Available online at www.sciencedirect.com
ScienceDirect
Performance Evaluation 64 (2007) 909–932
PERFORMANCE EVALUATION
An International Journal
www.elsevier.com/locate/peva

Remote analysis of a distributed WLAN using passive wireless-side measurement

Aniket Mahanti*, Carey Williamson, Martin Arlitt

Department of Computer Science, University of Calgary, Canada

Available online 21 June 2007

Abstract

This paper presents network traffic measurements from a campus-wide wireless LAN (WLAN), with the data collected using remote passive wireless-side measurement. We used commercially-available monitoring devices to collect wireless traffic concurrently from 9 selected locations on the campus WLAN for 6 weeks. The aggregate trace contains almost 1 billion wireless frames, representing the WLAN activity generated by 6775 users and 97 access points. Analysis of the dataset identifies similarities and differences in the user behaviours across the observed WLAN locations, as well as emerging trends in WLAN usage regarding application usage and session mobility. Our study extends existing WLAN measurement studies by providing deeper insights into how WLANs are used, and by developing models of WLAN usage characteristics that are applicable in capacity planning, network testing, and network simulation studies.
© 2007 Elsevier B.V. All rights reserved.

Keywords: IEEE 802.11; Passive measurement; Traffic characterization

1. Introduction

Wireless Local Area Networks (WLANs) are commonplace in many university campuses. For campus network administrators, WLANs are an economical solution for providing Internet access in both academic (e.g., classrooms, libraries) and non-academic (e.g., lounges, athletic facilities) areas of the campus. For network users, WLANs offer the convenience of Internet access from any location at any time. This flexibility is important as users become increasingly dependent on the Internet, and ever-more demanding in their usage of the WLAN. The usage trends observed on campus WLANs often transcend many other environments as well, including enterprises and commercial wireless hotspots.

As WLANs grow in size, scale, and complexity, the challenges for WLAN traffic measurement also grow. Network administrators require suitable network monitoring infrastructure for the purposes of network management (e.g., accounting, network intrusion detection, usage policies, traffic shaping). Networking researchers also require suitable infrastructure for network traffic measurement purposes.

* Corresponding address: Department of Computer Science, University of Calgary, 2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4. Tel.: +1 220 6015; fax: +1 284 4707.
E-mail addresses: amahanti@cpsc.ucalgary.ca (A. Mahanti), carey@cpsc.ucalgary.ca (C. Williamson), arlitt@cpsc.ucalgary.ca (M. Arlitt).

0166-5316/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.peva.2007.06.004

Table 1
Summary of main results and observations from WLAN measurement study

View	Observation	Section
User	Wireless laptops running the Windows operating system are the most prevalent devices observed using the WLAN.	6.1
	Usage exhibits strong diurnal patterns.	6.2
Application	In terms of byte traffic volume, Web browsing is the most popular network application, followed by P2P file sharing, media streaming, and electronic mail.	7.1
	Inbound traffic dominates outbound traffic and internal traffic.	7.2
Mobility	The number of APs visited per user is geometrically distributed.	8.1
	Roaming events are temporally and spatially correlated, with movement patterns between locations influenced by geographic proximity.	8.2
Session	The number of sessions per user follows a Logarithmic distribution.	9.1
	Stationary session durations can be modelled with a two-parameter Weibull distribution; mobile session durations follow an Inverse Gaussian distribution.	9.2
	While relatively few sessions (11%) involve mobility, the median mobile session (2 h) lasts much longer than the median stationary session (44 min).	9.3
Network	Overall usage of the WLAN is relatively light, but non-uniformly distributed amongst APs.	10.1
	The relationship between number of users and AP traffic load is weak.	10.2
Wireless	Frame transmissions occur on all 11 of the IEEE 802.11b/g channels, though channels 1, 6, and 11 account for over 99% of the activity.	11.1
	CRC errors (11%) and MAC-layer retransmissions (25%) are common on the WLAN.	11.2

The primary challenges for WLAN measurement include the geographic diversity of WLAN deployments, the physical proximity required for WLAN packet capture, and the need for a wireless-side (rather than just wired-side) view of the network. Coupled with the growing complexity of WLAN deployments (e.g., multi-channel IEEE 802.11a/b/g networks), the heterogeneity of user equipment (e.g., different devices, operating systems, and protocol stacks), and recent trends in Internet usage (e.g., messaging, gaming, P2P file sharing, video streaming), geographically-distributed campus WLANs present a challenging environment for network traffic measurement.

In this paper we demonstrate the viability and utility of a remote passive wireless-side measurement methodology in a large distributed production WLAN. We used commercially-available monitoring devices called Radio Frequency Grabbers (RFGrabbers) [25] to collect wireless traffic from 9 selected locations in 7 buildings on the campus WLAN. With judicious decisions regarding packet filtering and capture, our study collected wireless-side traces for about 6 weeks. The aggregate trace contains almost 1 billion wireless frames, representing the WLAN activity generated by 6775 users at 97 access points (APs). These traces provide a view of the WLAN user activity for about one-half of an academic term. Our approach demonstrates a practical and commercially-available solution for network administrators to manage a large campus WLAN. It also shows an approach to WLAN traffic measurement that is suitable for networking researchers.

To illustrate the capabilities of wireless-side measurement, we present a comprehensive multi-layer analysis of our WLAN datasets, from the application layer to the wireless link layer. Our WLAN measurements identify emerging WLAN trends, particularly those regarding application usage and session mobility, that have performance implications on capacity planning for future WLAN deployments. Where possible, we present a number of statistical models that best capture the salient features observed across our WLAN datasets. Such models are useful in many different situations. In Table 1, we summarize the high-level results and observations from our study.

The rest of this paper is organized as follows. Section 2 describes the WLAN environment for our study, and Section 3 the data collection methodology. Section 4 highlights the advantages of our wireless-side measurement methodology. Section 5 provides an overview of our traces. The subsequent six sections analyze the traces from different perspectives, including user (Section 6), application (Section 7), mobility (Section 8), session (Section 9), network (Section 10), and wireless viewpoints (Section 11). Section 12 discusses the implications and applications of our work. Section 13 relates our work to prior research efforts on WLAN traffic measurement. Finally, Section 14 concludes the paper.

2. Measurement environment

2.1. WLAN environment

The University of Calgary is a comprehensive, research-intensive university, whose campus consists of about 50 academic, administrative, residential, and athletic buildings. The university houses 60 departments and employs 2000 full-time faculty and staff. The university has 25,000 full-time students (undergraduate and graduate).

The Information Technologies (IT) division at our university offers *AirUC*, an IEEE 802.11 a/b/g non-encrypted infrastructure wireless network for use by students, staff, and faculty. Currently, AirUC consists of about 500 APs available in the main floors of all campus buildings. It allows for seamless roaming, where users can move from one building to another without disconnecting. Another 1000 APs are currently being added to expand coverage to all remaining floors in every building, interconnected walkways, and several outdoor areas.

2.2. WLAN configuration

Most of the APs in the AirUC network are Aruba 70 dual-band 802.11 a/b/g APs. The Aruba 70 is a dual-radio "thin" AP with built-in omni-directional high-gain tri-band antenna to support the 2.4 GHz and 5 GHz spectrums. Thin APs implement the minimal functionality required by the 802.11 standard. Upper-layer MAC processing functions are integrated into a central AP controller. The AP controller used on the campus is the Aruba 6000. Currently, the AirUC network has 6 AP controllers.

The AirUC network uses a 3 channel spectrum for the 'b/g' mode and an 11 channel allocation for the 'a' mode. The Aruba controllers handle channel allocation based on AP deployment. If the controller senses channel conflicts from another WLAN, then it may change the channels as needed. Almost all the APs operate on channels 1, 6, and 11, whose centre frequencies are sufficiently well-spaced for concurrent operation (i.e., non-overlapping channels in the frequency domain). The AirUC network has 14 different IP subnets, depending on the building. These IP subnets are allocated for AirUC client usage, and are separate from the corresponding wired subnets. All AirUC APs exist on a non-routable private network. All of the wireless subnets are connected to an Extreme Networks Summit X450 switch that acts as the core router for the AirUC traffic.

AirUC employs Web-based authentication. Once associated with an AP, a wireless device is dynamically assigned an IP address by a DHCP server. To utilize the AirUC network, the user then has to authenticate using an assigned username and password through the Aruba captive portal (once a Web browser is launched). Users are assigned IP addresses based on their "start" location, and retain these addresses wherever they roam, as long as they maintain a wireless connection.

3. Trace collection methodology

RF Monitoring (RFMON) is the most widely used technique for collecting traces of wireless activity [11,21,29]. This configuration places a wireless network interface card (NIC) into monitor mode, allowing the NIC to passively observe all nearby wireless traffic. NICs placed in RFMON mode can only sniff frames on a single channel. Furthermore, not all NIC chipsets and drivers support RFMON mode [12]. For those that do support RFMON mode, some chipsets may not function properly [12]. Some chipsets and/or operating systems may just support RFMON promiscuous mode, where only wireless data frames are captured. Also, not all drivers are supported on all operating systems (e.g., Windows, Mac OS). In almost all cases, those employing the RFMON design have used notebook computers with a wireless NIC, with a protocol analyzer (e.g., Ethereal, tcpdump) running to capture frames. This means that the placement points and operating range (if not using an external antenna) of the sniffer will be constrained.

To overcome the above-mentioned shortcomings, we use a specialized trace capture program called Airopeek NX [25]. Airopeek is a real-time 802.11 a/b/g WLAN analyzer used by network designers and administrators for performing site surveys, security audits, application-layer protocol identification, and troubleshooting. Airopeek works in conjunction with a network adapter (e.g., wireless NIC) to sniff frames from the air. In this work, we used off-the-shelf WLAN adapters called RFGrabbers [25]. Fig. 1 shows the trace collection infrastructure used for WLAN traffic measurement. The following subsections provide more details on Airopeek and RFGrabbers.

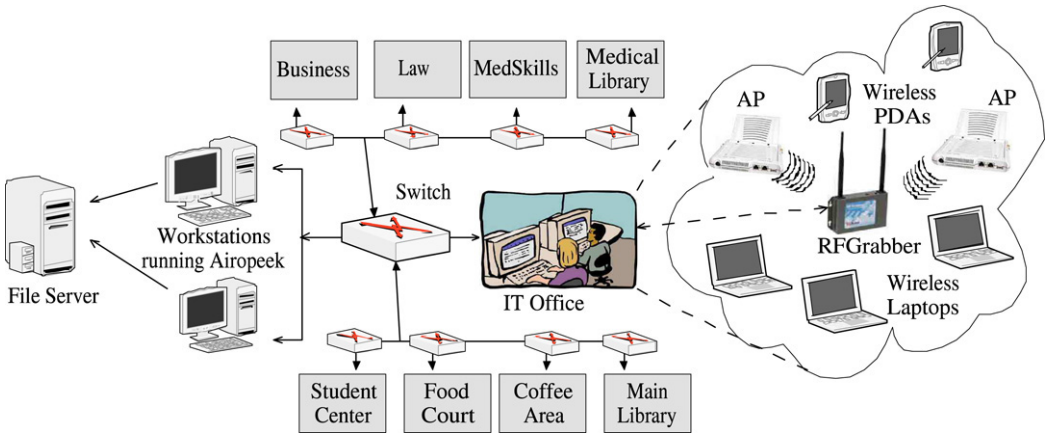


Fig. 1. WLAN trace collection infrastructure using Airopeek and RFGrabbers.

3.1. Airopeek

Airopeek can capture the MAC-layer and the higher-layer protocol headers of a packet. Airopeek has many features that make it attractive for WLAN data capture. First, Airopeek allows multiple simultaneous capture sessions, each using a different network adapter. Second, packet slicing can be used to capture only a portion of each packet instead of the whole packet. We set the slice size to 512 bytes allowing us to capture all headers and a portion of the packet payload. Third, Airopeek offers a wide selection of built-in filters that can be used to limit captures to packets that satisfy certain criteria. New filters can also be created. With packet slicing and filters, the size of the resulting trace capture files can be reduced significantly. Finally, triggers allow automatic management of the trace capture process. Users can set triggers to start and end a trace capture process based on the time of day, duration, trace size, or content bytes captured.

3.2. RFGrabber

The RFGrabber probe is an Ethernet-connected WLAN adapter that acts like a "listen-only" AP. With RFGrabber, one can capture 802.11a/b/g WLAN frames at a remote location and send copies of those frames encapsulated in UDP back to Airopeek running on any network-accessible computer. An RFGrabber probe connects to any existing Ethernet LAN and communicates with Airopeek using the Automatic device discovery protocol (Autocol), Simple Network Management Protocol (SNMP), and the Tazmen Sniffer Protocol (TZSP). RFGrabber probes do not suffer from vendor-specific AP loss or type-specific frame loss, which have been identified as issues in earlier wireless-side measurement studies [28].

Two versions of the RFGrabber probes were used in the study. The main difference between the two versions is how the firmware functions when the channel scanning feature is enabled in Airopeek. In version 1.1, the RFGrabber spends equal time listening on each selected channel. In version 2.0, the channel scanning order is random with the following properties [26]: two-thirds of the time the channel is chosen from the set of all selected channels, and one-third of the time the channel is chosen from the set of channels that have been identified by beacons. No channels are repeated until all channels in the set have been scanned. We configured the RFGrabbers to scan channels 1, 6, and 11 every 500 ms.

3.3. Locations

A guiding objective in our study was to capture wireless traffic from several locations, each representing a different user population. The involvement of the IT division was essential for the success of our study. Their staff's knowledge and experience were used regarding which locations were popular with WLAN users. In addition, they knew where the APs were installed. This information helped to determine where to deploy our 10 probes.

After consultation with IT, 9 locations in 7 buildings were identified. One sniffer was deployed at each location, and one was retained in the lab for testing purposes. The locations selected were:

(i) *Business*: This building houses the Faculty of Business.
(ii) *Coffee Area*: The lobby of the Engineering building has two coffee shops, a convenience store, vending machines, and a number of seating spaces.
(iii) *Food Court*: This is the largest Food Court on campus.
(iv) *IT Office*: This location houses the IT division.
(v) *Law*: This location is home to the Faculty of Law.
(vi) *Main Library*: This is the main library of the university, with many computer labs and study spaces.
(vii) *Medical Library:* The Faculty of Medicine has a library serving its students, as part of the Medical Centre, which is located about 2 km from the main campus.
(viii) *MedSkills*: MedSkills in the Medical Centre is for health sciences students and affiliated outside users.
(ix) *Student Centre*: This location houses the Student Union, convention facilities, and a restaurant.

IT technical staff deployed the probes on our behalf. We conducted a beacon scan for every sniffer, once deployed, and constructed a list of all AirUC APs visible to the probes. In total, 97 APs were identified, representing about 20% of the APs in the current AirUC.

The wireless traces were collected on two dedicated Dell OptiPlex GX270 2.8 GHz PCs. One PC had 3 GB memory and 80 GB disk, while the other had 2 GB memory and 60 GB disk. Both machines ran Windows XP, and were time synchronized using the Network Time Protocol (NTP). The captured traces files were regularly compressed and moved to a network file server.

3.4. Privacy issues

Ensuring the anonymity of AirUC users was a priority for us. The MAC address enables us to distinguish between users, but not identify a particular user. The IP addresses in the trace are temporary and cannot reveal the identity of individual users. Since user authentication takes place on a secure Web connection, the IT usernames are not decipherable in the trace. Finally, we do not search for any private information contained in the partial packet payloads that were captured, or allow anyone else to do so.

4. Advantages of wireless-side measurement

The two main enabling pieces of our wireless-side measurement methodology are RFGrabber probes and the Airopeek analyzer. The pragmatic advantages of our trace collection methodology over previous (wired and wireless) data collection techniques are many:

(i) *Remote passive capture*: RFGrabbers need not be physically co-located with the trace collection workstation. Multiple probes can be placed near APs in the WLAN. The probes connect to the existing Ethernet LAN and stream captured wireless packets to a workstation elsewhere on the network. The monitoring approach is passive from the viewpoint of the WLAN.
(ii) *Special-purpose measurement devices*: RFGrabber probes are compact devices that can be deployed easily at or near the same locations as APs. This approach provides an AP-centric view of the network, rather than a client-centric view as obtained with a wireless laptop in promiscuous mode. Furthermore, the probes can have a larger antenna than typical laptop NICs (providing a better operating range), and have the ability to scan the multiple physical channels used on multi-channel WLANs.
(iii) *Complete WLAN view*: RFGrabber probes can capture all frame types, including Data, Management, and Control frames. Because Management and Control frames do not enter the wired Distribution System (DS), these frames cannot be captured from the wired side.
(iv) *Link-layer headers*: RFGrabber probes can capture the full wireless frame, including LLC and MAC-layer information. This header information includes MAC addresses, frame directionality, data transmission rate, channel number, signal strength, and retransmission flags. Obtaining MAC/PHY information from the wired side of the network is difficult [29].
(v) *Multi-layer analysis*: Wireless-side traces provide all information required for multi-layer protocol analysis up to the application layer. No additional data is required. Previous studies [6,7,18,24] have collected syslog or authentication data for the entire WLAN, with wired traces collected for a portion of the network. Typically, syslog or authentication logs were used to identify the MAC addresses for each user's NIC, which were then used to extract the wireless traffic from the traces.

Table 2
Volume of AirUC traffic captured during the trace period

Channel	Data (GB)	Management (GB)
1	50.40	18.00
6	28.21	16.13
11	39.36	21.10
Total	117.97	55.23

(vi) *Intra-session activity*: Syslogs can be used for determining user sessions. The syslog records association, authentication, roaming, and disassociation events for wireless clients, and authentication logs record login and logout information. However, sessions calculated using syslog or authentication data do not provide insight into the activity level of a session. Wireless traces collected from the probes can provide this information.

As with most other wireless adapters, the probes cannot guarantee that all packets in the air are captured. The obvious reasons are physical and environmental limitations that affect antenna gain, operating range, signal quality, traffic intensity, and wireless interference. RFGrabbers have an indoor operating range of up to 80 meters [26]. Our tests indicate that these devices see 95%–99% of the traffic transiting a nearby AP. The methodology for measurement loss estimation is described next.

All data and management frames (except retransmissions) sent by an 802.11 wireless device can be distinguished by a sequence number in the MAC header. Every time a wireless station or an AP sends out a data or management frame, the sequence number counter is incremented by 1. Sequence numbers can have any value in the *mod* 4095 set. Once the maximum value is reached, the counter wraps. By counting the gaps in the sequence numbers of frames captured by a sensor, the number of missed data and management frames can be estimated.

5. Trace data overview

Our trace collection began on Friday March 3, 2006 at 12:00 AM and ended on Friday April 14, 2006 at 11:59 PM. As a result of network and power outages, several gaps occurred in the trace data. During the night of Friday March 17, a power surge resulted in the interruption of trace collection on 4 probes for 4 days. For consistency across traces, we ignored all data collected between March 18 and March 21. The second gap occurred on Wednesday March 29. A failure occurred in the automated aggregation of trace data to the file server, resulting in Airopeek being unable to save new data until the aggregation process was re-established. Thus, we ignored all data collected on March 29. After eliminating these gaps, the resulting trace length was 38 days. We used this trace data to analyze user, application, mobility, session, network, and wireless traffic characteristics on AirUC.

5.1. Trace characteristics

The trace captured 933,182,676 frames, of which 64.02% were Management frames, 35.97% were Data frames, and 0.04% were Control frames. Note that the design of our filters restricted the number of Control frames captured.

Table 2 summarizes the volume of AirUC traffic captured for the entire trace period. Beacons accounted for 99.7% of all Management frame bytes. Other Management frame subtypes observed in the trace were Authentication (0.025%), Association Request (0.031%), Association Response (0.033%), Disassociation (0.023%), Reassociation Request (0.003%), and Resassociation Response (0.003%). Our analyses focus on the Data frames. Data frames for which the CRC checksum failed were ignored. These accounted for approximately 11% of the total Data frame count.

Table 3 provides a general summary of the trace characteristics on a per-location basis, showing the number of APs and users observed at each location, as well as the session activity, peak hourly traffic rates and IP traffic volume. Further details on these characteristics follow in the remainder of the paper.

5.2. User classification

AirUC users were distinguished based on the MAC addresses of their NICs. Consistent with prior work [2,6], we assumed that each MAC address represented a unique user. We looked at the Address fields of all Data frames in

Table 3
Summary of trace characteristics

General category	Specific location	Num APs	Num users	Sessions per day	Peak IP traffic		IP Traffic Volume (GB)			
					pkt/s	Mbps	In	Out	Local	Total
Academic	Business	23	1215	307	160.8	0.8	9.1	5.9	2.6	17.5
	Law	18	1430	202	152.5	0.6	10.1	7.0	2.5	19.6
	MedSkills	6	193	33	5.9	0.0	0.2	0.0	0.1	0.3
Library	Main	4	2192	266	52.4	0.3	5.3	1.1	1.9	8.3
	Medical	14	456	95	73.0	0.3	1.6	0.6	1.2	3.4
Service	IT office	9	302	60	25.7	0.2	0.5	0.1	0.6	1.1
Social	Coffee area	8	1257	161	163.3	0.7	9.7	2.6	2.6	14.9
	Food Court	3	1645	156	84.3	0.3	6.2	4.2	1.8	12.2
	Student Centre	12	1958	273	161.7	0.8	15.0	5.4	4.3	24.7
WLAN	Total or Avg	97	6775	1481	444.0	1.9	57.6	27.0	17.3	102.0

the trace. These frames can be divided into two[1] groups: frames sent to the AP (To-DS = 1 and From-DS = 0), and frames sent from the AP (To-DS = 0 and From-DS = 1). For Data frames sent to each AP, all unique MAC addresses seen in the Address 2 field (transmitter station) were added to the list of users. Similarly, for Data frames sent from each AP, all unique MAC addresses seen in the Address 1 field (receiver station) were added to the list.

5.3. User sessions

Users generate sessions. Sessions occur when a user joins the WLAN, uses it for a certain time, and then leaves the network. The session duration is defined as the time spent between joining and leaving.

A session starts when the user's NIC sends an Authentication frame to the AP. After getting a positive response from the AP, the user NIC sends an Association Request frame. If the AP allows the user NIC to associate, the AP replies with an Association Response frame. The user NIC is now associated with the AP. Next, DHCP boot request and boot response packets are exchanged between the user station and a DHCP server. The DHCP server is part of the DS and hence all wireless packets must pass through the AP. After being assigned an IP address, the user can use the WLAN. When the user decides not to use the network any more, the NIC may send a Disassociation frame to the AP. Usually, NICs do not send out such frames when the stations are shut down. However, within minutes, the AP sends a series of Deauthentication frames indicating the end of a session.

In some cases, the RFGrabber may not capture all the packets required to determine the start and end of user sessions. In the absence of Authentication and Association frames, we started a new session whenever a packet from a new user was noticed in the trace. Similarly, in the absence of Disassociation frames, we chose an arbitrary session idle timeout of 30 min to differentiate between two sessions of the same user. The end time for the session was set to the time of the last packet seen. All active sessions that persisted beyond the end of the trace were ignored.

5.4. Roaming

Users may roam from one AP to another. Roaming can be identified by looking for Reassociation Request and Reassociation Response frames between the user NIC and the new AP. Again, the trace did not necessarily record all such frames. Hence, we maintained two state variables for each unique user: number of roams, and current AP (represented by its BSSID). Each time a user NIC used a new AP to send or receive packets, the roams counter was incremented, and the current AP was updated. When the user session ended, the roams counter was saved to indicate the number of roams during the session. Also, we used the 30 min session timeout for roaming users. Thus, if a user associated with a new AP within 30 min, it was assumed that the user had only roamed and not started a new session.

[1] In our trace, we did not observe any frames sent in ad hoc mode (To − DS = 0 and From-DS = 0), which completely bypass the wired Distribution System (DS).

5.5. Terminology

The following terminology is used in this paper:

- *WLAN*: Traces from the 9 locations were aggregated based on the timestamp of each packet. The aggregate trace was analyzed to understand the overall usage, roaming patterns, and traffic characteristics on AirUC. In the rest of this paper a reference to "WLAN" means the 9 locations collectively.
- *AP set*: To study user behaviour, an AP set is maintained for each user. This set contains the BSSIDs of unique APs with which the user NIC has associated, either during a session, an hour, or a day.
- *Intra-location and Inter-location roams*: Mobile users generate roaming events. A roaming event occurs when the user NIC associates with a new AP. When mobile users move between any two APs in the same location during a time period, this event is called an intra-location roam. An inter-location roam happens when a mobile user moves between APs situated in different locations.
- *Stationary and mobile users*: To designate a user as stationary or mobile during a time period, we looked at the AP set of the user. If all APs belonged to the same location and the cardinality of the AP set was less than 3, then the user was deemed stationary. Otherwise, the user was deemed mobile. These heuristics are intended to detect physical movement of users, rather than the transient roaming of the NIC itself. Because we did not have the exact location of the APs in our trace we were unable to use any heuristics that involved using location of APs.
- *Stationary and mobile sessions*: Sessions can be stationary or mobile. If a user was mobile between the start time and the end time of the session, the session was considered to be mobile; otherwise, the session was stationary.
- *AP association threshold*: This is the minimum time the user NIC should be associated with an AP, for the movement of the user NIC to that AP to be counted. We chose an AP association threshold of 2 min, which allowed us to remove excessive roams caused by aggressive NICs [13,22].
- *Incoming and outgoing traffic*: Incoming traffic refers to data being "downloaded" from the Internet, while outgoing traffic means users were "uploading" data.
- *Local traffic*: Any IP packet sent/received by a user to/from a host on the wired-side of the university network was considered part of the local (internal) traffic. Examples of such hosts are the university Web servers, DHCP servers, and mail servers.

6. User view

We begin our traffic analysis with a user-oriented view of the WLAN usage. We identified 6775 unique users in the trace, among which 805 users did not associate with any single AP for more than the AP association threshold. We conjecture that such users were on the periphery of the range of the RFGrabbers, so few packets sent or received by the users were captured. It is also possible that some of the users were denied access to the WLAN because of a failed authentication. Unless otherwise stated, the analysis presented in this paper is based on the traffic sent/received by the 5970 users that satisfied the AP association threshold.

6.1. User devices

Most laptops and handheld devices these days have built-in wireless NICs. For those using an older model laptop, buying a stand-alone wireless NIC is an inexpensive solution.

We used MAC addresses to classify the wireless NIC cards by vendor. A MAC address is 6 bytes long with the first 3 bytes representing the vendor ID and the last 3 bytes representing the card ID. We used the vendor ID list from IEEE to classify the 6775 user NICs [8]. We identified 96 unique vendors. We found that approximately half of the users have devices with built-in wireless NICs, such as Intel (38.4%), Apple (12.5%), and IBM (2.9%). The remaining users use stand-alone wireless NICs (e.g., GemTek (12.9%), Askey (7.3%), D-Link (3.5%), and Linksys (2.9%)) in their devices to access AirUC.

We used the pOf tool [20] to determine the operating systems of the user devices. The pOf tool exploits the differences in the TCP/IP stacks of operating systems to create a fingerprint. For pOf to determine the operating system of a device, the device must send at least one TCP SYN packet. In our trace, we captured TCP SYN packets for approximately 73% of the 6775 user NICs. We applied the pOf tool on all TCP SYN packets of these user NICs. If the tool consistently indicated the same operating system for all TCP SYN packets of a NIC, we assigned that operating

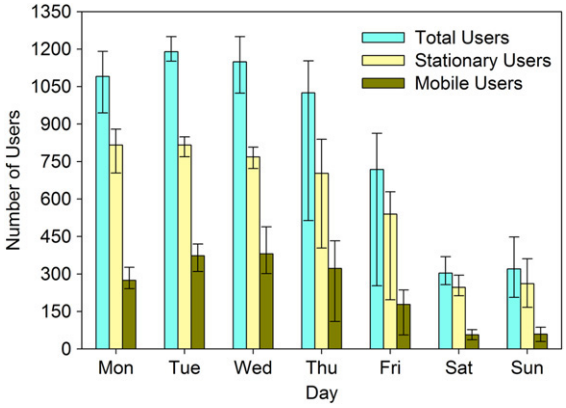


Fig. 2. Number of users seen daily in the WLAN.

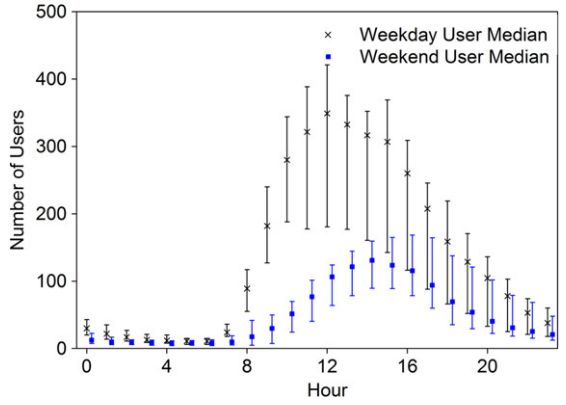


Fig. 3. Median number of users seen per hour.

system to it. If pOf returned two different operating systems that can run on the same CPU (e.g., Windows and Linux), a NIC was identified as a dual-boot system. On the other hand, if the operating systems were not architecturally compatible (e.g., Mac and Windows), then the NIC was left unidentified [6].

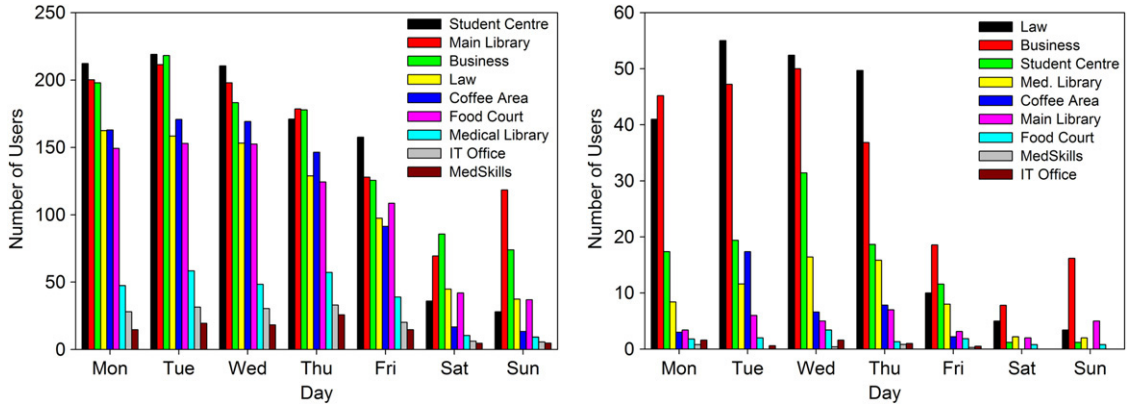
We found that over 60% of the users have laptops running Windows XP SP2 or Windows 2000. Macintosh devices (12%) are also popular among users. About 1% of the users ran Palm OS or Pocket PC. This may indicate that handheld wireless devices are not prevalent among users or that users are not utilizing the wireless capabilities of such devices (which are not as well suited to common Internet activities such as Web browsing). Other operating systems identified were FreeBSD (0.27%), Linux (0.19%), Dual-boot Windows/Linux (0.13%), and OpenBSD (0.03%).

6.2. *WLAN usage*

We were interested in understanding how AirUC was used. Specifically, we wanted to know how many users connected to the WLAN on a daily and hourly basis, whether they were stationary or mobile, and how usage differed between weekdays and weekends (or days and nights).

Fig. 2 shows the minimum, average, and maximum number of users seen during the days of the week. The graph shows the expected weekly cycle. On average, more users (1200) used the WLAN on Tuesday than on any other day of the week. User counts tapered toward the weekend. On each day, 20%–30% of the observed users are mobile.

Fig. 3 shows the median number of users seen per hour in the WLAN, as well as the 10th and 90th percentiles. Each x-axis data point in the graph represents a one hour time interval. For example, the tick labelled 8 represents the time interval 7–8 AM. A classic diurnal pattern is seen for weekdays. The user count increases steadily during the morning hours, and peaks near 400 around noon. The users tend to use the WLAN in high but decreasing numbers during the afternoon. The number of users gradually decreases until midnight, after which only a few users stay connected.



(a) Stationary users.

(b) Mobile users.

Fig. 4. Average number of stationary or mobile users at each location per day.

Similar patterns have been reported in previous studies (e.g., [5,13]), and are expected patterns of human behaviour, which is location independent. For weekend days, the diurnal pattern is still present, but less pronounced. On weekends, fewer users are on campus, and those who do come tend to arrive on campus later. The user count peaks in the early afternoon (2 PM), and gradually decreases as the day ends. The diurnal patterns observed here are quite consistent across all the 9 locations studied. The Main Library location differed slightly: activity persisted into the late evening, because of extended hours during the final exam period.

6.3. Usage at locations

We wanted to contrast user activity at the various locations. Also, we wanted to see how user behaviour at a location deviated from that of an average user of the WLAN. Specifically, we were interested in knowing which locations had the most users, and how usage varied at each location by the day or the hour and why.

Fig. 4 shows the average number of stationary users and mobile users at each location on the campus. We observe that on average all social areas and two academic areas, namely, Business and Law, boast high number of users during weekdays. Also, most users remain stationary and usage at each location follows the weekly work cycle. There are more mobile users in locations with higher number of APs.

The reason for some locations witnessing more users than other locations can be explained as follows:

- Student Centre and Food Court are the two most popular locations on the campus, where students come to socialize and eat at the Food Court. Coffee Area has open seating spaces and attracts a number of students during the day.
- Approximately 3000 students are enrolled in some business programs at our university. The Business location has large open spaces and sufficient seating areas for students to use their laptop to access the AirUC network. The computer labs in Business become crowded during peak hours. Therefore, the Faculty of Business recommends its students to purchase laptops (with wireless capability) for academic purposes.
- Main Library has approximately 4500 visitors per day. Also, the library has spaces where students can study privately or jointly. Notice that Main Library has the most users on Sundays. With midterms in March, the library seems an obvious choice for students to study during weekends.
- Law has the next highest number of APs (after Business) among all the locations. The location also has spaces for students where they can study or relax.
- Medical Library and MedSkills are unique because they are located on the South campus of the university. MedSkills is an academic area where Medicine students interact with approximately 80–100 volunteer patients and 100 professional actors who enact symptoms. The area also has a couple of classrooms. Medical Library is the dedicated library for Faculty of Medicine students. Thus, there is not a wide mix of students at this location leading to fewer users (mostly Medicine students).
- Users at IT Office are mostly composed of IT employees. The IT Office is located on the topmost floor of the Math Sciences building. It is rare for students to go there and access the AirUC WLAN, since there are no seating areas for students at this location.

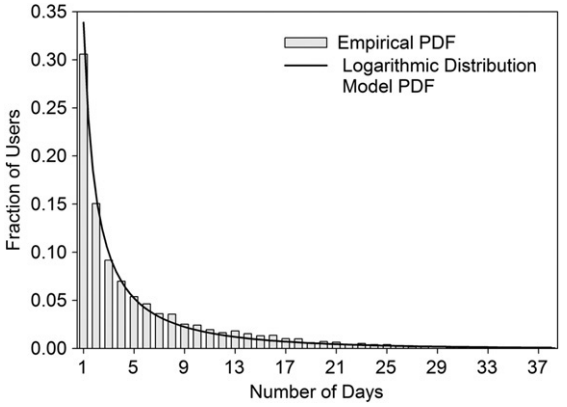


Fig. 5. Number of days users access the WLAN.

6.4. Usage regularity

We were interested in knowing how regularly users accessed the WLAN, to understand their reliance on the WLAN, and how often they use it in their daily academic life.

Fig. 5 shows the distribution for how many different days users accessed the WLAN during our trace. Of the 5970 total users, 30% used the AirUC network only one day during the trace period. The remaining "repeat users" (70%) showed widely varying activity: the median is 3 days, but 20% connected on 10 or more days. Three users connected to the AirUC network every day during the 38-day trace period. These all connected from the Student Centre.

The regularity of usage shows some location-specific differences. For example, 42% of the users at the Business location connected to the wireless network for 4 days or more. This difference can be attributed to more Business students having laptops, since the Faculty of Business encourages its students to do so. The IT Office was the least popular location among users; approximately 63% of the users at this location connected to the network only once. However, there was one user, probably an IT employee, who connected on 35 different days. The overall usage pattern in Fig. 5 follows a Logarithmic distribution (PDF: $f(x) = \frac{-\theta^x}{x \ln(1-\theta)}$) with a shape parameter, $\theta = 0.94$. The fitted distribution passed the Kolmogorov–Smirnov (K–S) goodness-of-fit test at all significance levels tested ($\alpha = 0.10$, 0.05, and 0.01).

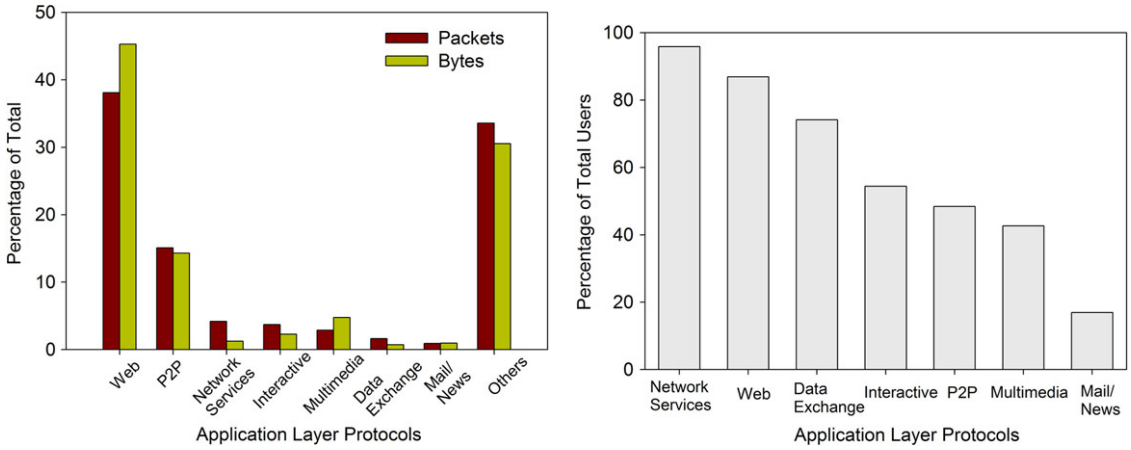
7. Application view

The next set of results provides an application-level view of what users were doing on the WLAN. These activities influence the network traffic workload, and are relevant to capacity planning of future WLANs. Our trace recorded about 103 GB of IP traffic attributable to users. Approximately 96% of the traffic was TCP. UDP traffic accounted for about 3%, while Internet Control Message Protocol (ICMP) constituted less than 1% of the traffic.

7.1. Application-layer protocols

We wanted to understand what applications were used while users were connected to AirUC. For this purpose, we created 8 application groups based on the protocols used:

(i) *Data Exchange*: This group includes file transfer applications (e.g., FTP), database applications (e.g., SQL), and network file systems (e.g., SMB, NetBIOS, NFS).
(ii) *Interactive*: This category includes chat applications (e.g., GTalk, IRC, MSN), session applications (e.g., SSH, telnet), remote desktop applications (e.g., GoToMyPC, RDP), and collaborative applications (e.g., Cu-See Me, Groupwise).
(iii) *Mail/News*: Email (e.g., IMAP, POP, SMTP) and newsgroup (e.g., NNTP) applications are included.
(iv) *Multimedia*: Streaming media (e.g., QuickTime, RTSP), audio/video (e.g., Windows Media Player), and VoIP (e.g., SIP) protocols constitute the Multimedia category.
(v) *Network services*: Bookkeeping protocols are grouped here (e.g., DHCP, DNS, Kerberos).



(a) Traffic count and volume.

(b) Users.

Fig. 6. Application-layer protocols by category.

(vi) *P2P*: This group includes peer-to-peer file sharing applications (e.g., Gnutella, BitTorrent, KaZaA).
(vii) *Web*: HTTP and HTTPS are included here.
(viii) *Others*: All unidentified protocols are in this category.

We used a simple port-number approach for an initial traffic classification. For each category, we built a comprehensive list of well known (0–1023 range) and commonly used (>1024 range) port numbers for application-layer protocols that run on either TCP or UDP [9].

Fig. 6(a) shows that about 46% of the user byte traffic was from Web surfing. About 15 GB (15%) of user traffic was from known P2P applications. These findings are consistent with other campus measurement studies, where P2P traffic was significant [6]. At our university, there is a traffic shaper in place for known P2P traffic. The traffic shaper does not prevent P2P traffic, but severely limits its bandwidth consumption.

About 31 GB (31%) of user traffic is in the Others category. This observation is common in campus WLAN studies [6,22]. In [22], the authors were unable to identify 35% of the wireless packets. Similarly, the authors in [6, 13] were not able to identify a substantial portion of the traffic. Much of our Others traffic was TCP (93%), with the remaining 7% being UDP. The applications generating such traffic often use random or unassigned port numbers, for which port-based classification fails.

To examine the Others issue in more detail, we collected a separate 1 h trace[2] of TCP traffic with full packet payloads on the morning of April 6. Traffic classification was done using signature matching in Bro [4]. The results of this analysis show that HTTP (55.35%) and P2P (32.45%) account for most of the byte traffic volume, followed by Others (6.58%), Mail (2.23%), and Interactive (1.38%). Since only TCP traffic was collected, Network Services contribute negligibly to the total. These results suggest that a majority of the "Others" traffic in our study is actually P2P traffic, as we suspected.

Fig. 6(b) shows application usage by users. We associated a user with a particular application category if we saw at least one packet attributable to protocols in that category. Because users are connecting to the wireless network, network services protocols (e.g., DHCP) are used by over 95% of users. A high percentage of users (86%) were involved in Web browsing activities. Approximately 75% of the users used some data file system application to access files remotely. Only 16% of the users used some email client to check messages, suggesting that users prefer Web-based email services to regular email clients. Almost half of the user population used some P2P application, and 40% used media streaming applications. These trends suggest growing WLAN traffic demands in the near future.

7.2. Traffic directionality

Fig. 7 shows the distribution of incoming, outgoing, and internal traffic for the WLAN as a percentage of the total traffic per application category. This analysis reveals distinctive profiles for different network applications. For

[2] For this trace only we utilized `tcpdump`, as Bro requires traces in `pcap` format.

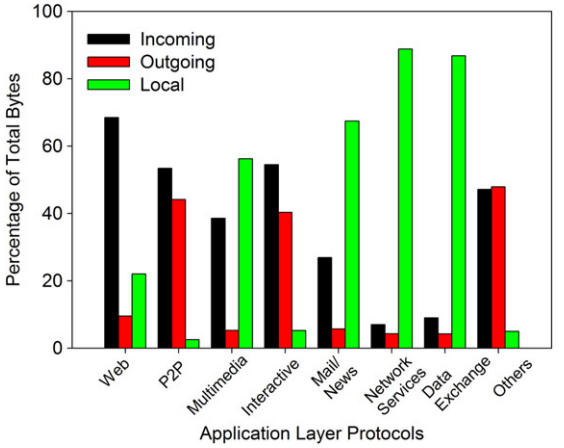


Fig. 7. Traffic directionality by protocol category.

example, about 90% of the network services were local (internal) to the campus network, for obvious reasons. A similar profile is seen for data file system usage, indicating that users were primarily accessing content from university file servers. Email applications showed 65% internal traffic, and 30% incoming traffic. Web browsing showed the opposite trend: users surfed off-campus Web sites (70%) more than local university sites (20%).

The balance between incoming and outgoing traffic volume for P2P traffic is not surprising, since newer P2P applications, such as BitTorrent, make the download rate proportional to the upload rate. Symmetry is also observed in the Others category, which has characteristics qualitatively similar to P2P.

Only 0.3 GB of P2P traffic was internal. This low value suggests that P2P applications do not exploit local network topology well, or that users have such diverse interests that local file sharing is rare. The growing volume of upload traffic has possible performance implications for wireless access networks, because of uploader/downloader fairness issues in WiFi networks [19].

Approximately 3 GB of Multimedia traffic was internal. Because DAAP was the dominant protocol in this category, we believe that users share iTunes among themselves. It could also be possible that users accessed audio/video files available on the university Web site.

7.3. Popular web sites

Considering that about 46% of the traffic was for Web surfing, we were interested in finding similarities in the browsing habits of users. In Fig. 8, we show the geographic distribution of all external IP addresses.[3] We identified 3,821,929 unique external IP addresses in the trace. We used the MaxMind GeoLite country database to map an IP address to a country [15]. MaxMind claims that the database is 97% accurate. The database is created by querying major Internet registries such as American Registry for Internet Numbers (ARIN), Réseaux IP Européens (RIPE), doing reverse DNS lookups, and conducting surveys of Web users about their location. Our tests done with a small set of IP addresses with known locations returned believable results. It is interesting to note that only 7% of the IP addresses seen were from Canada.

To understand what Web sites were popular among users, we did reverse DNS lookups for some of the most frequently seen IP addresses. We found that approximately 42% of 6775 users had visited the university Web site. Searching the Web was also a common activity among users. Approximately 43% of the users visited google.com. Hotmail.com was visited by approximately 37% of the user population. Other popular websites included tsn.ca (Canadian sports channel Web site), blackboard.ucalgary.ca (university course management system), aol.com, and msn.com.

[3] Any IP address that did not belong to the university IP address space was considered external.

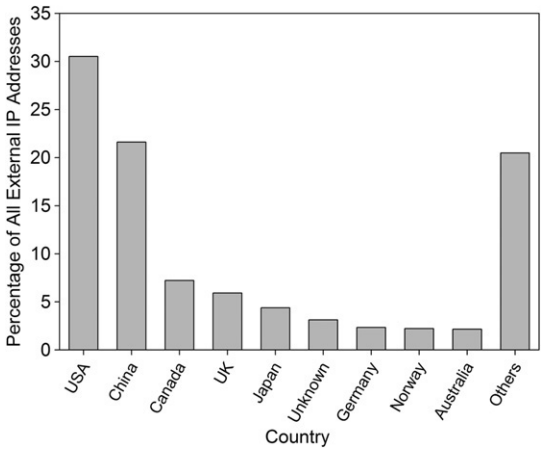


Fig. 8. Geographic location of external IP addresses.

8. Mobility view

One of the primary benefits of wireless Internet access is that it enables user mobility. Given our knowledge of the user population and application usage, we were interested in knowing how users were leveraging the flexibility offered by a campus-wide WLAN. Specifically, we wanted to know how many AirUC users moved, how they moved, where they moved, and how often they moved.

8.1. APs and locations visited

Fig. 9(a) shows user mobility with location granularity. About 54% of the users were seen at only a single location during the trace. The most locations visited was 7. Fig. 9(b) shows per-location analysis for the number of APs visited. The visit behaviour differs slightly across locations, since it is influenced by the number of APs available. The median number of APs visited by users at the Business location and the Medical Library was 2. At all other locations, 50%–75% of the users visited only a single AP.

Fig. 9(c) shows the overall distribution of how many APs were visited by each user. About 30% of the users were seen at only one AP. The median number of APs visited was 2, and the maximum was 29. About 90% of the users visited fewer than 8 APs. Few users were highly mobile; nonetheless, the distribution does have a pronounced tail. We found that the distribution of the number of APs visited in the aggregate trace is well-approximated with a geometric distribution (PDF: $f(x) = p(1-p)^{x-1}$), with a shape parameter, $p = 0.27$. The fitted distribution passed the K–S test for $\alpha = 0.01, 0.05,$ and 0.1. Similar models (with different geometric decay rates) apply for most individual locations as well.

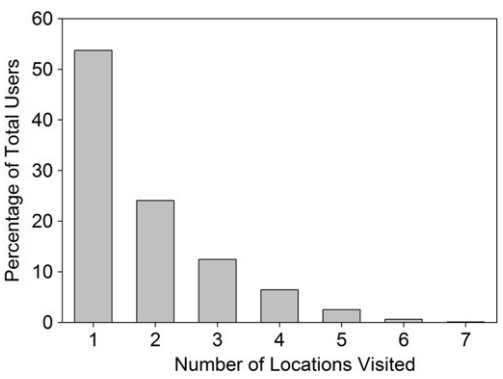
8.2. Roaming events

Fig. 10 shows mobility characteristics by time of day. During an hour, more users moved within a location than between locations. Most movements between locations happened during normal working hours. Inter-location roaming events peaked during lunch hour, indicating movement to and from the Food Court and the Student Centre and other locations on the campus. Such movements accounted for 40% of all inter-location roaming events.

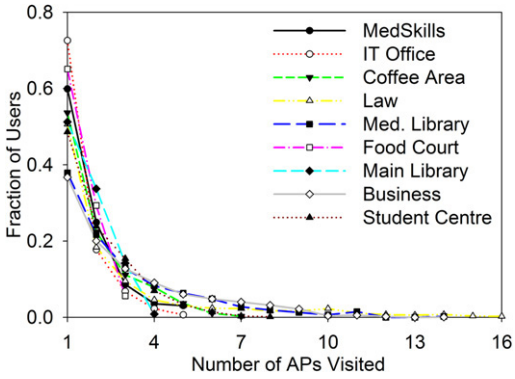
8.3. Mobility patterns

To better understand user mobility trends, we did a pairwise analysis of the number of observed users in common between any two locations. Fig. 11 shows the results of this analysis, using the bar heights to indicate the number of users in common between two locations. For example, the rightmost vertical bar indicates that 458 users from the Business location also visited the Main Library.

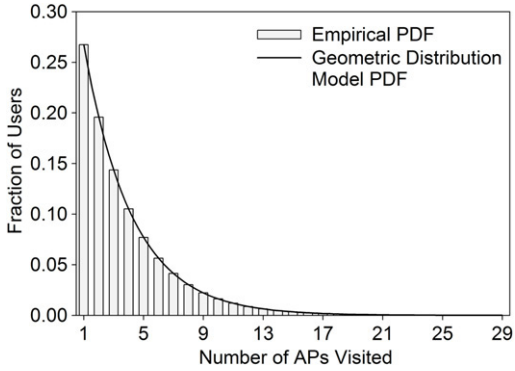
Fig. 11 can be further explained using the Medical Library as an example. The short vertical bars in the Medical Library column represent the number of users common between the Medical Library and all remaining 8 locations,



(a) Locations visited.



(b) APs visited per location.



(c) APs visited overall.

Fig. 9. Locations and APs visited by users.

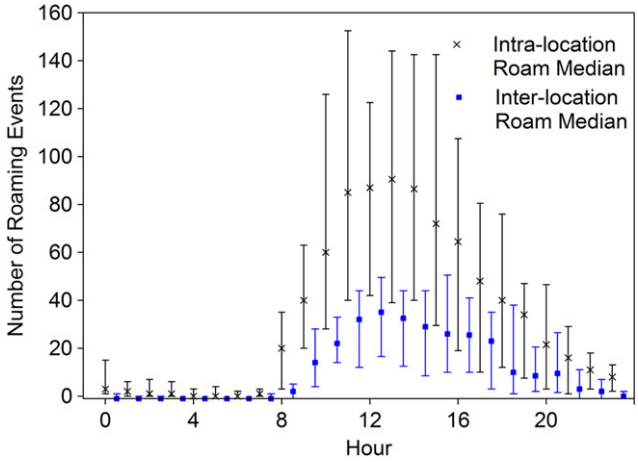


Fig. 10. Number of roaming events per hour.

considered one at a time. We found that the Medical Library had 149 users in common with MedSkills during the trace period. The fact that the Medical Library and MedSkills are distant (2 km) from the rest of the campus is apparent in the graph. In total, only 70 users from the two Medical Centre sites were observed using the WLAN at other campus locations.

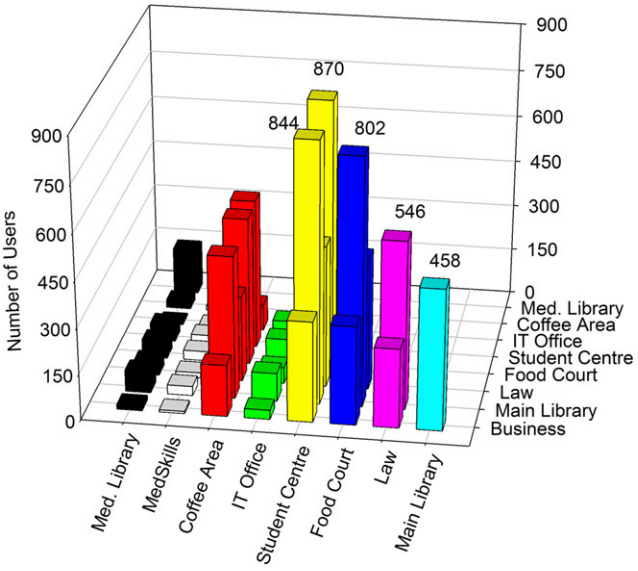


Fig. 11. Number of users common among any two locations.

From Fig. 11 we also observe that many users are common between the Student Centre, Food Court, Law, and Main Library, considered pairwise. These results reflect the popularity of these locations with users. Even with few APs, locations such as the Food Court and the Main Library attract many users. It seems that users prefer a known location that they like, as opposed to an area that has better wireless connectivity. Expansion of the AirUC network (currently underway) will certainly help such users.

The user mobility patterns observed are also influenced by geographic proximity. For example, we found 546 users in common between the Main Library and Law. These two buildings are situated next to each other. Expanding AirUC to walkways and outdoor areas between neighbouring locations could benefit these users.

9. User session view

Understanding user sessions is important for network planning. Sessions help us understand connection patterns, the amount of time users are active, and the traffic generated. These results can be used to develop synthetic workload generators for performance evaluation of the WLAN (e.g., Surge for performance evaluation of Web sessions [3]).

9.1. Sessions per user

The notion of sessions provides a deeper understanding of user activity. Analysis of sessions can tell us how often users connect to the WLAN, how long they remain continuously connected, and how much data is transferred. Users in the WLAN generated 47,945 sessions, of which 5040 (approximately 11%) were mobile sessions.

Fig. 12 shows how many sessions were started by each user during the trace period (the embedded graph zooms in on the initial portion of the CDF). These results highlight the prevalence of one-time users (27%), as well as repeat users. An average user started 9 sessions, while the median was 4. There is a pronounced tail to the distribution, showing the presence of a small number of "heavy" users. The maximum number of sessions generated by a user was 494. A Logarithmic distribution with $\theta = 0.97$ fits the session behaviour well. The fit passed the K–S test with a test statistic equal to 0.03. The graphical fit is shown using a CDF in Fig. 12.

We wanted to understand which locations were preferred by users when starting a session. To counter the influence of larger locations, we normalized the total count of session starts at each location by the number of APs. We found that 68% of all sessions were started in the Main Library, Food Court, or the Student Centre. On average, 633 sessions were started per AP in these locations during the trace period.

We also studied session inter-arrival times for the aggregate traffic from all users. The median session inter-arrival time was 20 s. About 90% of all sessions were started within 2 min of the previous session start. The maximum session

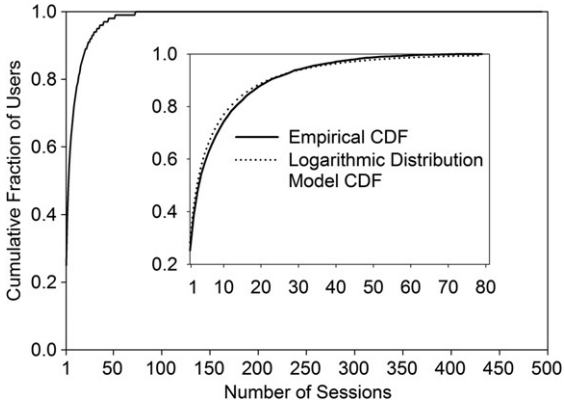


Fig. 12. Number of sessions started per user.

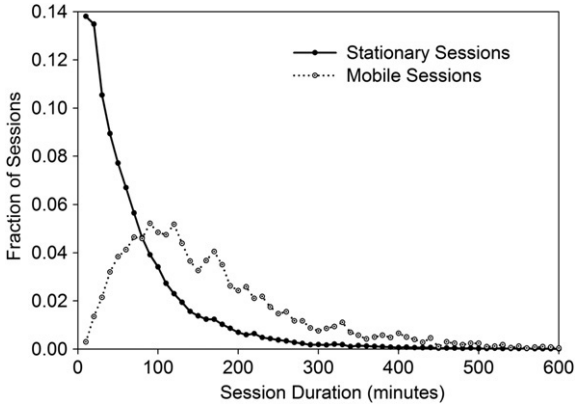


Fig. 13. Session durations.

inter-arrival time observed was about 4 h. We also looked at the hourly trends for session inter-arrival times. The median inter-arrival times were time-varying, and followed the expected diurnal patterns. Session inter-arrival times can be modelled using the two-parameter Lognormal distribution (PDF: $f(x) = \frac{e^{\frac{-(\ln x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma x}$) with shape parameters, $\sigma = 1.4$ and $\mu = 3.01$. The K–S statistic for the fitted distribution was 0.02, which was lower than the critical values at $\alpha = 0.01, 0.05$, and 0.1.

9.2. Session duration

Fig. 13 shows the distribution of durations for all sessions. The median session duration was approximately 50 min, and the mean session duration was about 1.4 h. About 90% of the sessions ended within 3 h. Mobile sessions tend to last longer than stationary sessions. The median duration for stationary sessions was 44 min, while the median for mobile sessions was 2 h. About 90% of all mobile sessions ended within 6 h.

Stationary session durations follow a two-parameter Weibull distribution (PDF: $f(x) = \frac{\alpha x^{\alpha-1}}{\beta^\alpha} e^{-\frac{x^\alpha}{\beta}}$) with shape parameters, $\alpha = 0.93$ and $\beta = 66.94$. The K–S statistic for the fitted distribution was 0.03, which was lower than the critical values at $\alpha = 0.01, 0.05$, and 0.1. Mobile session durations can be modelled using the Inverse Gaussian distribution (PDF: $f(x) = \sqrt{\frac{\lambda}{2\pi x^3}} e^{\frac{-\lambda(x-\mu)^2}{2\mu^2 x}}$) with shape parameters, $\lambda = 346.83$ and $\mu = 157.39$. The fitted distribution had a K–S statistic of 0.01, which was lower than the critical values at $\alpha = 0.01, 0.05$, and 0.1.

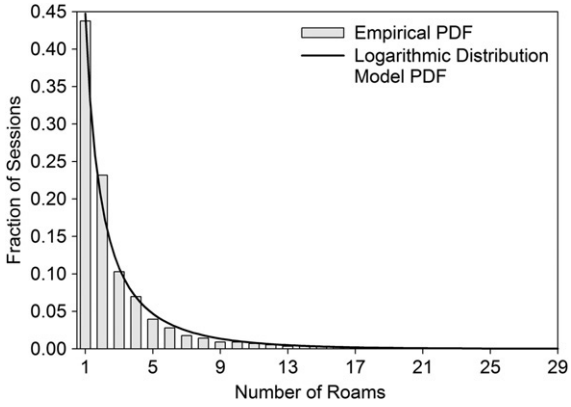


Fig. 14. Number of roams per mobile session.

Table 4
Average session weights

Session duration (d) (h)	Average session weight (MB)	
	Stationary	Mobile
$0 < d \leq 1$	0.55	1.16
$1 < d \leq 2$	2.03	3.15
$2 < d \leq 3$	3.96	5.03
$3 < d \leq 4$	5.83	9.40
$4 < d \leq 5$	6.03	10.91
$5 < d \leq 6$	11.25	13.21

9.3. Session mobility

Fig. 14 shows the distribution of the number of roams during a (mobile) session. About 40% of all mobile sessions had a single roam. Only 10% of the sessions had more than 6 roams. The most roams observed per session was 30. The number of roams per session is logarithmically distributed with the shape parameter, $\theta = 0.85$. The K–S statistic for the fitted distribution was 0.03 which was lower than the critical values at $\alpha = 0.01, 0.05$, and 0.1.

Since mobile sessions tend to be longer, the number of times a user associates with an AP can be high. We observed one user NIC that roamed 87 times during a session. Usually such excessive roams are caused by anomalies in the NIC firmware, and/or the user sitting near multiple APs [13,22].

9.4. Session activity

To better understand the relation between session duration and how active users are during the session, we use a measure called session weight [1]. Session weight is the total volume of user traffic transferred in a session.

Table 4 lists the average weights for sessions with durations up to 6 h, classified using 1 h bins. Note that 90% of the sessions end within 6 h. The average weights of mobile sessions are higher than those of stationary sessions for three reasons. Mobile sessions on average are longer than stationary sessions. Mobile sessions remain less idle than stationary sessions. Finally, when users move from one AP to another they generate network management traffic that does not occur when the sessions are stationary.

The median session duration for the top 10% of the heavy sessions was 2.2 h, while for the remaining 90% of the sessions the median was 44 min. In general, heavy sessions were long, but long sessions are not necessarily heavy. The amount of transferred data was low for most session durations. We found that the top 10% of the heavy sessions accounted for 85% of the total user traffic seen in the trace. Such observations are common in Internet traffic and WLAN measurement studies [13,27]. Also, the top 10% of long sessions accounted for approximately 40% of the total traffic.

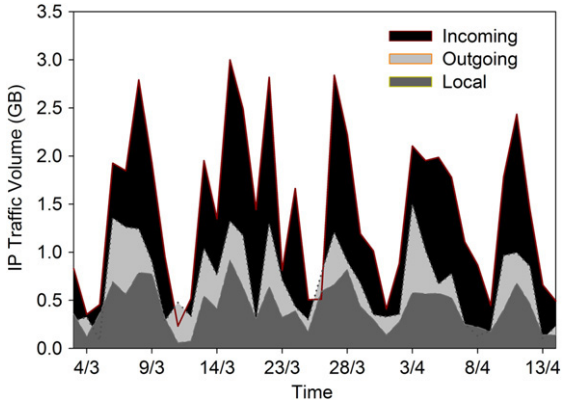
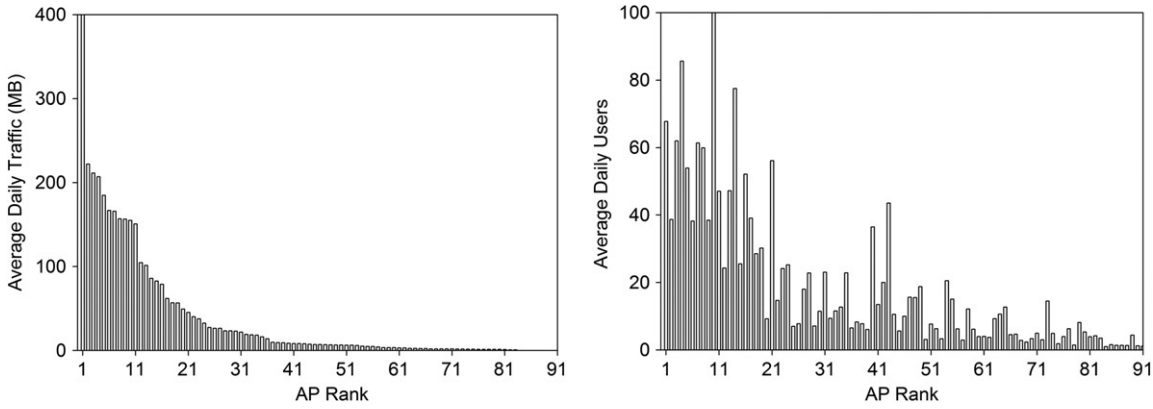


Fig. 15. Data transferred daily.



(a) Average daily traffic.

(b) Average number of user.

Fig. 16. AP load.

10. Network view

This section presents a network view of the trace data analysis. We were interested in knowing what network-layer protocols are seen, how much user traffic the network handles, how this traffic varies with time, how much traffic is accounted for by each location, and how much traffic is transferred in each direction. We also wanted to understand the load on the WLAN and individual APs.

10.1. User traffic load

Fig. 15 shows the variation of daily user traffic load for the trace period. Clearly, users downloaded more than they uploaded. The dips represent weekends when there were few users connected to the network. The network load was fairly low, with the peak daily throughput observed below 1 Mbps. Note that the throughput calculations are based on user traffic only. The actual network load would be higher if Management and Control frames were considered.

10.2. AP load

To understand AP load, we ranked all 97 APs seen based on the average daily traffic they handle. Fig. 16(a) shows the load distribution of APs based on average daily traffic. Fig. 16(b) shows the average daily number of users these APs (in the same rank order) handled. Two observations can be made from Fig. 16. First, load is unevenly distributed across APs in the WLAN. Second, the traffic load on APs is loosely related to the number of users. These observations match those made for a public WLAN [1]. In particular, traffic load depends on the type of users, and the applications they use. Non-uniform AP usage seems to be an inherent characteristic of WLANs.

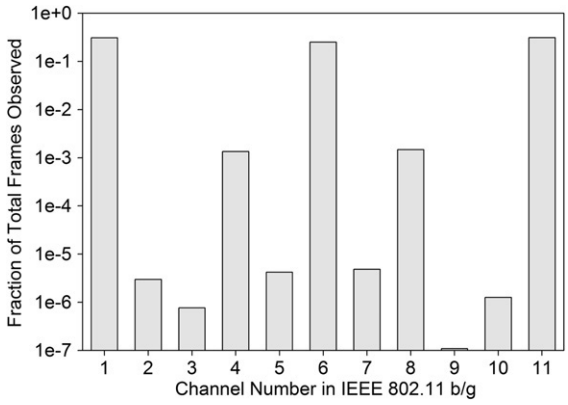


Fig. 17. Distribution of wireless channel usage.

11. Wireless view

Data capture using RFGrabbers enables analysis across all layers of the protocol stack. We finish our analysis with a brief look at several wireless link-layer issues.

11.1. Wireless channel usage

Fig. 17 shows the observed distribution for usage of the 11 physical channels available in IEEE 802.11b/g. As expected, most frame transmissions (99.67%) occur on channels 1, 6, and 11. Load is roughly balanced on these channels.

Valid frame transmissions were observed on every one of the other available channels. This phenomenon can be explained as follows. Each 802.11 b/g channel is 22 MHz wide, but there is only a 5 MHz separation between the centre frequencies of each successive channel. The implication is that we often observe adjacent channel interference or channel bleeding. For example, we captured 0.001 GB of traffic on channels 5 and 7. Similarly channel bleeding can be observed on channels 2 and 10.

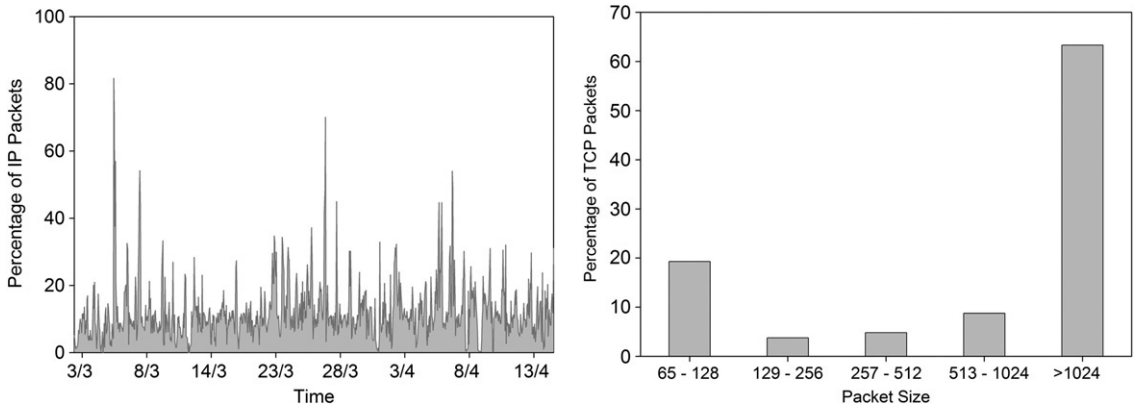
The secondary peaks on channels 4 and 8 are due to some APs in the Student Centre configured (either manually or dynamically) to use these channels. Recent measurement work in the literature indicates that concurrent operation on slightly overlapping channels (e.g., 1, 4, 8, 11) is practically feasible [16,17]. The channel scanning capability of the RFGrabbers indicates that this property is sometimes exploited in our current WLAN deployment.

11.2. Wireless error rate

The overall error rate observed on our WLAN was higher than expected. Approximately 11% of the Data frames captured by the RFGrabbers had CRC errors, representing 29 GB of traffic volume. This can be attributed partly to the placement of the RFGrabbers. Due to poor link quality between the RFGrabbers and APs/clients, the RFGrabbers can capture packets with error. The CRC error rates are fairly consistent across most locations in the WLAN.

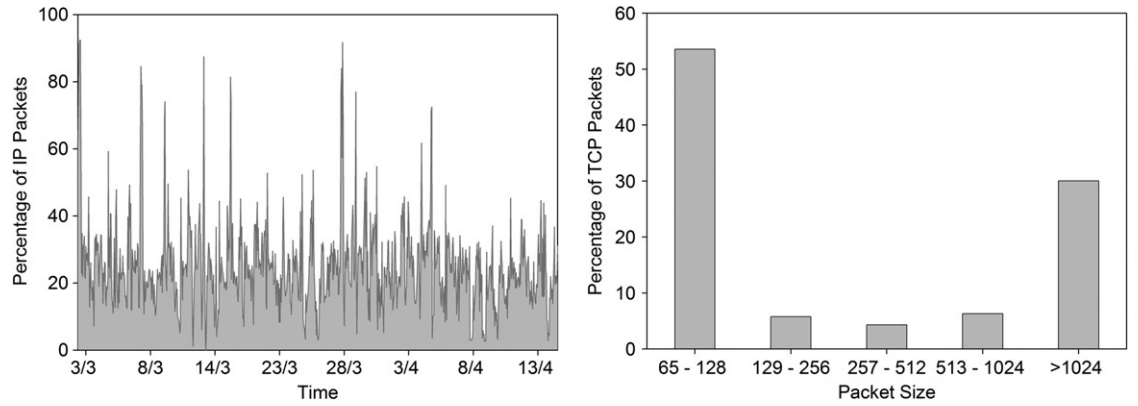
Fig. 18(a) shows the percentage of IP packets that were in error during any hour. By looking at these error statistics we find that in some hours more than 80% of the packets were in error. However, these values are only noticed during the night when there is little network traffic. CRC errors are caused by interference from nearby traffic on the channel, poor radio link, and channel noise. We did a directionality analysis of the CRC error rates and found that 41% of the Data frames in error were sent by wireless clients to the AP, while the APs sent 59% of the Data frames in error. For Management frames, 98% of the errors happened when an AP transmitted the frame. Options for reducing CRC error rates include relocating APs (i.e., to improve line of sight to wireless devices) and changing the transmission frequency.

Fig. 18(b) shows the packet size distribution of TCP packets with error. The majority of the errors were observed for packet sizes that were between 65–128 bytes and packet sizes greater than 1 KB. Our analysis shows that errors are concentrated on those packet sizes that are dominant; in fact, in our network, 52% of the TCP packets had a packet



(a) Error rate. (b) Packet size distribution.

Fig. 18. Packets with CRC errors.



(a) Retransmission rate. (b) Packet size distribution.

Fig. 19. Retransmitted packets.

size of 65–128 bytes and 31% of the TCP packets had a packet size greater than 1 KB. Additionally, the probability of packet corruption increases with an increase in packet size.

Fig. 19(a) shows the percentage of IP packets that had their wireless retry bit set to 1. Approximately 25% of the Data frames observed were flagged as MAC-layer retransmissions and accounted for 23 GB of traffic volume. Wireless retransmissions are caused when the sender does not receive an ACK from the receiver of the frame or receives a corrupted ACK. Similar to the CRC error rate, the wireless retransmission rates are quite consistent across most locations. We noticed that the hourly percentage of wireless retransmission during work days varied between 10% and 80%. About 47% of the Data frame retransmissions and 68% of Management frame retransmissions occurred while wireless clients were transmitting packets to an AP.

Fig. 19(b) shows the packet size distribution of retransmitted packets. We observe that more than 50% of the TCP retransmitted packets were small (<128 bytes). Note that CRC errors are one of the many reasons for packet retransmission. Thus, there is no direct correlation between the results in this figure and those in Fig. 18.

12. Discussion

Although we have only collected and analyzed data from a single academic WLAN, we believe that many of our observations are applicable in other settings. For example, many of the characteristics are an artifact of human behaviour, which we expect to extend beyond our local campus. For example, the choice of wireless devices (by OS) reflects the distribution in much of North America and other parts of the world. The usage patterns clearly reflect

users' daily habits, which are also similar elsewhere. The choice of applications are similar in other locations. Session characteristics are also reflective of global phenomena. For example, session lengths are affected by issues such as available time (e.g., between classes, lunch break), battery life of the wireless device, etc. Finally, the mobility habits of our local users are also expected to be similar to users in other locations. Although WLANs provide users with the freedom to access a network from almost any location, the applications and devices typically restrict mobility. Again, these issues apply in most environments, not just ours. Thus, we expect that our observations and models apply broadly. Although the particular parameter choices will vary, the underlying models may be more widely applicable.

Based on these observations, we believe that our results are helpful to a wide audience. Administrators could utilize the models for capacity planning purposes, perhaps after determining parameters for their own environment. Tool developers could incorporate the models into their capacity planning products, which would make the models accessible to more people. The models may also be helpful for synthetic workload generation. Tools that provide this capability are useful to equipment vendors, network administrators, and researchers alike, who often need to test a product or network under realistic conditions.

13. Related work

In recent years a number of measurement studies have examined WLAN usage, user mobility, and other traffic characteristics. Initially, most WLAN measurement studies analyzed data collected from the wired portion of the network. For example, Henderson et al. [6] used SNMP polling logs, syslog, and tcpdump, Balazinska and Castro [2] used SNMP polling logs, and Schwab and Bunt [22] used authentication logs and Etherpeek traces.

More recent WLAN studies have used passive wireless-side measurement. Yeo et al. [28,29] used multiple sniffers in controlled experiments to merge traces and get an aggregate view of the WLAN. Using this methodology they studied PHY/MAC-layer characteristics of a department WLAN. Jardosh et al. [10,11] used three laptop sniffers to capture wireless packets from an Internet Engineering Task Force meeting and studied link-layer behaviour in a congested WLAN. Rodrig et al. [21] took wireless measurements using five PC sniffers from the SIGCOMM 2004 conference WLAN to study the operational behaviour of the 802.11 MAC protocol. Cheng et al. [5] developed a system called Jigsaw that provides large scale synchronization of wireless traces from distributed sniffers. Mahajan et al. [14] developed a tool (Wit) to merge traces from multiple monitors, infer missed frames, and evaluate the WLAN performance. Sheth et al. [23] built a system consisting of multiple wireless sniffers, a data collection mechanism, and an inference engine to detect anomalies at the physical layer.

Our work is mostly orthogonal to these. We find that the commercial tools simplify the collection and aggregation of wireless measurements, and thus our work instead focuses on the analysis and modelling of WLAN usage.

14. Conclusions

This paper presented a measurement study of a campus WLAN environment, with the data collected using remote passive wireless-side measurement. Our measurements were gathered and aggregated using commercially-available tools. These tools help overcome the challenges of measuring and monitoring a large, geographically-distributed, and heterogeneous WLAN, making WLAN measurement more accessible to network operators and researchers alike.

Our study demonstrated the feasibility and effectiveness of remote non-intrusive wireless-side measurement in a campus WLAN environment. This data collection approach enabled multi-layer analysis from the wireless layer to the application layer. Analysis of our traces identified several trends consistent with prior campus WLAN measurement studies, including diurnal usage patterns, diverse network application usage, and limited user mobility, while offering new observations on session activity, mobility patterns, wireless channel usage, and more. The analysis of our traces, both in aggregate and on a per-location basis, identified similarities and differences in WLAN behaviour across heterogeneous sets of users. Where possible, we statistically modelled the salient features observed across our WLAN datasets.

Our study provides a useful snapshot of the current usage of a large, geographically-distributed campus WLAN. Our analysis identifies several emerging trends in application usage, user mobility behaviour, and WLAN deployment, as well as some performance-related issues at the wireless layer.

Acknowledgements

Financial support for this research was provided by Informatics Circle of Research Excellence (*i*CORE) in the Province of Alberta, as well as by Canada's Natural Sciences and Engineering Research Council (NSERC). The authors thank David Yang, Michael Martini, and Terry Bellward of the University of Calgary Information Technologies Division for deploying the RFGrabber probes in the campus.

References

[1] Anand Balachandran, Geoffrey Voelker, Paramvir Bahl, Venkat Rangan, Characterizing user behavior and network performance in a public wireless lan, in: Proc. of ACM SIGMETRICS Conference, Marina del Rey, USA, June 2002, pp. 195–205.
[2] Magdalena Balazinska, Paul Castro, Characterizing mobility and network usage in a corporate wireless local-area network, in: Proc. of Conference on Mobile Systems, Applications and Services (MobiSys), San Francisco, USA, May 2003, pp. 303–316.
[3] Paul Barford, Mark Crovella, Generating representative web workloads for network and server performance evaluation, in: Proc. of ACM SIGMETRICS Conference, Madison, USA, June 1998, pp. 151–160.
[4] Bro. http://bro-ids.org/.
[5] Y. Cheng, J. Bellardo, P. Benko, Alex Snoeren, Geoff Voelker, S. Savage, Jigsaw: Solving the puzzle of enterprise 802.11 Analysis, in: Proc. of ACM SIGCOMM Conference, Pisa, Italy, September 2006, pp. 39–50.
[6] Tristan Henderson, David Kotz, Ilya Abyzov, The changing usage of a mature campus-wide wireless network, in: Proc. of Conference on Mobile Computing and Networking (MobiCom), Philadelphia, USA, September 2004, pp. 187–201.
[7] Ron Hutchins, Ellen Zegura, Measurements from a campus wireless network, in: Proc. of IEEE International Conference on Communications (ICC), New York, USA, April 2002, pp. 3161–3167.
[8] IEEE OUI and Company ID Assignments. http://standards.ieee.org/regauth/oui/oui.txt, June 2006.
[9] Internet Assigned Numbers Authority. http://www.iana.org/, June 2006.
[10] Amit Jardosh, Krishna Ramachandran, Kevin Almeroth, Elizabeth Belding-Royer, Understanding congestion in IEEE 802.11b wireless networks, in: Proc. of Internet Measurement Conference (IMC), Berkeley, USA, October 2005, pp. 279–292.
[11] Amit Jardosh, Krishna Ramachandran, Kevin Almeroth, Elizabeth Belding-Royer, Understanding link-layer behavior in highly congested IEEE 802.11b wireless networks, in: Proc. of ACM SIGCOMM Workshop on Experimental Approaches to Wireless Network Design and Analysis (E-WIND), Philadelphia, USA, August 2005, pp. 11–16.
[12] Kismet Wireless FAQ. http://www.kismetwireless.net/documentation.shtml.
[13] David Kotz, Kobby Essien, Analysis of a campus-wide wireless network, in: Proc. of Conference on Mobile Computing and Networking (MobiCom), Atlanta, USA, October 2002, pp. 107–118.
[14] R. Mahajan, M. Rodrig, David Wetherall, John Zahorjan, Analyzing the MAC-level behavior of wireless networks in the wild, in: Proc. of ACM SIGCOMM Conference, Pisa, Italy, September 2006, pp. 75–86.
[15] MaxMind. GeoLite Country Database. http://www.maxmind.com/app/geoip_country, June 2006.
[16] Arunesh Mishra, Eric Rozner, Suman Banerjee, William Arbaugh, Exploiting partially overlapping channels in wireless networks: Turning a peril into an advantage, in: Proc. of Internet Measurement Conference (IMC), Berkeley, USA, October 2005, pp. 311–316.
[17] Arunesh Mishra, Vivek Shrivastava, Suman Banerjee, William Arbaugh, Partially overlapped channels not considered harmful, in: Proc. of ACM SIGMETRICS Conference, Saint-Malo, France, June 2006, pp. 63–74.
[18] Timo Ojala, Toni Hakanen, Timo Mäkinen, Veikko Rivinoja, Usage analysis of a large public wireless lan, in: Proc. of Conference on Wireless Networks, Communications, and Mobile Computing (WirelessCom), Maui, USA, June 2005, pp. 661–667.
[19] S. Pilosof, R. Ramjee, D. Raz, Y. Shavitt, P. Sinha, Understanding TCP fairness over wireless LAN, in: Proc. of IEEE INFOCOM Conference, San Francisco, USA, March 2003, pp. 863–872.
[20] pOf. http://lcamtuf.coredump.cx/p0f.shtml.
[21] Maya Rodrig, Charles Reis, Ratul Mahajan, David Wetherall, John Zahorjan, Measurement-based characterization of 802.11 in a hotspot setting, in: Proc. of ACM SIGCOMM Workshop on Experimental Approaches to Wireless Network Design and Analysis (E-WIND), Philadelphia, USA, August 2005, pp. 5–10.
[22] David Schwab, Rick Bunt, Characterising the use of a campus wireless network, in: Proc. of IEEE INFOCOM Conference, Hong Kong, China, March 2004, pp. 862–870.
[23] Anmol Sheth, Christian Doerr, Dirk Grunwald, Richard Han, Douglas Sicker, MOJO: A distributed physical layer anomaly detection system for 802.11 WLANs, in: Proc. of Conference on Mobile Systems, Applications and Services (MobiSys), Uppsala, Sweden, June 2006, pp. 191–204.
[24] Diane Tang, Mary Baker, Analysis of a local-area wireless network, in: Proc. of Conference on Mobile Computing and Networking (MobiCom), Boston, USA, August 2000, pp. 1–10.
[25] WildPackets. http://www.wildpackets.com/.
[26] WildPackets, Airopeek NX User Manual, 2003.
[27] Carey Williamson, A tutorial on internet traffic measurement, IEEE Internet Computing 5 (6) (2001) 70–74.
[28] Jihwang Yeo, Moustafa Youssef, Ashok Agrawala, A framework for wireless lan monitoring and its applications, in: Proc. of ACM Workshop on Wireless Security (WiSe), Philadelphia, USA, October 2004, pp. 70–79.
[29] Jihwang Yeo, Moustafa Youssef, Tristan Henderson, Ashok Agrawala, An accurate technique for measuring the wireless side of wireless networks, in: Proc. of Workshop on Wireless Traffic Measurements and Modeling (WiTMeMo), Seattle, USA, June 2005, pp. 13–18.

Aniket Mahanti is a Research Associate in the Department of Computer Science at the University of Calgary. He holds a B.Sc. (Honours) in Computer Science from the University of New Brunswick, and an M.Sc. in Computer Science from the University of Calgary. His research interests include IEEE 802.11 networks, measurement, modelling, characterization, and performance evaluation.

Carey Williamson holds an iCORE Chair in the Department of Computer Science at the University of Calgary, specializing in broadband wireless networks, protocols, applications, and performance. He holds a B.Sc. (Honours) in Computer Science from the University of Saskatchewan, and a Ph.D. in Computer Science from Stanford University. His research interests include Internet protocols, wireless networks, network traffic measurement, network simulation, and Web performance.

Martin Arlitt is a Senior Research Associate in the Department of Computer Science at the University of Calgary and a Senior Scientist at HP Labs. His research interests include measurement, characterization, and performance evaluation of computer systems and networks.