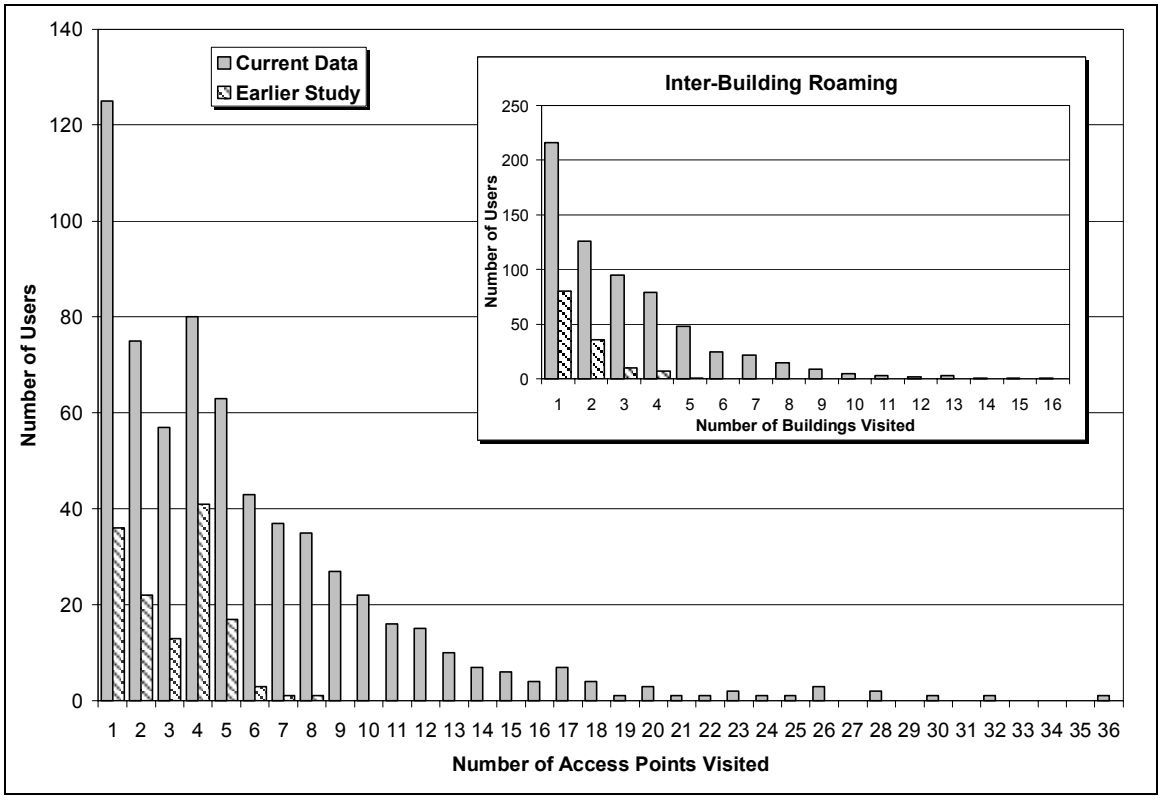


Figure 4. Distribution of Access Points Visited per User (Inset: Buildings Visited per User)

It is difficult to tell from Figure 4 whether or not the difference between the current data and the earlier data is due merely to the growth in both network size and user population. By normalising the cumulative distributions of the data over both the number of active access points visited and the total number of users in each dataset we can factor out the change in network size and popularity when comparing the two studies.

In Figure 5 we can see that the overall roaming behaviour *has* changed significantly. The fraction of users who do not roam at all (zero on the horizontal axis) has decreased by more than 5% between the two studies. The distributions cross above 30% of active users, as the increase in network size tends to shift lightly roaming users to the left. The most actively roaming users at the top of the two distributions show remarkably similar coverage of the active wireless network. In both studies, the top 3% of roaming users visit more than 22% of the wireless network and the furthest roaming users reach more than 40% of the access points. As the wireless network continues to expand, we expect to see a further decrease in the fraction of users who do not roam, but little change in the overall roaming behaviour of average and highly roaming users.

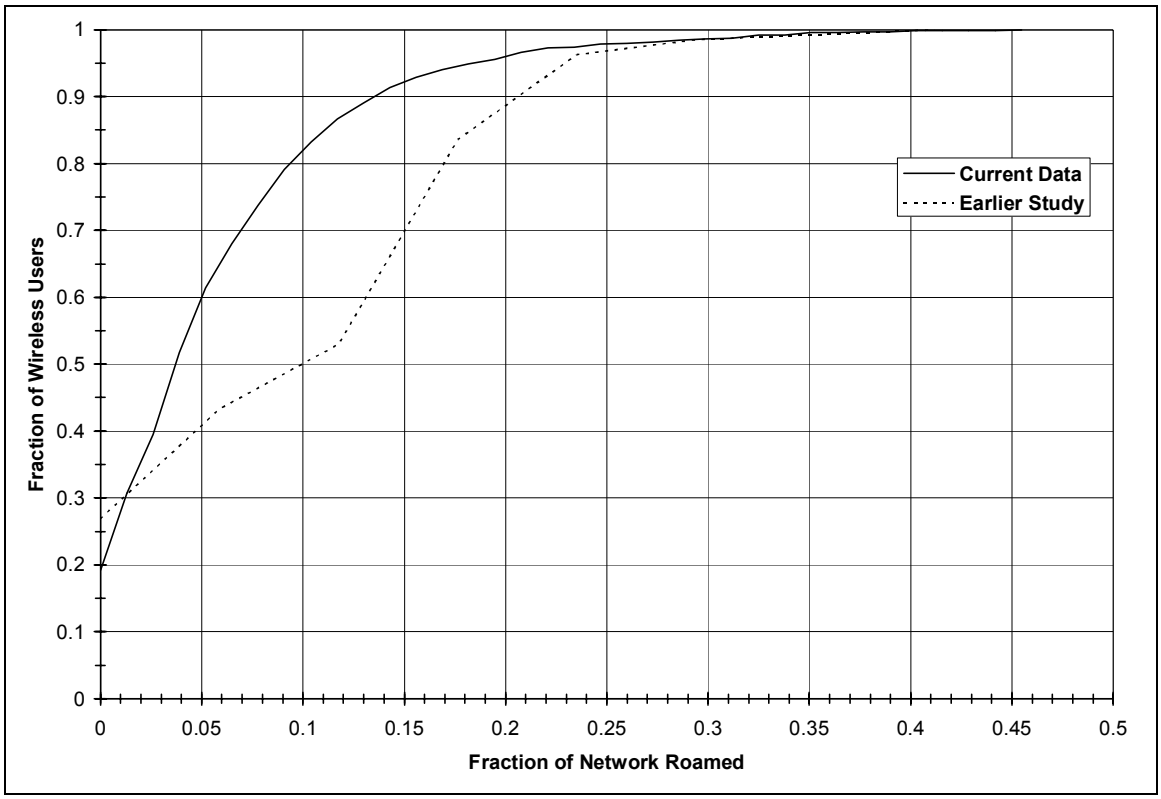
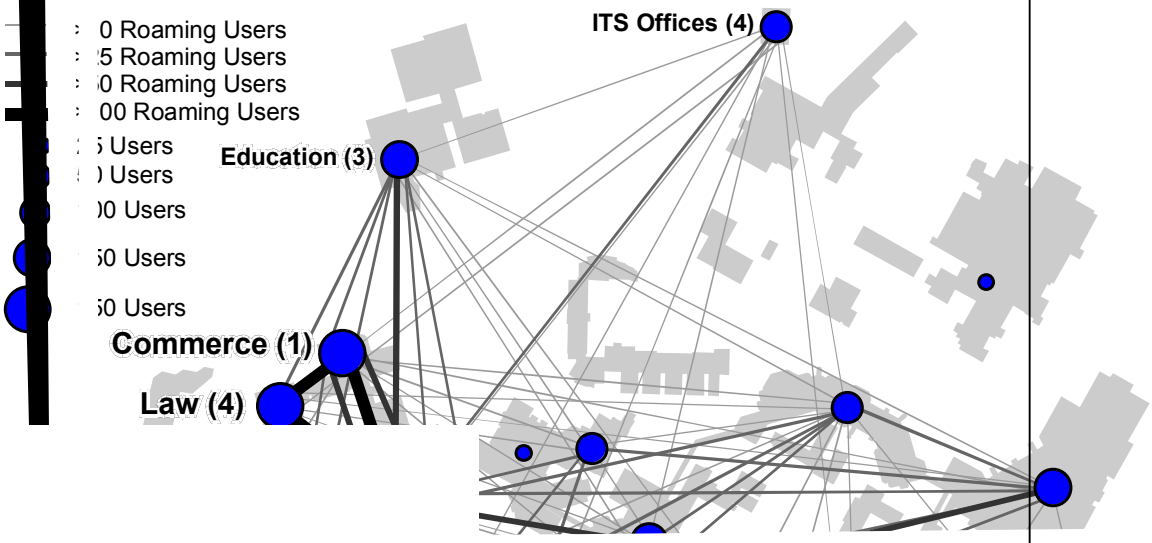


Figure 5. Comparison of Roaming Behaviour (Access Points Visited per User)

Visualizing the ACS log data on a map of the campus (Figure 6) gives a clearer picture of where our users are roaming. Each wireless-equipped building is marked with a circle, the radius of which is proportional to the number of unique users seen at that location. The thickness of a line connecting a pair of buildings indicates the number of users who visited both locations. The building names are printed adjacent to their circles, with the number of access points in the building in parentheses.


0 Roaming Users
5 Roaming Users
0 Roaming Users
00 Roaming Users
5 Users
) Users
)0 Users
60 Users
50 Users
ITS Offices (4)
Education (3)
Commerce (1)
Law (4)

roaming map shows that this skewed distribution corresponds to the underlying layout of the campus, with the most popular access points located in five adjacent, inter-connected buildings.

The second degree roaming links more densely inter-connect the five most popular locations and add connections to buildings with 100-150 users (Geology, Engineering, Health Sciences, and Education). Newly installed access points in locations such as Computer Science, Kinesiology and Agriculture were visited by 25-50 roaming users. More remote buildings, such as the ITS offices, St. Andrew's College and Animal Sciences saw even fewer roaming users. Newly installed access points in Royal University Hospital, Chemistry and Veterinary Medicine were used by only a small number of users, fewer than ten of whom roamed to other buildings. The relative disuse of the newest access points is consistent with the skewed distribution of users per access point observed in the ACS log comparison.

The roaming map suggests that a building's popularity with wireless users depends on user familiarity and location (newer and less accessible access points saw less usage) rather than the amount of wireless coverage available in a building. The single access point in Commerce, for example, was visited by several times more users than Kinesiology's nine.

Trace Data

As described in the methodology section, packet header traces were gathered at a number of campus locations in order to gain more detailed information on wireless usage. We used the Coral Reef toolset [11] to determine which protocols and applications are most commonly used.

Figure 7 shows the percentages of packets and bytes transmitted using each of the five IP protocols we saw in our trace data. Non-IP traffic accounted for 14% of the packets, but only 4% of the bytes, which indicates that almost all actual user data was transferred out as IP packets. ICMP (Internet Control Message Protocol), commonly used by network utilities and routers, accounted for 1.22% of the packets recorded. A small amount of local multicast management (IGMP) was also observed.

The primary IP protocols, UDP and TCP, were used in 85% of the packets, which carried 95% of the bytes on the wireless network. The significant use of UDP may be attributed to various forms of online entertainment, such as network games, peer-to-peer networks, and streaming media.

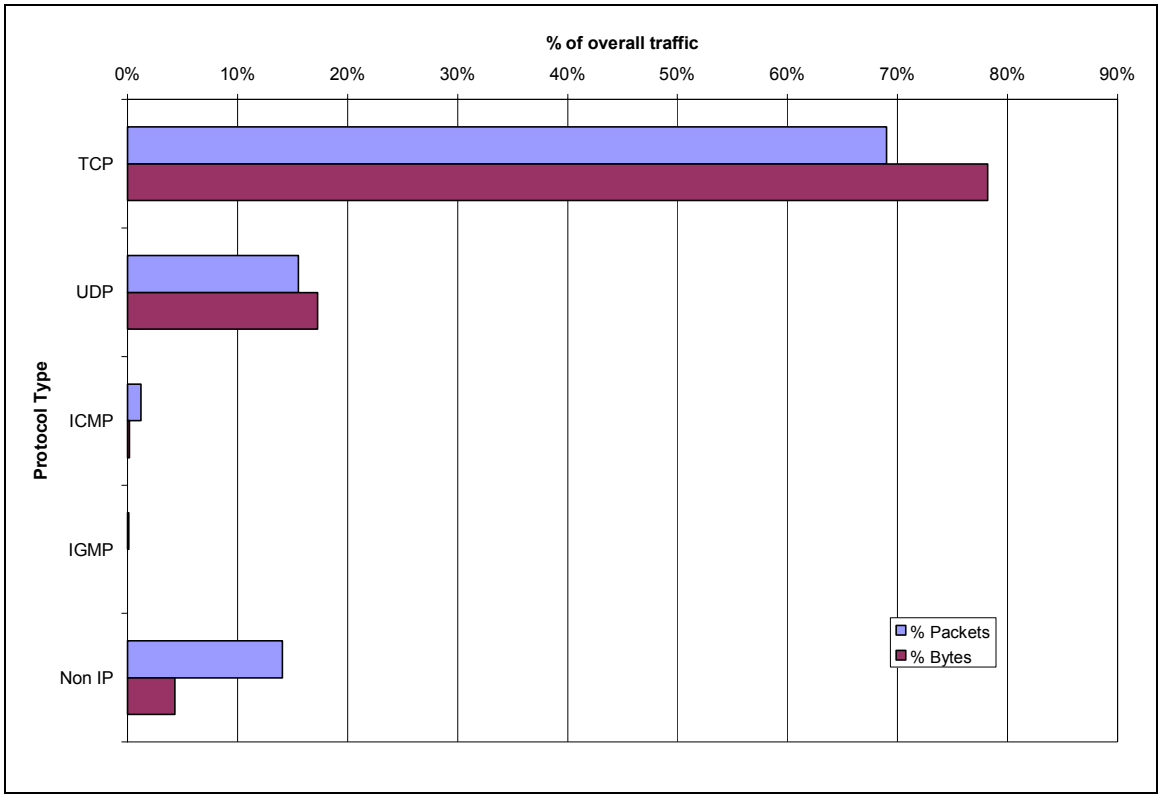


Figure 7. Trace Traffic by IP Protocol

Examining UDP and TCP traffic in further detail (Figure 8), we find that web access (HTTP and HTTPS) is by far the most common use of wireless networking — responsible for 71% of the bytes. Applications that use unallocated ports (above 1024) generated over 17% of bytes and 27% of packets. Audio and video sent via the Real Time Streaming Protocol (RTSP) was the second most common application, followed by network management (SNMP) information. All IP addresses on the campus are assigned dynamically using DHCP. Standalone (non-web) e-mail applications sent and received just over 1% of the packets and bytes traced, most of which was done through our IMAP-based student e-mail server. Roughly 7% of the packets seen on the wireless network were from various forms of network file access (such as NetBIOS, Appleshare, SMB, NFS) and file transfer (such as FTP).

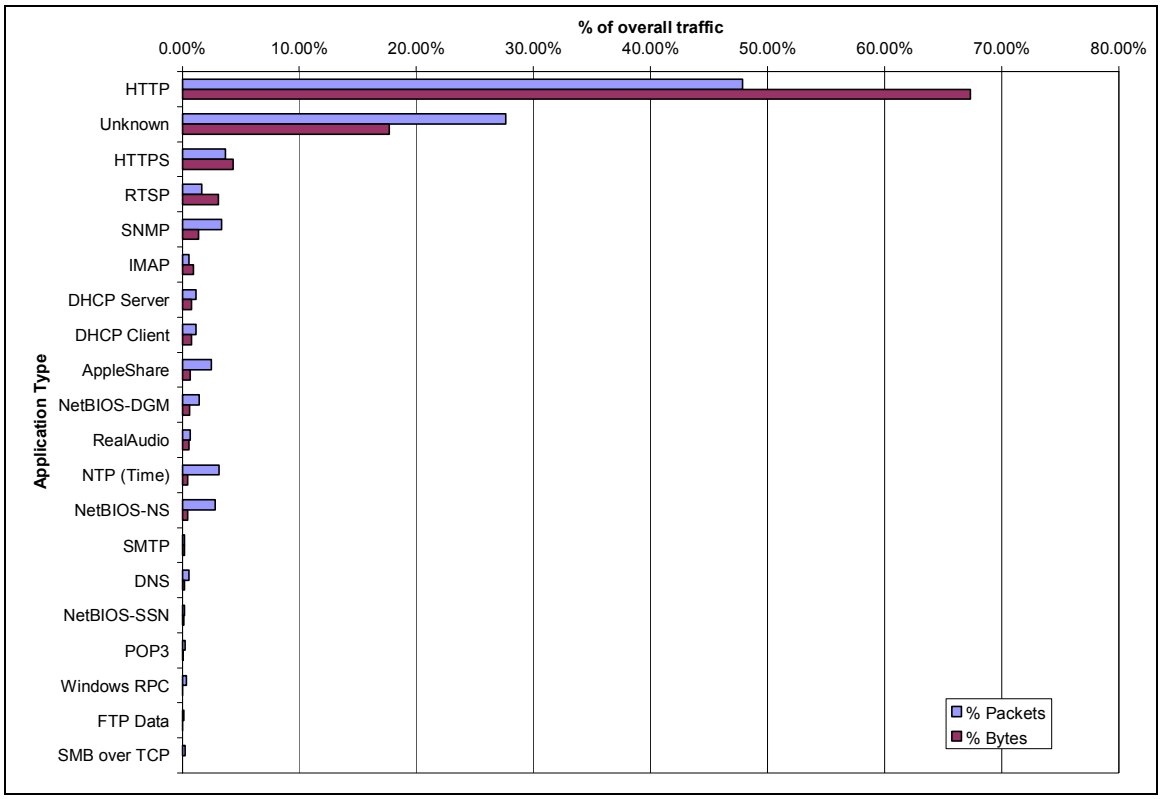


Figure 8. Top 20 Traced UDP and TCP Applications

Conclusions

As we deploy our campus-wide wireless network incrementally it is important that we understand the needs of our users. Through on-going analysis we seek to determine where, when, how much, and for what our network is being used, and how that usage is changing over time. In this paper we have described the methodology we are employing to study our usage patterns and some of the results we have obtained from our studies to date. Unlike other studies of wireless networks, our data is collected in a centralized manner made possible by the LEAP authentication system and the network environment

we have in place at our University. We augment anonymised ACS log data with localized packet header traces to enable a more complete analysis of user behaviour.

Both our wireless network and our wireless usage continue to grow. In the time elapsed between our first study and the present (roughly a year), the number of access points and the number of users have both more than doubled. We now provide at least partial coverage in about half of our campus buildings. The increase we are seeing in user demand certainly warrants continued expansion of this service. We also see a clear need to continue to study usage patterns to guide this expansion.

As we study usage patterns we are particularly interested in the roaming behaviour of our users, now and in the future. Our results to date suggest that the expansion of the wireless network over the past year has changed the roaming patterns of many users, and skewed the distribution of users per access point. Although most of our users still access a limited number of access points in a limited number of buildings we are seeing an increase in roaming behaviour with a larger fraction of our users roaming between buildings and the most active roamers visiting an ever increasing number of locations. We take this as clear evidence of growing demand for mobility support. The low use of some of our newest access points emphasizes that popularity is a function of familiarity and underlines our need to be more proactive in publicizing where coverage is available.

In terms of applications, web access via HTTP and HTTPS is by far the most common, accounting for more than 70% of our wireless traffic. Our next most popular application, audio and video via RTSP, is far behind. Non-web e-mail through our IMAP-based student server contributes very little of our wireless traffic.

The data capture methodology we have developed is successfully providing us the means to gain valuable insight into the usage of our campus-wireless network. Our results are guiding our planning by telling us two important things:

1) which buildings need more wireless coverage, and
2) which access points need more promotion.

This information will be very useful to us as we continue to evolve our service. Our current research is focussing on a more detailed examination of the roaming patterns of our users, and we hope to have more to say about this soon.

References

[1] A. Balachandran, G. Voelker, P. Bahl and V. Rangan. Characterizing User Behaviour and Network Performance in a Public Wireless LAN. In Proceedings of ACM SIGMETRICS '02, pp. 195-205, Los Angeles, CA, June 2002.

[2] TCPDUMP, http://www.tcpdump.org.

[3] D. Kotz and K. Essien. Characterizing Usage of a Campus-wide Wireless Network. In Proceedings of ACM MobiCom '02, pp. 107-118, Atlanta, GA, September 2002.

[4] D. Tang and M. Baker. Analysis of a Local-Area Wireless Network. In Proceedings of ACM MobiCom '00, pp. 1-10, Boston, MA, August 2000.

[5] F. Chinchilla, M Lindsey and M. Papadoupouli. Analysis of Wireless Information Locality and Association Patterns in a Campus. In Proceedings of IEEE Infocom 2004, Hong Kong, March 2004.

[6] S. Convery and D. Miller. SAFE: Wireless LAN Security in Depth – version 2. White Paper, Cisco Systems Inc., San Jose, CA, March 4, 2003.

[7] D. Schwab and R. Bunt. Characterising the Use of a Campus Wireless Network. In Proceedings of IEEE Infocom 2004, Hong Kong, March 2004.

[8] EtherPeek, http://www.wildpackets.com/

[9] NetBSD, http://www.netbsd.org.

[10] Tcpdpriv, http://ita.ee.lbl.gov/html/contrib/tcpdpriv.html.

[11] D. Moore, et. al. The CoralReef Software Suite as a Tool for System and Network Administrators. In Proceedings of the 15th Systems Administration Conference (LISA 2001), pp. 133-144, San Diego, CA, December 2-7, 2001.

Footnotes

[1]ACS data covers dates between August 20, 2003 and April 17, 2004. August and April averages are based on available data.