Abstract—We present the results of an analysis of the usage of our new campus-wide wireless network. A week-long traffic trace was collected in January 2003, recording address and protocol information for every packet sent and received on the wireless network. A centralised authentication log was used to match packets with wireless access points. The trace was analysed to answer questions about where, when, how much, and for what our wireless network is being used. Such information is important in evaluating design principles and planning for future network expansion.

Keywords—Wireless network measurements; Wireless LANs; Traffic analysis; Network design & planning

I. INTRODUCTION

The University of Saskatchewan 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 recently introduced 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 small number of Cisco access points (18) placed strategically in a number of locations. The demand for wireless access continues to grow, and the network is being expanded to offer higher capacity and greater coverage to the campus.

In order for us to plan for this expansion, it is important that we understand current usage patterns— that we understand where, when, how much, and for what our wireless network is currently being used. This paper describes the methodology we employed to collect data on usage, and presents the results of our analysis. Although our wireless network is small at the present time, a description of our approach to this analysis will be useful to others planning similar studies.

Usage data was collected in co-operation with our Information Technology Services Division (ITS) over the period of one week in January 2003. Although no effort was made to ensure this week was representative of overall usage patterns, we feel that the data we collected represents a useful snapshot of the usage of a campus wireless network.

The paper is organized as follows. Section II reviews related work in wireless network measurement. A brief description of the wireless network in place at the time of our data collection follows in Section III. In Section IV we describe the methodology followed when gathering and analysing the data, including steps taken to ensure user confidentiality. Section V contains the results of our analysis. We conclude in Section VI with a summary of our findings.

II. RELATED WORK

The design of this experiment was based largely on work done by Balachandran et al. [1]. Their analysis and characterization of the traffic generated by attendees of an ACM conference 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 trace was done using tcpdump 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 [2] 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 decentralised 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 [3] 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.

III. NETWORK ENVIRONMENT

At the time of data collection our campus wireline network operated as a switched network\(^1\). This means that all the equipment connected to the University of Saskatchewan campus network was part of the same subnet. The campus network is connected to the internet via a Cisco router.

For security purposes, our wireless network consists of a virtual network, existing on a separate subnet from the rest of the campus. Packets sent from and to wireless devices travel on the same physical network as normal campus traffic. Upon connection, wireless devices on the network are assigned internal IP addresses by a DHCP server. All wireless traffic is then sent to the campus router, which routes it either to the internet or back onto the campus network in the normal subnet. This prevents unauthorized wireless users from connecting directly to campus servers and the internet.

Since wireless and non-wireless packets travel on the same physical network, capturing only the wireless traffic proved to be much more of a technical challenge than in previous studies. Fortunately, since the wireless traffic is routed to and from a distinct subnet, it was possible for us to distinguish wireless packets from normal traffic. By re-programming the campus router, we were able to mirror those packets originating from or travelling to the wireless network. The wireless traffic was mirrored to a router port that was monitored by a trace gathering computer.

At the time of measurement the first 18 Cisco access points were up and running. The selected locations spanned the campus covering a wide range of possibilities, from public spaces (lounges, libraries, even a coffee shop), to classrooms and laboratories, to office spaces, to allow ITS to gauge the nature of the user demand. The availability of this technology was not well-advertised, although wireless PCMCIA cards were offered at a special price through the Campus Computer Store and the locations of the access points were provided on our student computing web site\(^2\).

IV. METHODOLOGY

A. Trace Collection

We ran a software package known as EtherPeek on a dedicated computer to collect our trace. Although EtherPeek is designed to allow network administrators to monitor activity, its trace gathering capabilities were deemed sufficient to meet the needs of this project. While other applications might have been more suitable for trace gathering, EtherPeek’s ability to record MAC addresses allowed us to analyse the usage patterns in greater detail.

A short trace was recorded as a trial run for the trace gathering system. Doing this allowed us to determine the traffic level on the wireless network, which would dictate the bandwidth and storage requirements of the trace gathering computer. The traffic level was low enough that, with occasional dumps to CD, storage and bandwidth were not problematic for the size of trace we wanted.

Trace collection started on Wednesday, January 22, 2003 at 9:07 AM local time. Each packet sent from and to the wireless network was mirrored to our trace gathering computer. EtherPeek analysed each packet individually, and recorded information such as the date, time, origin, destination, and protocol. Trace collection stopped one week later on Wednesday, January 29 at 8:37 AM. The trace data was later exported from EtherPeek as a series of comma-separated-value (CSV) files.

B. Authentication Logging

For security our network administrators have deployed Cisco’s proprietary LEAP authentication system [4] to control access to the wireless network. This requires that any potential user provide a username and password, which are verified by a central Radius server before a connection is established. The LEAP Radius server keeps track of every wireless user currently connected to the network. This information is also logged for security monitoring purposes. The log includes a record of each authentication, including the date, time, username, client MAC address and the IP address of the access point the user is connected to. To allow us to determine where users were connecting to the network, we were provided with a one week anonymised portion of the authentication log corresponding to the week of the trace. By matching the MAC addresses in the authentication log with the MAC

---

1 We have since converted to a routed network.

2 http://studentcomputing.usask.ca/
addresses on the packets, we could sort the data by access point.

C. Anonymisation

Since the packets we were monitoring were being sent from and to ordinary users, every precaution was taken to ensure that their identities would remain anonymous and that no private information would be revealed in the trace. The authentication log was stripped of user identifications, leaving only machine addresses as identifiers. Since these machine addresses are assigned by the network card manufacturer, they cannot be used to reveal the identities of individual users. The IP addresses present in the trace were only temporary addresses assigned by the campus DHCP server. Since the trace was being analysed weeks after it was recorded, the IP addresses were no longer current, and therefore could not be used to identify individual users. The trace itself contained only information gathered from the headers of the wireless data packets. Since no message bodies were included in the trace, no private information was revealed.

D. Analysis

The analysis of the trace files and authentication log was done primarily with custom-written Perl [5] scripts. Perl’s simple file i/o, associative arrays, powerful string handling, flexible data types and regular expression matching capabilities make it especially well-suited to this application.

Our initial analysis of the trace files was a simple validation and error-checking pass. This revealed that several portions of the trace contained mis-ordered or erroneous data. The erroneous data was due to a small number of malformed or non-standard packets being incorrectly identified and analysed by EtherPeek. The resulting trace entries contained binary data which the analysis scripts could not parse correctly. Due to human error when transferring trace data, some files were named incorrectly and parsed in the wrong order. This was detected by calculating the time difference between the last packet of one file and the first packet of the next.

The problems were resolved by sorting the mis-ordered files and removing the erroneous data. The end result of this validation was 7 days worth of wireless packet trace with no apparent gaps, mis-ordered packets or errors. This was what we expected, since the wireless network experienced no outages during the period of the trace, and the data was gathered directly from the router and authentication log.

The second stage of the analysis focused solely on the trace files. Studying the aggregate traffic patterns and the protocol mix gave us an initial understanding of the characteristics of the data. The results of this analysis are discussed in Section V.A. Next, we began studying the authentication log. Looking at only authentication times, network card addresses and access points addresses revealed characteristics of the users on the wireless network and the access points they connect to during the traced week. Results from the authentication log are presented in Section V.B.

```
While (Read Next Trace Line)
  Split Trace Line into Packet Fields
  While(Packet Time > Next Auth Time)
    Update Auth Table (Next Auth)
    Read Next Auth Line
    Split Auth Line into Auth Fields
    If (Auth Table contains Packet From)
      Count as Wireless Packet
    Else If (Auth Table contains Packet To)
      Count as Wireless Packet
    Else
      Count as Non-Wireless Packet
```

Figure 1: Trace and Authentication Log Parsing Algorithm

The final stage of analysis combined the trace data and authentication log to match packets with access points. In order to map packets to access points, a user/location lookup table had to be constructed and updated from the authentication log. The algorithm used to simultaneously traverse both the authentication log and the trace files is outlined in Fig. 1. This stage revealed the most detailed information about the wireless network, presented in Section V.C.

V. RESULTS

A. Trace Data

In Table I we summarize the information contained in the trace log. Due to a configuration error, packet sizes were recorded for only a small fraction of the packets traced. As a result, we have characterised the wireless traffic in terms of the number of packets. Fig. 2 shows the traffic over the entire length of the trace. The traffic level rises each day at around 9:00 AM and remains high until the evening. Between each day, the traffic level remains relatively steady at around 15 packets per second. As we will show, this base level of traffic was due to non-wireless traffic that was multicast onto the wireless network as part of automated network maintenance. On Saturday and Sunday, the traffic did not increase as early or as much as on the weekdays. This is expected on a campus network, since students and staff use the campus much less on weekends.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Packets</td>
<td>24,431,794</td>
</tr>
<tr>
<td>Total Seconds</td>
<td>603,054</td>
</tr>
<tr>
<td>Average Traffic</td>
<td>40.5 packets per second</td>
</tr>
<tr>
<td>Total Time</td>
<td>6 days, 23 hours, 30 minutes, 34 seconds</td>
</tr>
</tbody>
</table>
B. Authentication Data

Looking at the authentication log alone, we were able to determine several key characteristics of the wireless network’s users and access points as summarised in Table II. Over the week-long period of the trace, 134 unique users (machine addresses) connected to the network. The authentication log contained 24973 records of wireless users authenticating. The average number of authentications per user (186.4) seems high. A graph of the cumulative distribution function shows that most users authenticated many fewer times. From Fig. 3 we can see that over half our users authenticated more than 50 times during the week. This seems like a large number of authentications, but there are several contributing factors. Cisco’s wireless network card drivers store the username and password information permanently and authenticate automatically whenever the computer is near a wireless access point. Additionally, the authentication log reveals that users often re-authenticate at the same access point several times a minute. This is likely due to low, fluctuating signal strength at the edge of an access point’s signal range. When a user is just within range of two or more access points, the software will often switch connections repeatedly on the basis of the perceived signal strengths at each access point. These factors cause an artificial inflation of the number of authentications per user, meaning that authentications cannot be literally interpreted as distinct sessions of network usage. This rapid AP switching behaviour was also observed in previous studies [2].

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Authenticizations</td>
<td>24973</td>
</tr>
<tr>
<td>Unique Users</td>
<td>134</td>
</tr>
<tr>
<td>Mean Authentications Per User</td>
<td>186.4</td>
</tr>
<tr>
<td>Mode Authentications Per User</td>
<td>5</td>
</tr>
<tr>
<td>Median Authentications Per User</td>
<td>54.5</td>
</tr>
<tr>
<td>Access Points</td>
<td>18</td>
</tr>
<tr>
<td>Mean Authentications Per AP</td>
<td>1387.4</td>
</tr>
</tbody>
</table>

Fig. 4 shows the total number of authentications at each of the 18 access points observed over the period of the trace. Once again, the distribution of authentications is heavily skewed, with a small number of access points accounting for the vast majority of authentications.
There is particularly heavy usage in our College of Law. The reasons for this are discussed in C(3) below.

C. Combined Data

1) Non-Wireless Traffic

Approximately 38% (a total of 9,230,131) of the packets recorded in the trace data could not be associated with any of the users found in the authentication log. These non-authenticated packets were present in every second of the trace. They arrived at a near-constant rate of approximately 15 packets per second throughout the week, and did not vary with the number of wireless users on the network. Furthermore, the average arrival rate of these non-wireless packets cycles between a high and low rate with a period of approximately 300 seconds (5 minutes). These characteristics led us to conclude that this traffic was not being generated by wireless users. Since this non-wireless traffic did not vary with the time of day it must have been generated by automated equipment on the network. Given our network environment we conclude that this traffic is automated network-maintenance traffic generated by switches and other network devices on the campus and flooded onto the wireless subnet. This conclusion is supported by the protocol makeup of the unauthenticated traffic presented in the next section.

If this traffic is indeed being flooded from the wired campus network onto the virtual network which carries the wireless traffic, then it might represent some degree of unnecessary overhead. Since the wireless network is configured independently, and since the router which connects the two has been configured properly, local network maintenance traffic should not be passed from one network onto the other. Duplicating multicast maintenance packets onto the virtual wireless network needlessly doubles the maintenance overhead on the underlying physical network. While the size of these packets may be trivial, a continuous arrival rate of 15 packets per second means they represent a significant amount of traffic. Eliminating these duplicate packets from the wireless subnet could improve the performance of the wireless network.

At this time we have been unable determine why these non-wireless packets are present in the trace. It might be because they were actually routed onto the wireless subnet or simply because the router mirrored both wireless and multicast traffic to our monitored port.
Fig. 6 shows the protocol mix for the actual wireless traffic. As would be expected, web browsing (HTTP, HTTPS) and other common applications, including file sharing (CIFS, SMB, NetBIOS, NB), file transfer (FTP), e-mail (IMAP, POP3, SMTP), instant messaging (AOL, MSN), peer to peer (Gnutella), remote shell (SSH, TELNET), and network services (PING, DNS) dominate the user-generated traffic. Other TCP traffic from unidentified applications makes up 34.6% of these wireless packets. The fact that these protocols are those of end-user applications offers further confirmation that this traffic is user-generated and that the non-wireless traffic is not.

3) Traffic vs. Authentications

In Fig. 7 we compare the number of authentications at each access point to the number of packets sent to and from users authenticated at that access point over the course of the trace. Although the access points in Law generated the most authentications, the average number of packets associated with each of those authentications is much lower than elsewhere on the campus. This means that the number of authentications is not directly related to the packet rate. The difference between Law traffic and other traffic could be attributed to a difference in the usage patterns of our law students. It could also be due to poor signal strength or high levels of interference in the Law Building causing repeated disconnections and reconnections.

4) Daily Traffic Patterns

Fig. 8 shows the average traffic in packets per second for each day of the trace at each access point. The graph is separated into three parts for clarity. Several distinctive features can be observed. Generally speaking, older access points with which users are familiar experienced higher usage levels than newly installed access points.

Access points located in and near offices, such as those in the Administration Building, at the ITS Help Desk, in the ITS offices, and in the Computer Science
offices in the Engineering Building show very little traffic over the weekend. The Animal Sciences building houses another part of the Computer Science Department, including some research labs and associated graduate students. The low but constant minimum packet rate observed at Animal Sciences was likely caused by automated programs used by these students. Saskatchewan Hall is a residence, and the access point there was new and not very well known in January, so the small amount of traffic seen there is likely from a single resident. Browsers is a campus coffee shop. Since it is closed on weekends, its traffic drops to zero.

Popular public areas for socializing and studying such as Place Riel (our student centre), the Learning Commons in the main library and a lounge outside Arts 145 (a busy Computer Science undergraduate lab) were used on every day of the trace. Arts 110 (another student computing laboratory) is situated in the classroom wing near a large computer lab. And while this would appear to be a logical place to install wireless access, the lack of any non-classroom tables and chairs gives laptop users few places to comfortably connect from while not in classes, causing the minimal usage observed.

5) Roaming

We were especially interested in the degree to which users roam between access points and we were able to study this using the authentication log. Roaming is an important characteristic since it reflects the degree to which users are taking advantage of the unique benefits of wireless networking. Mobile computing is facilitated by a wireless network not only by allowing users to remain connected while travelling between adjacent access points, but also by allowing users to connect while stationary in a variety of separate locations.

Fig. 9 shows the distribution of the number of access points visited by each unique user. This graph shows that individual users visited at most 8 different access points over the period of the trace. On average, users visited 3 access points that week.

By counting the number of unique users (network card id’s) that were authenticated at two different access points, we gain an understanding of the number of users who roamed between the two locations. Graphing these on a map of the campus reveals where the most mobile users connected. In Fig. 10 the large blue circles contain the number of unique users who authenticated at each access point. Each circle is connected to the location of
increased use and more roaming. advertised and easily identifiable will encourage sure the locations of wireless accessibility are well covered area they will not attempt to connect, even if access point locations. If they believe they are in a non-connected areas of the campus. For these users, use roaming is largely dependent on their knowledge of access point locations. If they believe they are in a non-covered area they will not attempt to connect, even if there is an access point available. Therefore, making sure the locations of wireless accessibility are well advertised and easily identifiable will encourage increased use and more roaming.

6) The College of Law
For a number of reasons, data associated with access points located in the College of Law present special opportunities to gain insight into student usage patterns. The access points in the Law Library, the Moot Court (a classroom for practising trial law) and the Law Student Lounge are all heavily used, accounting for almost 86% of the entries in the authentication log. The College of Law generated over one third of the total number of wireless packets in the trace. Users who connected in the Law Building also roamed between access points far more frequently than users in other parts of the campus. There are several factors contributing to this heavy usage.

Even before the introduction of wireless technology, the Law Library has long been a central location for law students to read, write, study and work. After wireless networking became available, law students found the ability to stay connected while in the library study space invaluable. The College of Law has since made a major commitment to wireless technology on campus. The college has supported and expanded wireless access within the Law Building. As a result of this shift to wireless connectivity, traditional wired computer labs in the College of Law have been closed. Wireless networking offers much greater flexibility and connectivity for students and much lower maintenance costs for the college.

Another significant factor influencing the popularity of wireless access within the College of Law is a paradigm shift within the legal community as a whole. In the span of a few years the legal profession has transitioned to digital technology in an unparalleled way. A new emphasis on online files has made laptop computer ownership a de facto requirement for law students. For students who work with online legal documents before, during, between and after classes each day, wireless connectivity allows free and flexible access that no number of computer labs could ever achieve.

7) Design Principles
Our approach to wireless service deployment at the U of S is based on several guiding principles. These principles were derived from constraints such as the geography of the campus and the technology chosen. In order to maximize the usefulness of the network, our first access points were installed primarily at locations which best met the following criteria:

- In our view, mobile computing should focus on location rather than movement. The usefulness of a wireless network is directly proportional to the probability that a user will be able to get a connection in places he/she normally visits. On a highly distributed campus such as ours, it is more important to install access points in classrooms, libraries, labs and lounges than it is to have complete coverage between locations. For this reason, we will continue to implement “islands” of connectivity rather than “continuous corridors” of wireless coverage.

- As more course notes, announcements, textbooks, reference materials and other material becomes available online, the need to have ready access to that material will increase. Therefore, preference is given to placing access points in classrooms, libraries, labs and lounges than it is to have complete coverage between locations. For this reason, we will continue to implement “islands” of connectivity rather than “continuous corridors” of wireless coverage.

- Wireless technology is designed for users of modern mobile computers. In order to capitalize on existing user owned devices, priority should be given to areas which are already home to a large number of mobile users. More specifically, students in high-tech or professional programs are most likely to own mobile devices which would benefit from wireless access. By improving and expanding wireless access in professional and high-tech
 colleges we can further capitalize on existing user-owned mobile devices.

Our experience in the College of Law confirms that these guiding principles were an appropriate choice for effective deployment in our environment. A high level of availability was achieved for law students with only a small number of access points. The demand for wireless connectivity came from both existing student behaviour and decisions by the College of Law to focus on online resources and wireless resources. Accessibility was facilitated by the pre-existing popularity of mobile computing in the college. The high level of wireless adoption experienced in Law offers proof that these principles are sound and gives a clear direction for future wireless development.

VI. CONCLUSIONS

In this paper we have presented a methodology for capture and analysis of traffic patterns on a campus-wide wireless network. Unlike previous studies of wireless networks, our trace was collected in a centralized manner made possible by the LEAP authentication system and the network environment in place at the University of Saskatchewan. Our centralized network design made possible a more complete and error-free trace than those analysed in previous projects. The goal of our analysis was to determine where, when, how much, and for what our campus wireless network is being used.

While a single week might not be representative of overall usage patterns, we feel that the results presented here do offer some real insights into how campus wireless networks are used. We were able to show that an average wireless user on our campus connected only a small number of times in the week from a limited number of access points. The popularity of a given access point was largely determined by its accessibility and familiarity to users. In our results, a small number of our access points generated a majority of authentications, but a greater number of authentications did not correspond to a higher traffic level. Roaming levels between access points in our trace were related to the relative proximity and popularity of the access points.

The usage patterns in the College of Law provide an excellent example of the success of our wireless network design. By making wireless access available in popular locations, increasing the demand for connectivity and capitalizing on user owned devices, the College of Law achieved a high rate of wireless adoption and maximized the value of wireless networking for its students. This success confirms our guiding principles for deployment and enables us to maximize the usefulness of future expansions.

Moving forward, we are expanding the wireless network and continuing to study usage patterns. Ongoing analysis will continue to be vital in guiding the expansion of our network to enable us to better serve a growing mobile user base. In early 2004 we plan to perform a longer, more in-depth trace of the wireless network. This larger study will perform more in depth analysis of the traffic at key access points on campus, and will examine roaming behaviour over a much longer period in order to provide a more representative characterisation of the use and performance of an expanding campus wireless network.

We are developing new metrics and tools to measure the performance of the network and the behaviour of mobile users more accurately. In the long term, we hope to develop a methodology and a toolset which can be used to characterize the use and performance of a campus sized wireless network easily and accurately.

ACKNOWLEDGMENTS

We would like to thank Glenn Hollinger, Gary Berg, and Daniel Brad of Information Technology Services for their advice and help throughout this project. Financial support for this research is provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) and by TRLabs, Saskatoon.

REFERENCES