Measurements From a Campus Wireless Network

Ron Hutchins Ellen W. Zegura
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280

***Abstract*— In this paper we examine a high speed wireless access network and present traffic patterns and on-line behavior for wireless users. To date, wireless network studies have largely focused on cellular voice technologies and architectures. Here we present an analysis of data collected from an authenticated campus area access network providing both wireless 802.11b and walk-up Ethernet capabilities through an authenticated access network service. We present an analysis of session data, transport layer flow data, and movement data taken from 109 wireless access points spread across 18 buildings. Local Area Wireless Network (LAWN) wireless services support about 444 wireless hosts, 765 current users, and have sustained more than a million TCP flows over the nearly two month collection period.**

I. Introduction

Wireless access network capabilities enable users to move quickly and easily between different locations, with the potential for uninterrupted network activity. In this study we examine a high speed wireless access network and present traffic patterns and movement characteristics for wireless users.

To date, wireless network studies have largely focused on cellular voice technologies and architectures (c.f. [1]). As an exception, Tang and Baker investigated a metropolitan wireless network[2] and a wireless implementation inside a single building[3]. Here we present an analysis of data collected from a campus area network providing wireless 802.11b[4] access capabilities, called the Georgia Tech Local Area Wireless Network (LAWN). Access to the campus network and the Internet from the LAWN is initiated by user log-in, which registers the user's current host IP address with the firewall, and terminated by idle timers which remove the host entry from the firewall. This authentication to time-out period defines a LAWN *session* in our discussion.

In this paper, we present an analysis of 138 days (January 2 through June 2, 2001) of session data, 54 days (April 9 through June 2, 2001) of transport layer flow data, and movement data taken from 109 wireless access points, spread across 18 buildings. LAWN wireless services currently support about 444 hosts utilizing wireless 802.11b network attachment. The LAWN has supported more than a million TCP flows over a nearly two month collection period. This data is unique in that sessions, similar in characteristics to PPP [5] sessions on dial-up networks, are captured, anonymized, and associated with flow and movement data for the wireless user. The use of these sessions in the analysis permits comparison to session characteristics of other access network technologies [1].

We present basic session statistics including session counts based on time-of-day and session length distribution. We observe linear growth in number of users in this early phase of the LAWN project. We also observe diurnal behavior of users, with a peak usage time in early evening and minimum usage in early morning. Short hold times strongly dominate session lengths for wireless users.

We provide aggregate flow data consisting of flow length distribution and flow counts on a time-of-day basis. We provide counts and mean flow lengths for the most popular ports based on observed usage and discuss the differences between interactive protocols like secure shell (SSH) and transactional protocols like HTTP.

Finally, we define and document inter-move time for wireless users. We describe an example user's behavior and the interactions between sessions, flows, and movement, showing that movement in the LAWN causes minimal disruption of flows and sessions.

This data presents a snapshot in time of a new and growing, mobility-enabled campus network with over 15,000 potential users. It also presents early documentation of traffic patterns of wireless users across multiple buildings. The results may be useful in differentiating wireless data traffic from cellular voice network traffic and for planning future wireless and mobile supporting systems. Although the number of users today is small, this data, especially the movement data, may be useful in simulation and modeling of wireless networks.

The next section (II) presents more details on the dataset collected. Section III-A presents an aggregated user view of our analysis; while, Section III-B delves into individual characteristics of users on the wireless system. Section III-C presents an example user's movement patterns and an indepth analysis of port usage. Mean flow length on a per-port basis is presented in Table I. Related work is presented in more detail in Section IV, and we summarize conclusions in Section V.

II. Background

In this section, we describe the details of the data collection system, and the process used in anonymizing the data [2].

[1] For an examination of dial-up access network data, see [6].

[2] We point the interested reader to a longer version of this paper [7], which contains details on the wireless architecture and implementation, as well as more detailed plots of the analysis results.

The LAWN consists of about 110 Lucent Orinoco 802.11b wireless access points placed for maximum coverage across 18 buildings that form the Wireless Corridor across the Georgia Tech campus. Inter-building wireless connectivity is only available today on a small scale. Users are required to log-in with Georgia Tech Kerberos credentials through an SSL encrypted tunnel to the LAWN authentication system. Upon authentication, the user's host is registered with the firewall and traffic can pass to the campus network or the Internet. This local authentication enables any of the approximately 15,000 users on the campus to access the wireless network while preventing access by others.

Three different data sets are currently being collected on the LAWN:

- Session information, collected from the Authentication System.
- Flow information, collected between the firewall and the LAWN router.
- Movement information, collected via SNMP polling of tables on the access points themselves.

LAWN *session* information is continuously collected and logged. This data consists of event records containing time-of-day of event, Kerberos user identifier, IP address of the communicating host, and an event code. Events include firewall open (FWOPEN), firewall close (FWCLOSE), and multiple error conditions including authentication errors. Session length is the time between an FWOPEN requested for the authenticating host and the FWCLOSE request issued by a polling process. Polling consists of multiple ping requests sent to each authenticated host 60 seconds after traffic was last seen from this host. Unresponsive hosts are removed from the host access table and an FWCLOSE event is logged.

The first step in processing the collected data involves anonymizing the data by assigning a sequential number (UID) to each user identifier on first appearance. This unique mapping is maintained throughout the processing period via a local database. Anonymized authentication records are next matched by UID to form FWOPEN/FWCLOSE pairs which are then reduced further to authentication records consisting of start time, stop time, length, UID, and dynamically assigned host IP address. This information forms the basis for the session time-of-day and session length analyses. We utilize about 150 days of session data for this analysis.

Transport layer flows are defined as a five-tuple of source IP address, destination IP address, protocol, source port number, and destination port number. Flows are defined in a single direction; a bi-directional conversation has two uni-directional flows associated with it, which may have different starting and ending times. Flow data is captured on the LAWN between the firewall/authentication system and the LAWN router using a collection device running an in-house developed program. The flow data is captured in five minute snapshots indexed by the five-tuple. Any flow that spans multiple five minute intervals appears in more than one five minute sample snapshot [3].

No TCP SYN/FIN bit information is present in the currently available flow data. Since the protocol-defined beginning and end of the flow are not available, a heuristic time-out process is implemented to reassemble flow segments. Claffy and Braun [8] discuss a method for iteratively processing flow data using increasing time-out values to associate flow segments, and cite other research which suggests time-out values of five and fifteen minutes. The expanded version of our paper [7] details the timeout used and rationale.

In the 802.11 protocol suite, wireless hosts choose a single access point as network point of attachment based on criteria such as RF signal strength. Wireless access points create datalink bridge tables containing the associated MAC addresses to build the minimum spanning tree for the wireless network. We poll these bridge tables on a fifteen minute interval to collect this host data. A *move* is identified by observing a host that changes access points between two successive polls.

IP addresses for the LAWN are dynamically distributed via a DHCP server on the authentication system. About 500 addresses, drawn from two contiguous class C blocks, provide IP addresses for all devices on the LAWN, including the wireless access points. These addresses have permanent reverse lookup names and no dynamic DNS mapping is performed. Several users may utilize the same IP address over time, complicating later analysis. We do not provide any source or destination IP address information in order to ensure the privacy of wireless users. Data on errors, including authentication errors, are not processed in this research.

A complete data analysis requires associating related data elements from each of the three data sets. Since DHCP[9] may reissue an IP address to multiple users over time, session and flow data together are necessary to relate the user and IP address with the wireless host during the appropriate time intervals. SNMP polling data completes the picture by providing the location of the host for this time.

This study includes session and flow data from walk-up user ports in the Georgia Tech Library and other buildings. Walk-up users directly connect laptop computers to the network via Ethernet ports, authenticating through the same LAWN authentication system and drawing IP addresses from the same DHCP pool. MAC addresses from the SNMP polling of the bridge tables in the access points contain only wireless MAC addresses and so the movement data includes only wireless users.

III. Data Analysis

In this section, we present aggregate and individual user views of traffic statistics on the LAWN. We present overall data on sessions, transport level flows, and movement of users. We show length and time-of-day count statistics for sessions and flows, and growth data for LAWN users over the period of the

[3] The five minute capture interval is a compromise in the collection device that permits logging of both TCP and UDP data. Though TCP flows can be captured based on the SYN/FIN bit combination, UDP data with no SYN/FIN delimiters would require "interpretation" during collection, decreasing flexibility in the analysis stage.

analysis. Data on individual users includes session characteristics, flow length and count, and individual movement characteristics. Sessions are grouped by the user identifier (UID). Flows are identified by five-tuple. Movement is identified by host MAC layer (hardware) address.

A. *Overall Characteristics of Traffic*

From an overall perspective, this data analysis documents a strong diurnal cycle for LAWN users with a peak session time between 2pm and 5pm and a minimum usage time between 4am and 6am. Flow peak occurs between 3pm and 6pm and the minimum is between 5am and 7am. We show that both session and flow lengths are dominated by short hold times and that there are more than an order of magnitude more flows than sessions. We find that 65% of the wireless users exhibit movement during the analysis period, with four times as many users moving between buildings as there are moving within the same building.

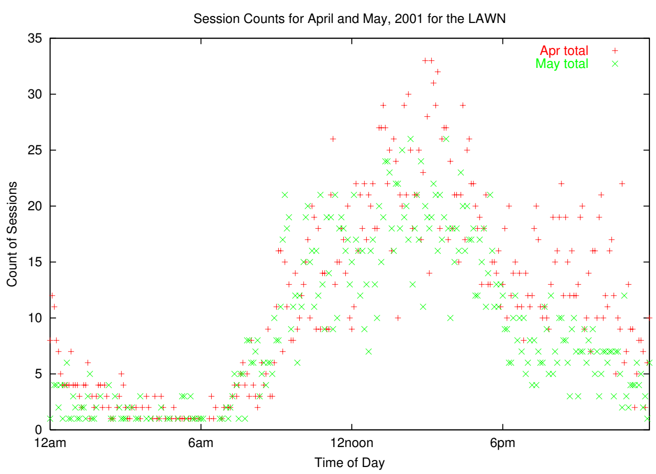

Fig. 1. Session Counts by Time of Day for Authenticated Sessions

Aggregate Count of Sessions Based on Time-of-day. Figure 1 shows the distribution of session starts for each five minute interval during the day. This plot represents the aggregate of all data for the months of April and May, 2001. In this data a diurnal cycle is obvious, exhibiting a low value of one session starting in a five minute interval between 4am and 6am, and a high count of 29 sessions starting between 2pm and 5pm. The ratio of peak count to mean count for April and May are 2.6 to 1 and 2.7 to 1 respectively. Weekends show a significant dropoff in session starts with weekday mean session count about 1.5 times greater than weekend count. Semester break occurs in May, possibly accounting for the lower session count for the month.

Session Length Distribution. Session length on the LAWN is the time between the FWOPEN requested for an authenticating host and the FWCLOSE request issued by the polling process. Session lengths are collected using a five minute bin size. Session length is dominated by short sessions with a 529 second mean in April, and a 611 second mean in May. Of the total 2349 sessions in April, 21 extended past 24 hours. In May, 14 of the 1860 total sessions were longer than a day. A very few sessions extended over the entire collection period, perhaps due to an error in the terminating polling process.

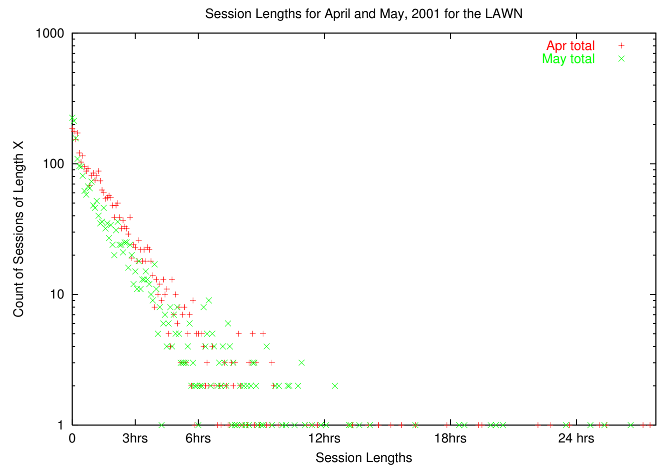

Fig. 2. Session Lengths by Time of Day for Authenticated Sessions

In Figure 2, the y-axis shows the counts for sessions with a given length for two months of the data collection, April and May, 2001. The x-axis shows session lengths in ten minute increments from zero to about 28 hours. The plot shows a high percentage of sessions lasting less than 30 minutes, but significant counts are seen lasting up to six hours. Very few sessions extend beyond the twelve hour mark. The LAWN does not currently implement an enforced time-out for authenticated users and has no immediate plans to terminate sessions automatically.

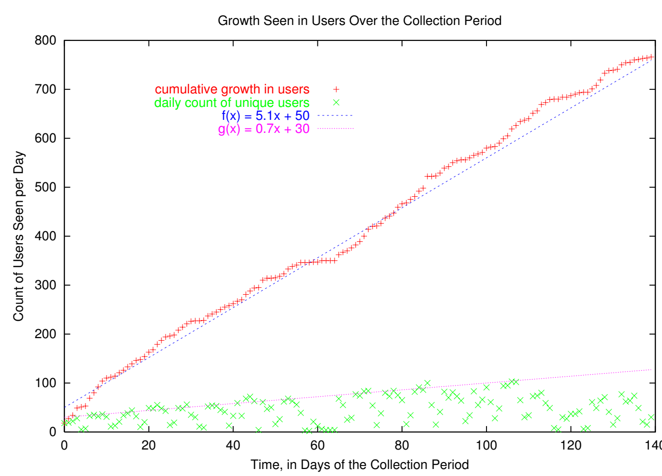

Fig. 3. Growth in Users, Cumulative and Unique per Day, January - May, 2001

Growth of Users. Figure 3 plots the cumulative and unique per-day user counts between January 2 and June 2, 2001. The x-axis shows consecutive days of the collection period and the y-axis shows count of users. The data plots show one data point per day representing the count of users for each dataset on that day.

During this period, the count of users grew from less than 30 in early January to 765 on June 2. The lower curve shows a large difference in the number of active users on weekdays as contrasted to weekends. The data points fall to the tens of users on weekends versus about a hundred during weekdays. This is expected on a university campus. The effect of holidays and end-of-term can be seen in the middle (March 5-9, Spring Break, around day 60) and near the end of the graph (May 5-11, end-of-term break, around day 110). The cumulative user count shows steady, approximately linear growth. The rate of growth is represented by f(x) in the graph which has a slope of 5.1, indicating an increase of about five users per day over the period. For the count of unique users per day, as indicated by g(x) in the graph, the rate increases at less than one per day.

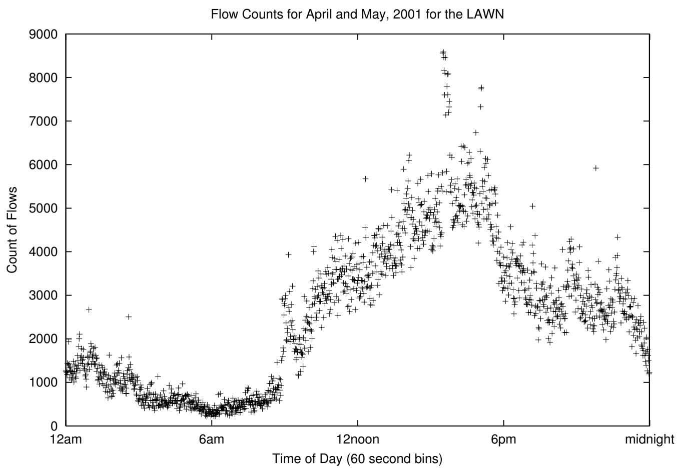

Fig. 4. Counts by Time of Day of All Flows

Flow Counts. For the months of April and May, 2001, Figure 4 closely follows the distribution of sessions over time-of-day, but with flows having a greater-than-two-orders-of-magnitude higher peak count during a day. An active session is required in order to create a flow on the LAWN, mandating a minimum one-to-one relationship between sessions and flows. Like sessions, the low count for flows occurs between 5am and 7am, however, the peak flow count is between four and five pm, about an hour later than peak session count. One possible explanation is that users increase the number of flows that they utilize over time in a single session.

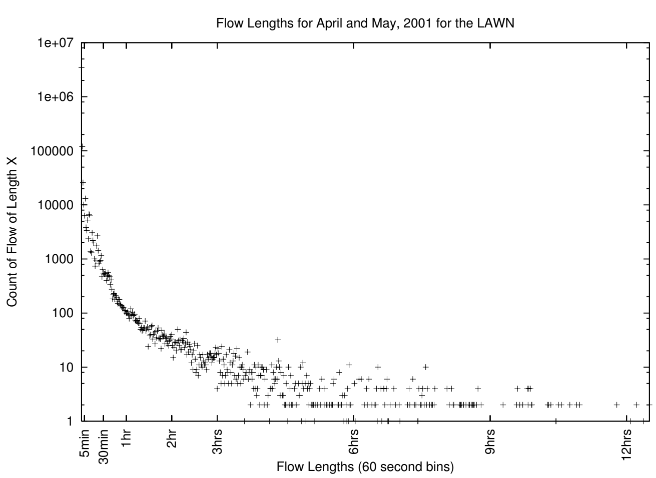

Fig. 5. Lengths of All Flows

Flow Length. Figure 5 presents aggregate flow length data for April and May, 2001. The x-axis identifies flow lengths from 60 seconds through more than 12 hours quantized into 60 second increments. The y-axis shows the counts for flows of length x. Flow lengths are again related to session lengths. Short flow hold times, under thirty minutes, strongly dominate with a large percentage under 5 minutes in length and nearly one million at sixty seconds or less. Flow lengths of several hours are visible, with a small number approaching nine hours in length.

TABLE I

MEAN FLOW LENGTH AND COUNTS FOR SELECTED TCP PORTS

Service Name	Port Number	Mean Flow Length (sec)	Count of Flows
FTP-data	20	55.879	19323
FTP-control	21	135.402	7065
SSH	22	935.896	2285
Telnet	23	1067.100	1976
SMTP	25	6.868	2846
DNS	53	0.109	4232
HTTP	80	15.571	1336592
POP3	110	7.188	57215
IDENT	113	0.219	37469
NNTP	119	268.514	1039
NETBEUI	139	585.654	1676
IMAP	143	38.919	6414
SSL	443	42.927	141733

Specific port counts and mean flow length for selected ports and services are presented in Table I. This data shows the high percentage of short sessions and identifies certain interactive applications, such as SSH and Telnet, that dominate the longer session lengths.

Movement of Hosts on the LAWN. The early nature of this wireless implementation and the associated availability of movement capabilities allow us to investigate movement patterns for users on the network. Having the capability to move does not imply that users will move. While some may readily adopt the movement capabilities over a wide range of locations, some may never take advantage of these capabilities. We investigate these patterns and document individual movement of users.

In this study, we observe that for the 444 wireless MAC addresses the following applied:

- 158 did not move
- 58 moved only within the same building
- 228 changed buildings

Figure 6 shows the distribution of all inter-move times for all hosts (by MAC address) over the period from April 11 through June 2, 2001. The x-axis shows inter-move times in five minute bins. The y-axis shows the count of wireless hosts that exhibit inter-move times of x seconds at some time during the collection period. Both the x and y-axes are displayed using log scales

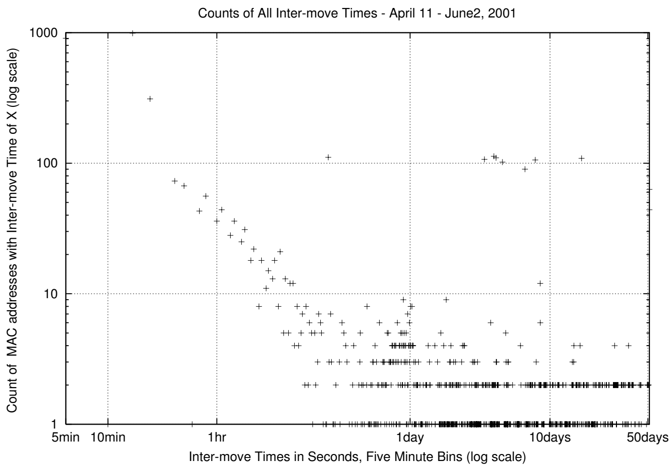

Fig. 6. Inter-move Time for MAC Addresses, Log Scale

to show the extremes in the shortest and longest time intervals for host movement [4]. The movement data was quantized into five minute buckets. In this study, a move mandates a change in access point. The end of one session and the creation of another on the same access point does not constitute a move.

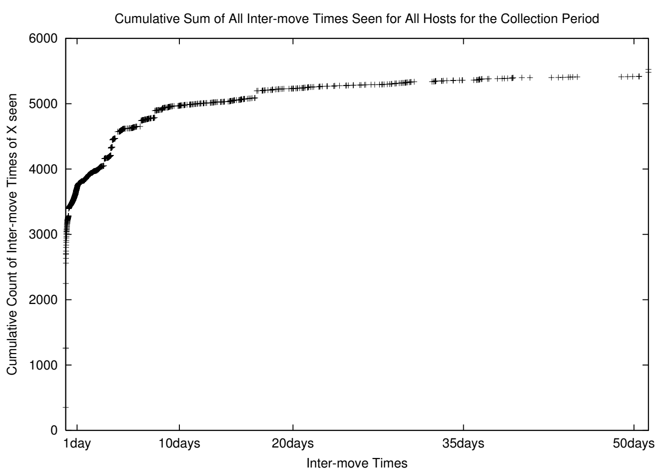

Fig. 7. Cumulative Distribution of Inter-Move Times

The plot shows significant numbers of moves in the 0-to-5 minute and 10 - 15 minute intervals. One host, which oscillated regularly between two access points, may in part explain this data, since that host was seen to move more than a hundred times between the two access points. It is interesting to note that, although the distribution of data falls generally in a line on this log-log scale, there are significant points at around eight hours, at three to five days, and around twenty to twenty five days. This is likely due to some users following a daily work schedule moving from lab facilities to offices, or from weekly and monthly behavior. This data shows a great variety in user movement patterns, but still supports a large percentage of moves falling in the 15 minute to one day range. More work is needed in this area to understand user movement patterns on wireless data networks.

The cumulative distribution plot in Figure 7 shows the cumulative sum of individual inter-move times for all hosts over the collection period. The x-axis shows inter-move times and the y-axis shows the running sum of counts for these inter-move times. The plot shows that the largest percentage, more than half, of the inter-move times are less than one day with a very significant portion of that data above the one hour mark. This supports a hypothesis that laptop computers with keyboards do not lend themselves to rapid movement, and indicates that users may move across several different locations but work while stationary much of the time.

B. *Individual User and Host Data*

Now we move to more detailed information on individual users and hosts. We attempt to answer questions about how individual users differ in their specific use of the wireless network. Understanding differences between users is important in provisioning wireless networks, modeling and simulation of user movement patterns, and in designing protocols which support mobility.

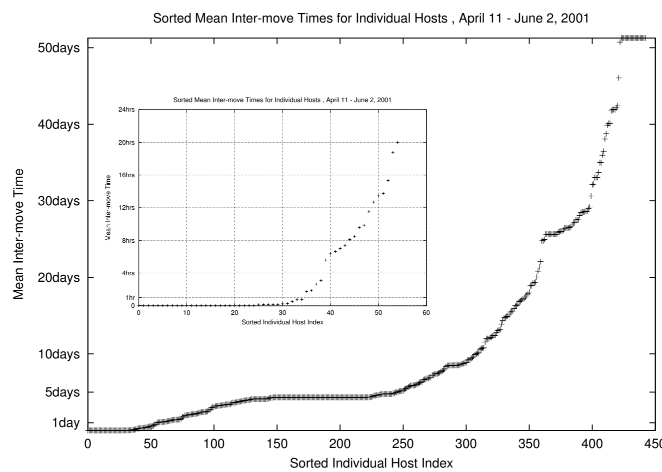

Fig. 8. Inter-move Time: Overall and Expanded Scale

We present data compiled from the individual wireless access points in the network representing the wireless hosts. Inter-move times for individual hosts as a cumulative sum of individual times and as a sorted mean value per host are shown. The counts of access points visited by individual hosts and hosts seen by the individual access points are presented to explore relationships between access points and hosts in the network. This data may also be important for load balancing and for provisioning wireless networks. We present flow lengths and counts for four different individual hosts which may represent a long term view of wireless use.

Mean Inter-move Times for Individual Hosts. Figure 8 shows the mean inter-move data for each wireless user. The

[4] Although the polling time for the SNMP movement data is fifteen minutes, we observe that if a host moves after a poll of its current access point to another access point sequentially next in the polling cycle, it is possible to see move times shorter than fifteen minutes. In fact, very short move times, in the tens of seconds range, are seen in the data. This is a result of polling and the polling interval.

x-axis shows individual host index (anonymized MAC address) while the y-axis presents mean inter-move time. The data is presented as sorted by inter-move time to show the predominance of mean inter-move times under five days time. The inset plot expands the data for host indexes with inter-move times less than 24 hours. The expanded data shows 24 inter-move times of zero indicating that 24 hosts appeared only once in the access point data. Since inter-move time is directly related to the time a host is associated with an access point, long inter-move times generally appear to be from stations that did not move during the data collection, but did appear more than once in the SNMP data.

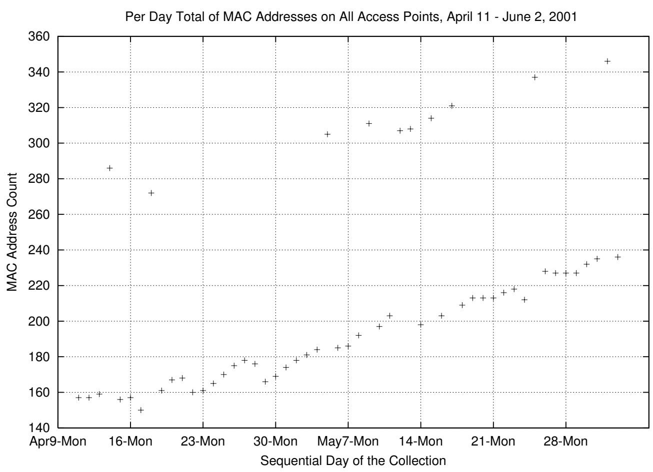

Fig. 9. Growth in MAC Addresses

Growth from the MAC Address Perspective. Another way to observe growth in use of the LAWN system is to examine the counts of unique MAC addresses seen on a per day basis. Shown in Figure 9, we present sequential days of the collection on the x-axis and the count of unique MAC addresses seen on the y-axis. Like the authentication data, this plot shows dips on weekends and a steady increase over the nearly eigth-week period. Finals week is somewhat noticeable during the week of April 30. The growth shows a general increase of about 1.7 new wireless hosts per day over the collection period. We note significantly increased counts on ten days of the period, where the counts exceed the expected value by approximately a hundred hosts. The data points follow nearly the same difference in hosts on each of the ten days where the high count appears. Examination of this data suggests that a number of hosts are participating in project work in the same building at around the same times.

Individual Flow Data. Figure 10 presents flow data for four hand-selected IP addresses on the LAWN. Since DHCP is used, an IP address may represent more than a single user. These addresses were selected specifically for the length of time that they used the wireless network, and therefore, the quantity of data available on them individually. Since the network is a new service, many users do not have enough history on the LAWN to show any trends in their use. In these plots the x-axis shows flow length quantized into sixty second bins. The y-axis shows

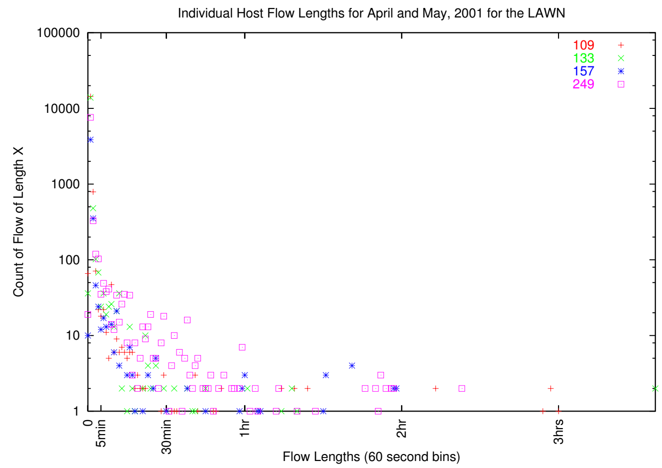

Fig. 10. Individual Flow Lengths

the counts of all flows for each user aggregated across April and May, 2001. Each of the four addresses shows slightly different characteristics but all exhibit a high percentage of flows with lengths of five minutes or less. User 249 shows groups of flow lengths in the less-than-thirty minute range. User 157 shows fewer flows, with the majority grouped around the five to fifteen minute range. As discussed earlier, certain applications like SSH and Telnet tend toward longer hold times on flows due to their interactive nature. The mix of applications that a user utilizes has a strong influence on the length of flows.

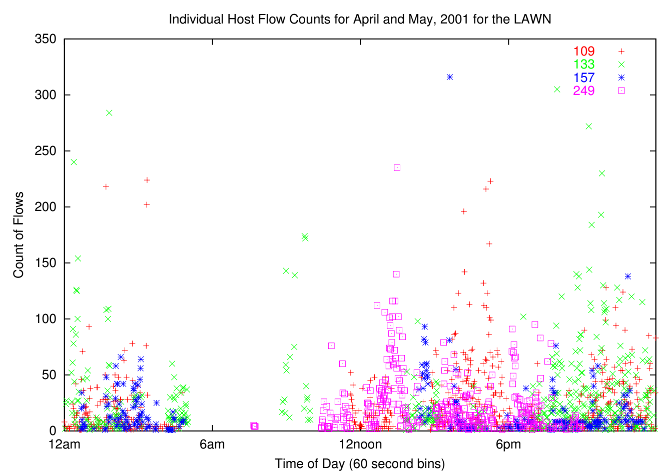

Fig. 11. Individual Flow Counts

Individual flow counts are shown in Figure 11 for the same four addresses above. Again, the four exhibit different time-of-day characteristics with user 249 appearing predominantly in the early afternoon and evening. User 133 shows activity over most of the day except between 5am and 8am but appears to have some regularity in flow activity around a four or five hour interval. User 157 shows activity in the early afternoon at around 3pm, the evening between 8pm and 11pm, and again in the early morning hours between 1am and 5am. User 109 appears in two major time intervals: 3-6 pm and 2-4 am. Again, differences in peak time-of-day for a single address may be due to different users acquiring the same DHCP address over time.

C. An Example of a Moving User

Finally, to investigate the coupled relationship between sessions, flows and movement, we present data from a single moving user relating all these aspects of the data. We attempt to answer the question, "At this point in time, does movement disrupt sessions and flows for connected users?" For this study we programatically simulate the session and flow start and end times and impose the movement impulse function at appropriate times to observe interactions. For the example user, chosen based on a relatively high move count, we observe fifty-two total sessions, 1691 unique flows, interspersed with seventeen moves. Over the study we see two moves take place inside of sessions, interrupting 8 total flows. The resulting impact on the user, though noticeable, does not appear significant. For the moment, keyboard data entry may be a limiting factor for user movement. Voice services appear to be inherently different from data services, and voice over IP on wireless networks will be impacted in significantly different ways as compared to non-voice services.

IV. Related Work

Most closely related to our work are studies of wireless access networks and user populations. Tang and Baker [3] describe an analysis of a 12-week study consisting of seventy four users utilizing thirteen wireless access points within a single campus building to access campus and Internet resources. They document movement of users across the access points, time-of-day information for usage, and information on specific applications used on the network. In looking at how often users are active, their results generally agree with ours, in that more users are active on fewer days, while a few users are active on many days, though the limited size of their user population makes this trend less clear.

Tang and Baker also present an analysis of data [2] from a commercial wireless metropolitan area implementation, Metricom's Ricochet Network service. In this analysis, seven weeks (February and March, 1998) of trace data from the packet radio network are analyzed in order to understand how wireless network users take advantage of the mobile environment. They find that users tend to use wireless during non-work hours, hypothesizing the use of faster wired technologies at work. This finding may support the low usage time between 7am and 11am in our study. More than half of the users in the study move between multiple locations during the collection period. This compares with our finding of 158 stationary users of 444 total on the LAWN.

V. Conclusions

The Georgia Tech Local Area Wireless Network offers a unique opportunity to study a new access technology during its early implementation and growth phases. The size of the dataset is not sufficient today to conclusively characterize wireless campus users. The results offer insight into how this new technology is being used currently and may over time provide a view of changing traffic and movement characteristics as users take advantage of new capabilities. In this study we find that:

- Basic time-of-day session and flow counts resemble the diurnal behavior discussed in previous work, possibly relating to campus activity as opposed to the commercial work world or home usage. A peak usage period around 4pm suggests end-of-workday use.
- Session and flow lengths are dominated by short hold times but exhibit significant counts of longer sessions and flows. The count of flows is at least an order of magnitude greater than the count of sessions.
- Users exhibit variability in movement, with nearly half of the users never moving and others moving regularly and across multiple buildings. Current technologies do not easily support movement while sessions and flows are in progress but future technologies may be more heavily impacted by a lack of transparent handoff during movement. The mean time between moves is in general longer than mean session and flow lengths.

References

[1] D. Lam, Y. Cui, D. Cox, and J. Widom, "A location management technique to support lifelong numbering in personal communications services," in *Mobile Computing and Communications Review*, January 1998, vol. 2, no.1, pp. 27–35.

[2] Diane Tang and Mary Baker, "Analysis of a metropolitan-area wireless network," in *Proceedings of MOBICOM 1999*. August 1999, pp. 13–23, ACM Press.

[3] Diane Tang and Mary Baker, "Analysis of a local-area wireless network," in *Proceedings of MOBICOM 2000*. August 2000, pp. 1–10, ACM Press.

[4] "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," IEEE 802.11 Standard, ISBN 1-55937-935-9, 1997.

[5] W. Simpson, "The point-to-point protocol (PPP)," Internet Request for Comments 1661, July 1994.

[6] R. Hutchins, E. Zegura, A. Liashenko, and P. Enslow, "Internet user access via dial-up networks - traffic characterization and statistics," in *ICNP 2001: 9th International Conference on Network Protocols*, November 2001.

[7] R. Hutchins and E. Zegura, "Measurements from a wireless campus network," College of Computing, Technical Report, August 2001, ftp://ftp.cc.gatech.edu/pub/coc/tech_reports/2001/GIT-CC-01-21.

[8] K. C. Claffy, Hans-Werner Braun, and George C. Polyzos, "A parameterizable methodology for internet traffic flow profiling," http://www.nlanr.net/Flowsresearch/Flowspaper/flows.html.

[9] R. Droms, "Dynamic host configuration protocol," Internet Request for Comments 2131, 1997.