Characterizing the duration and association patterns of wireless access in a campus

Maria Papadopouli[a] and Haipeng Shen[b] and Manolis Spanakis[c]
IBM Technical Contact: Dr. Mark Pozefsky, Pervasive Computing, SWG

***Abstract*— Our goal is to characterize the access patterns in a IEEE802.11 infrastructure. This can be beneficial in many domains, including coverage planning, resource reservation, supporting location-dependent applications and applications with real-time constraints, and producing models for simulations. We conducted an extensive measurement study of wireless users and their association patterns on a major university campus using the IEEE802.11 wireless infrastructure. We characterized and analyzed the wireless access pattern based on several parameters such as mobility, session and visit durations.**

We show that the mobility and building type affect the session and visit durations. As the mobility increases, the visit duration tends to decrease stochastically. The opposite happens in the case of the session duration. Moreover, there exist different stochastic orders among visit durations of different building types when conditioning on session mobility. A family of BiPareto distributions can model the visit and session duration.

I. Introduction

IEEE802.11 networks are becoming widely available in universities, corporations, and residential areas to provide wireless Internet access. Such networks are also increasingly being deployed in airports, hospitals, shopping centers, and other public areas. The deployment of the wireless infrastructure in all these environments impacts the way users access the information. For how long is a wireless client associated with an access point (AP)? What is the duration of its continuous wireless access to the Internet and for how long does it stay disconnected? What is its AP trajectory as the device roams through the wireless infrastructure? How do association patterns of different types of clients (with respect to the device, usage pattern, location, setting, mobility) differ?

Currently, most of the simulation studies on wireless networks and protocols consider simplistic association and mobility patterns for the wireless users [1], [2]. There are a few studies on the mobility and association patterns in cellular networks [3], [4], [5]. However, the rapid deployment of the IEEE802.11 infrastructures in various environments triggers new applications and services, that in turn, generate a richer set of traces for analysis. There is a need for more realistic models of user communication and association patterns. This can be beneficial in capacity planning, administration and deployment of wireless infrastructures, protocol design for wireless applications and services, and their performance analysis.

The key issue that drives this study is the characterization of the continuous wireless access and user mobility pattern. For that, we define the *session* of a client to identify the *continuous* associations of this client to the wireless infrastructure. We characterize the session based on several features and distinguish two main components of the movement pattern, namely, its continuous access to the wireless infrastructure and its disconnection periods. Each session can be described by a trajectory to a sequence of APs and the duration spent at each AP.

Both application and infrastructure designers can exploit the trajectory and duration estimations to support caching, prefetching, graceful handoffs, resource reservation, and capacity planning at APs. Access points, proxies, and servers can use the estimation of their clients's visit duration and next association to prepare the handoff, share clients or traffic load with each other, and ensure a better quality of service characteristics.

This research extends our earlier study [6], the studies by Kotz and Essien [7], Balachandran *et al.* [8], Tang and Baker [9], and Balazinska and Castro [10] by focusing more closely on the association and mobility patterns of individual clients rather than on the entire population of mobile clients and in a finer time granularity. We monitor the behavior of each wireless user with respect to its association patterns and carry out user-behavior analysis more accurately. We focus on the analysis and modeling of session and visit durations, and apply our methodology on extensive wireless traces.

a. Department of Computer Science, University of North Carolina at Chapel Hill.

b. Department of Statistics & Operations Research, University of North Carolina at Chapel Hill.

c. Department of Computer Science, University of Crete and ICS-FORTH.

We show that as the session mobility increases, the visit duration tends to decrease stochastically. The opposite happens in the case of the session duration. This distinction between visit and session durations gives new insights into the wireless access characteristics. Moreover, there exist different stochastic orders among visit durations of different building types when conditioning on session mobility. The mobile sessions tend to be "imbalanced" with respect to their visit durations. A family of BiPareto distribution can model approximately the visit and session durations. As an example, we parametrized the stationary session duration using the BiPareto distribution. Finally, we propose a new set of metrics to describe a session and characterize its mobility, transient nature, spatial properties, and the disconnection periods. These contributions nicely tie with our earlier results that model the trajectory of the sequence of APs as a markov-chain and predicting with high probability (86%) the AP of the next association of a client [6].

Section II describes briefly the wireless infrastructure and our techniques for acquiring the data. Section III focuses on the session generation and its features for characterizing the mobility and access patterns. Section IV presents our measurements and provides insight about the associations and movement of users on the campus. Section V of this paper describes previous related research. In Section VI, we summarize our main results and discuss future work.

II. Wireless infrastructure

The UNC wireless infrastructure provides coverage for nearly every building in the 729-acre campus and includes a diverse academic environment. The majority (232) of the access points (APs) on campus were configured to send syslog events to a server in our department between 12:00:00 am on February 10, 2003 and 11:59:59 pm on April 27, 2003. During this tracing period, we recorded 8,158,341 syslog events for 7,694 clients, and 222 APs distributed among 75 buildings. A client is a device that communicates with the campus wireless infrastructure and is identified by a unique id based on its anonymized MAC address. In our earlier work [6], we describe in detail about how clients communicate with APs, the events that allow us to log the clients' activities, and the measures taken to ensure privacy.

The campus primarily uses Cisco Aironet 350 802.11 access points (APs) in a VLAN to provide the wireless network service [11]. An AP generates log messages for IEEE802.11 MAC level events, which indicate when a user associates, authenticates, deauthenticates, or disassociates with the AP. The majority of APs on campus were configured to send this data via *syslog* messages to a *syslog* server in our department. The messages sent by the APs are detailed in [7], [6].

A session consists of a sequence of visits (one or more) without any disconnection between these consecutive visits. Each client may have one or more sessions. We will present the session generation process in detail in Section III.

In addition to each AP's unique IP address, we maintain information about the building the AP is located in, its *type*, and its coordinates. The possible building types are academic, administrative, athletic, business, clinical, dining, library, residential, and theatre. We have map coordinates for each corner of most buildings. This allows us to estimate the centroid of each of these buildings. For the buildings that we do not have exact coordinates, we estimate their coordinates by visually inspecting a campus map and calculating the distance between the center of the building and the center of a building with known coordinates, and then scaling that distance according to the map's scale.

III. Session generation

The syslog messages are ordered based on their timestamp (i.e., the time when they are received from the server in our department). Our parser reads these syslog ("info-type") messages [12] sequentially for each client, interprets each event with respect to the Cisco documentation [12], creates some state information for each client, and generates each client's transitions from one AP to another or to its disconnection from the infrastructure. Its main task is to construct the visits and sessions for each client.

A. Session generation process

The parser maintains for each client a *state array* that indicates the *state of the client with respect to each AP* in the infrastructure. This state corresponds to the IEEE802.11 state variables, namely state 1 or "unauthenticated, unassociated", state 2 or "authenticated, unassociated", and state 3 or "authenticated and associated" (page 22, IEEE802.11b, 1999 Edition [13]). We also introduce the state -1 or "undefined" state which is the initialization state for each client's state array. The parser also maintains the *current state* of a client that indicates the AP, if any, with which the client is currently associated. As the parsing of the syslog trace proceeds, the parser updates each client's current state and state array. It also maintains a status information that indicates whether or not these transitions are consistent with the IEEE802.11 state diagram[13]. For example, when we receive a deauthentication event from an AP i and the i-th entry is the state 2, the i-th entry changes to state 1 with a "successful deauthentication" status.

At the beginning of the tracing period, the session generation process assumes that each client has no visits or sessions, and initializes all the entries of their state arrays to the undefined state. The first visit of a session starts with the first association message after a period of disconnection or the start of the tracing period. Each of the remaining visits (if any) in the sequence are triggered by a (re)association event. Essentially, the visits in a session represent the continuous roaming of the client.

The session generator completes a visit when it receives any of the following events, namely, a (re)association, a deauthentication from the *current* AP [1], a disassociation from *any* AP, or reaches the end of the tracing period. The (re)association event (of the list above) extends the sequence of visits, and therefore the session, by one additional visit. Its timestamp indicates the start of the new visit and the completion of the previous visit. All the other events (in that list) complete the current visit and terminate the session (i.e., the current visit becomes the last visit of the session). Their timestamps mark the end of the current visit and session. A period of disconnection for that client follows until the receipt of a new (re)association event from an AP that will start a new visit (i.e., the first visit of a new session) or the end of the trace.

The status of each visit indicates whether or not the event that terminated that visit and initiated a transition (i.e., a new visit in the session or the completion of the session) is a successful transition. To be a successful transition, it must conform to the IEEE802.11 state transition diagram and not introduce any conflict or inconsistency with the states that the session generator maintains for that client. When the session generator parses a disassociation event for a client from an AP different from the current state of that client, it generates an unsuccessful transition. Similar outcome has an association event, if prior to that, the AP that sent it (e.g., j) has state 1 or -1 in the state array of that client. The session generator extends the session by creating a new visit (e.g., at j) and updates the state array accordingly. However, since the session generator is not aware about any (pre)authentication of this client with that AP j, it considers the transition invalid. This could happen when the authentication syslog message gets lost or the (pre)authentication takes place prior to the tracing period start.

The *visit duration* corresponds to the period from the start of the visit until its completion. The *session duration* is the sum of all the durations of the visits in that session. We merge the *consecutive (re)associations* of a client with the *same* access point during a session. Each new *merged visit* has as duration the sum of the duration of all the (re)associations that compose this merged visit. Throughout the following Sections, we refer to a "merged visit" simply as "visit".

B. Conditions for well-defined sessions

We have a large set of sessions (235,885) and would like to focus our analysis on those with the most reliable session information. Unreliable information can be mainly due to syslog packet losses, sporadic AP failures, partial knowledge about the configuration or IEEE802.11 implementation of the AP or client, or events that may have happened before or after the tracing period but we cannot verify.

We decided to select the well-defined sessions and focus our analysis mainly on them. A session is well-defined when it satisfies certain criteria regarding its completion and inter-AP and inter-building transitions. Any not well-defined session or transition is an invalid one. We describe these conditions in the following paragraphs.

1) Completion conditions: A session must have finished before the end of the tracing period. This condition is necessary to compute accurate session durations. The disassociation event that completes a session (of a client) must come from the same AP as the current state of that client prior to this disassociation.

The disconnection must take place with a *disassociated* or *deauthenticated* event with the reason *"Successful"* or *"Sender is Leaving (has left) ESS"*. When a session comes to an end with a *deauthenticated "Inactivity"* event, we consider the session to be not well-defined. Although, most of the APs' inactivity period has been set to 30 min, we found sessions that ended due to inactivity with an invalid duration. Since we analyze the duration of visits and sessions, we decided to filter out these sessions that ended with a deauthentication due to inactivity event.

2) Inter-AP transition condition: In order for a session to satisfy the transition condition, during the time that the session generator forms each transition of the session, the client status and its state array should "reflect" (re)associations that comply to the IEEE802.11 state diagram. As mentioned in Section III-A, the parser maintains some status information for each visit that

[1]When a deauthentication event from a different than the current AP is received, the client state with that AP becomes "deauthenticated" but the event is ignored in the session generation.

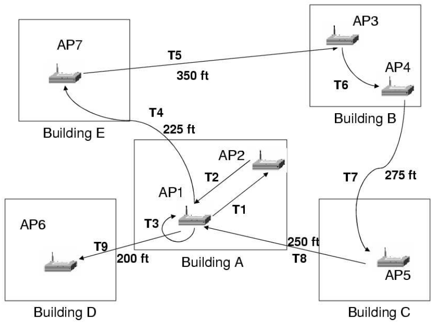


Fig. 1. Example of a session. Transitions are marked T1 through T9 and occur in ascending order. The distances that appear are the euclidean distances of the centroids of the buildings that are involved in the transitions.

indicates whether or not the transition to the next visit satisfies this requirement.

3) Inter-building transition condition: We found some sessions that cover very large inter-building distances (a few were even above 2,000ft). Given the transmission range specifications, these transitions cannot be valid. We speculate that in such case, the syslog server did not receive all the events properly, otherwise, the session generator would have formed either more visits for that session (with shorter inter-building transitions) or even multiple sessions.

With the support of an accurate positioning or tracking system, we could have validated the distances of all inter-building transitions. However, its deployment is costly and would have complicated our experiments in several ways. Given the lack of such mechanism, it becomes difficult and tedious to detect the invalid transitions. Finally, we decided to consider invalid all these sessions that have an interbuilding transition with euclidean distance above 460ft. The threshold was computed based on the transmission range specifications in [14] and our estimations that a typical maximum distance between any point in a building of the campus and its centroid is around 100ft.

C. Session features

We would like to capture how far a wireless client travels during a continuous wireless access. The following metrics provide indications about the mobility of a session. We will use them to classify the sessions and clients and also to investigate he impact of mobility on the session and visit duration.

AP-Path: Each session with at least two visits to different APs has an AP-path. We create the AP-path for each session as follows: Each node in the AP-path corresponds to a visit of the session. There is a one-to-one mapping of the nodes in the AP-path and the visits of the session. Consecutive visits of a session at the same AP have been merged (as mentioned in Section III).

We connect two nodes with an edge, if they correspond to consecutive visits in the session. For example, if a wireless client that was originally disconnected connects to APs 1, 2, 1, 1, 7, 3, 4, 5, 1, and 6 before disconnecting, its AP path is $1 \rightarrow 2 \rightarrow 1 \rightarrow 7 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 1 \rightarrow 6$. The *length* of an AP path is the number of its inter-AP transitions or edges of the AP-path. In this example, the AP-path length is eight. Fig. 1 is a visual representation of this example.

Building-path: For each session, we define the building path to be the sequence of its inter-building transitions. To construct the building-path, first, we find all consecutive visits to APs that belong to the same building and map them to a “building node”. We, then, connect two nodes with an edge, if they correspond to consecutive visits at APs that are located in different buildings in the session. The *building path length* is the number of its inter-building transitions or *building path edges*. In the above example, APs 1 and 2 belong to building A, APs 3 and 4 to building B, AP 5 to building C, AP 6 to building D, and AP 7 to building E. Thus, the building path is $A \rightarrow E \rightarrow B \rightarrow C \rightarrow A \rightarrow D$ and its (building path) length is five.

Longest cycle-free subpath: A cycle in a building path is a return to a building already visited in the session. The longest cycle-free subpath (LCFS) is the cycle-free building subpath in a session with the largest number of inter-building transitions. Ties are broken by choosing the first subpath in the session. In the example above, the LCFS of the session is four and corresponds to the subpath $E \rightarrow B \rightarrow C \rightarrow A \rightarrow D$.

We also have position information for all building in campus.

Range: The range of a path is the maximum pairwise euclidean distance of any two nodes in the building path. The two nodes need *not* be consecutive APs traversed in the path. In our example, the range is the euclidean distance between the centroids of building D and building B.

The range reveals *how far* a client traveled with continuous wireless access. On the other hand, the building path length and ap-path length reflect the total length of the roaming during a continuous wireless access, including potentially visits back to the same AP or at the same building. The larger their values, the higher the mobility tends to be.

D. Session and client classification

Sessions that visit *only one AP* are called *stationary* and have zero length AP-paths. *Mobile sessions* are

Metric	Mean	Median	Max	Min
LCFS	0.02	0	5	0
Unique APs	1.22	1	12	1
Unique bldgs	1.02	1	7	1
AP path length	2.68	0	8052	0
Building path length	0.23	0	248	0
Range(ft)	6.46	0	861	0
Duration(s)	7,395	633	4,147,354	0

TABLE I

STATISTICS ON WELL-DEFINED SESSIONS.

Metric	Mean	Median	Max	Min
LCFS	1.10	1	5	1
Unique APs	2.79	2	12	2
Unique bldgs	2.10	2	7	2
AP path length	21.06	4	1,755	1
Bld path length	11.17	3	248	1
Range (ft)	305.49	303.91	861	79.63
Duration (s)	8,708	1,701	1,613,443	0

TABLE II

STATISTICS ON WELL-DEFINED MOBILE SESSIONS.

the ones that have *at least one* inter-building transition. Sessions without any inter-building transitions can be stationary or ones that only visited APs in a single building. Unless otherwise stated, in this paper, we use the term *session mobility* as the number of inter-building transitions. Fig. 2 shows how the mobile sessions are distributed among clients. The mobile sessions correspond to only a small percentage of the total sessions for most of the clients.

We have classified the wireless clients based on their inter-building mobility, duration at each building, and frequency of their sessions in the trace. Depending on the inter-building mobility, we define as *short-range* clients the clients that always have zero-length building path sessions but may visit multiple buildings on different sessions, as *stationary* clients the subset of the short-range clients that visit APs in only one building throughout the entire trace, and as *mobile* clients the clients that visit two or more different buildings in the same session in at least one of their sessions.

We have 4,115 clients that contributed with 141,653 well-defined sessions. The 24% of them are mobile and the 69% of them have stationary sessions and no mobile sessions. 75% of all clients are short-range clients (i.e., had only sessions, each session with visits in a single building) whereas 20% of them have associated with only one AP throughout the tracing. All the results in our analysis in the following Sections are based on the well-defined sessions unless otherwise stated.

IV. MEASUREMENTS AND ANALYSIS

To explore just how mobile these sessions are, we computed some statistics on the mobility metrics defined in Section II, which are presented in Tbls. I, II and III.

To concentrate on mobile sessions that exhibit higher mobility, we select the ones that are above the 90-th percentile of their range (i.e., range larger than 380ft). There are 585 such sessions generated by 274 clients. Tbl. III summarizes the percentiles of several variables of the movement pattern.

Metric	25%	50%	75%	90%	99%
Visits	4	9	24	61	166.2
Unique APs	2	3	4	5	9
Unique bldgs	2	2	3	3	4
AP path length	3	8	23	60	165.20
Blg path length	2	6	17.25	42	114.95
LCFS	1	1	2	2	3
Duration(s)	480	2,040	5,210	10,460	131,090

TABLE III

PERCENTILE STATISTICS ON WELL-DEFINED SESSIONS WITH RANGE GREATER THAN 380FT.

A. Session duration

Balachandran *et al.*[8] considered a conference setting with four APs and modeled user session durations under minute resolution level. They found that the sessions have durations that can be modeled by a General Pareto distribution with a shape parameter less than 1. The users have a behavioral pattern that revolves around the conference schedule with the longest duration to be three hours. Our current academic setting is completely different from their conference setting. In our study, we categorize the sessions according to user mobility. Our sessions are much longer; about 9.7% of them last more than three hours and the longest duration is about 34 days.

As mentioned in Section III-D, we classify the sessions as stationary and mobile. The mobile sessions can be further divided into those with building path length equal to one ("one-edge") and all the others ("multiple-edge"). Stationary sessions are sessions with

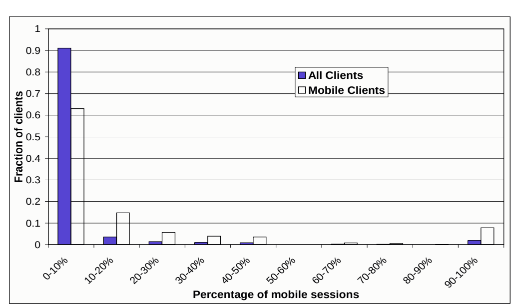
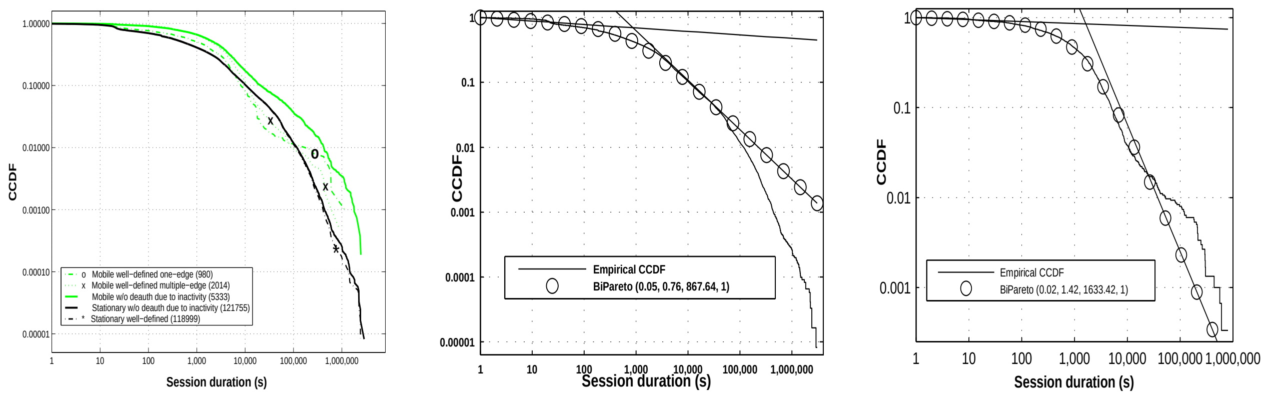


Fig. 2. Fraction of clients with mobile sessions.



(a) (b) (c)

Fig. 3. In (a): stationary vs. mobile session duration and impact of the mobility (number of edges) on the session duration. In (b) and (c): stationary and mobile session duration empirical vs. model, respectively.

zero building-path length. As the building path length increases, the client is considered to be more mobile. We started by first computing the session duration medians and found them to be 9, 18, and 34min, respectively.

Fig. 3(a) presents the log-log plots of the *complementary cumulative distribution function* (CCDF) of the session duration for these three types of well-defined sessions, respectively. The numbers in the brackets indicate how many sessions are used to derive the empirical CCDFs. It shows that there exists a stochastic order among the three types of sessions except in the tails. The CCDF for stationary session duration is uniformly smaller than that for one-edge mobile session duration, which is uniformly smaller than multiple-edge mobile session duration. This means that stationary sessions are *stochastically shorter* than mobile sessions. As the session mobility increases, the session duration increases as well stochastically. A random variable X is *stochastically larger* than another random variable Y if $P(X > t) \geq P(Y > t)$ for every t and $P(X > t) > P(Y > t)$ for some t [15].

Fig. 3(a) also considers all sessions excluding the ones that ended with a deauthentication due to inactivity message (a superset of the well-defined sessions). As in the case of the well-defined sessions, mobile sessions excluding the ones completed with a deauthentication due to inactivity are stochastically longer than the corresponding stationary ones ("Mobile w/o deauth due to inactivity" vs. "Stationary w/o deauth due to inactivity").

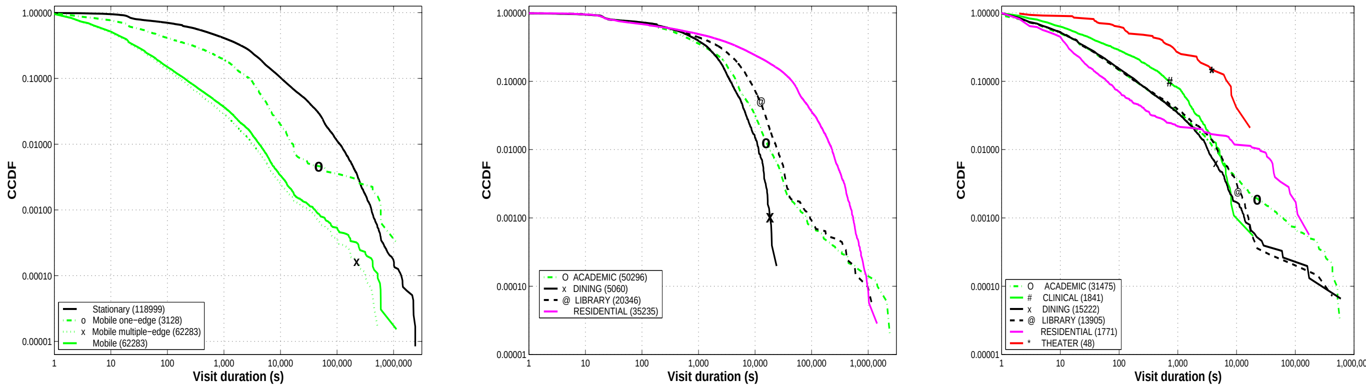
The "hollow" in the well-defined mobile sessions (in the interval $[10^3, 10^4]$, Fig. 3(a)) does not exist in the set of sessions that include all mobile sessions except the ones ended with deauthentication due to inactivity. We found that it is introduced by a group of four clients that exhibit a certain distinct behaviour. They have mobile sessions that violate the valid transitions condition, have very long AP and building paths with (mean,median) equal to (131,30) and (6.5,3), respectively. Also, a very large number of these sessions had visits in three certain dorms.

On these log-log plots, a CCDF of the form $x^{-\alpha}$ would appear as a straight line with slope $-\alpha$. The CCDF of the stationary session duration ("stationary w/o deauth due to inactivity" in Fig. 3) has two nearly linear regimes. After observing this, we propose to model the stationary session duration using a *BiPareto* distribution, whose CCDF is given by

$$\left(\frac{x}{k}\right)^{-\alpha}\left(\frac{x/k+c}{1+c}\right)^{\alpha-\beta}, \quad x \geq k.$$

$k > 0$ is the minimum value of a *BiPareto* random variable, which is a scale parameter. The CCDF initially decays as a power law with exponent $\alpha > 0$. Then, in the vicinity of a breakpoint kc (with $c > 0$), the decay exponent gradually changes to $\beta > 0$. The parameters (α, β, c and k) can be estimated via maximum likelihood. Saniee *et al.* [16] provide more details on the *BiPareto* distribution and its estimation method.

We fitted the BiPareto distribution to the stationary session duration, and the parameters are estimated to be (0.05, 0.76, 867.64, 1) using the maximum likelihood method. Figure 3 (b) plots the empirical log-log CCDF superimposed by the theoretical log-log CCDF of the fitted BiPareto distribution. The two linear regimes are also highlighted. The two CCDF closely follow each other with a coefficient of determination R^2 of 0.99. The major difference appears in the tails, which only concerns 1% of the sessions. One possible explanation for this discrepancy is due to censoring of our data collection period. Because of this, we did not get to observe those stationary sessions that are longer than the collection period. Otherwise, those long session durations will



(a) (b) (c)

Fig. 4. (a): visit duration for stationary vs. mobile sessions. (b) and (c): visit duration for different building types in stationary and mobile sessions, respectively.

bring the tail closer to the BiPareto tail. We also tried several other common parametric distributions such as Lognormal, Weibull and Gamma. The BiPareto gives a much better than the others. The fit becomes even better, if we aggregate the durations into minute resolution level and fitted with a BiPareto with parameters (0.34, 1.37, 258.94, 1).

The fitted parameters for the mobile session durations (0.02, 1.42, 1633.42, 1) (as shown in Fig.3)(c). The log-log CCDFs for mobile sessions also exhibit two linear regions except the tails starting from 3 hours. We propose to truncate the mobile session durations at 3 hours and model them using a truncated BiPareto distribution. The truncation percentage is about 9%.

B. Visit duration

The visits can be grouped into three categories according to the sessions they belong, namely, stationary, one-edge mobile, and multiple-edge mobile. Fig. 4 (a) compares the log-log CCDFs of durations for these three types of visits. In all Figs. 4, the numbers in brackets are the number of visits in each category. As one can see, there again exists a stochastic order among them. As the session mobility increases, the visit duration tends to decrease stochastically. The result is exactly opposite to what one sees in Fig. 3 for the session durations. This stochastic order (as in Fig. 4 (a)) also holds when conditioning on building types.

In stationary sessions, the visit duration matches with the session duration. Therefore, the visit durations of stationary sessions can be modeled by a BiPareto distribution (as discussed in Section IV-A). Based on visual inspection of the log-log CCDF, in plots not shown here, the BiPareto distribution can also be used to model visit durations in stationary sessions for several building types. As for the visits in mobile sessions, a truncated BiPareto distribution seems reasonable after one truncates the visit durations at 3 hours.

C. Visit durations vs. building type

Since visits occur at different buildings, we are interested in finding if there is any relationship between the building type and the visit duration. Fig. 4 (b) shows log-log CCDFs of visit durations in stationary sessions at several types of building. The figure indicates an increasing stochastic order among dinning halls, libraries and dormitories, and also among classrooms, libraries and dormitories (except in the tails, where the variability is high). The orders are also true for stationary session durations. Because stationary sessions and visits in stationary sessions are essentially the same. This stochastic order is consistent with the expected user duration in these environments. For example, residential buildings exhibit longer durations than dining or libraries. In addition, the vast majority of the stationary sessions lasts 1.5 hours or less. As for the visits in mobile sessions, as shown in Fig. 4 (c), visits in classrooms, dinning halls, and libraries are very similar, while there is an increasing order among dormitories, classrooms, (dinning halls, libraries), clinics, and theaters (except in the tails). Users that leave their wireless-enabled laptop on continuously in their office contribute to the very long stationary sessions.

D. Distribution of visit duration in session

Earlier, we modeled the session and visit durations. In this section, we focus on the distribution of the duration across different visits within a session. Are most of the sessions composed of relative short visits? Are the visits well-balanced? Does the first visit differ statistically from the last?

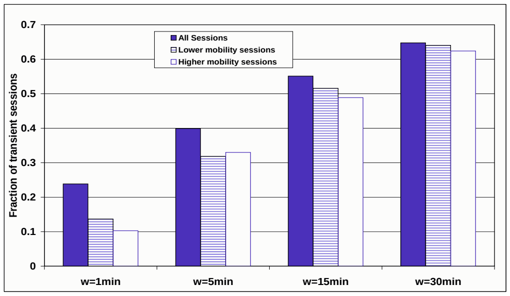


Fig. 5. Fraction of transient sessions (i.e., all visits to a building in the building path last less than w).

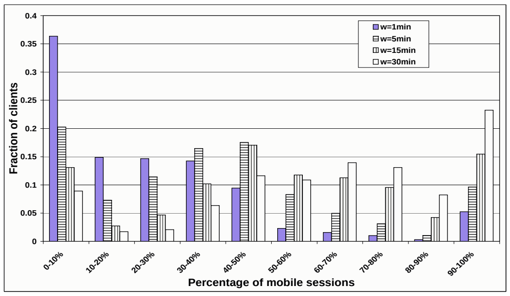


Fig. 6. Fraction of clients that have a certain fraction of their sessions transient.

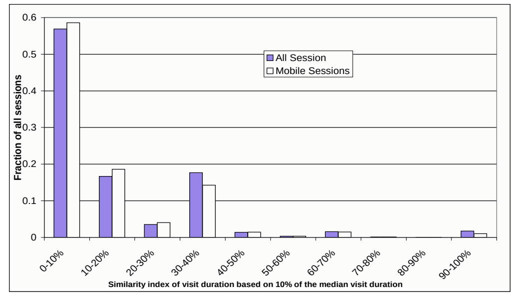


Fig. 7. Percentage of visits in a session with duration within an interval of +/-10% from the median visit of that session.

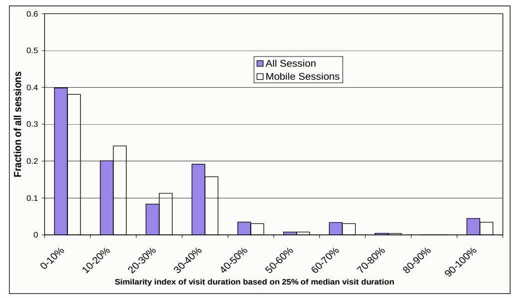


Fig. 8. Percentage of visits in a session with duration within an interval of +/-25% from the median visit of that session.

First, we distinguish the sessions that tend to be composed of relative short durations. Based on the visit duration at each *building* in the building path of a session, we identify the *transient* sessions as the ones that do not have any visits to a building that last more than w min. Fig. 5 illustrates the distribution of transient sessions for different time period w varying from 1min to 30min. We also distinguished the sessions based on the mobility. The lower mobility sessions are the ones with *only one* interbuilding transition. The higher mobility sessions are the ones with two or more interbuilding transitions. As expected, when the threshold increases the fraction of transient sessions also increases. However, it is interesting to observe that for low thresholds (e.g., 1min or 5min), mobile sessions tend to be less transient than the rest. Fig. 6 reveals a clustering of clients based on their percentage of transient sessions. Notice that more than 20% of the clients have a very high percentage (above 90-th percentile) of sessions in which all their visits last 30 min or less.

For those sessions with multiple visits, one interesting question is if the first visit differs statistically from the last visit. Our analysis shows that they look very similar statistically and both of them are stochastically shorter than visits in stationary sessions.

E. Individual durations within a session

Another method of investigating how the session time is distributed among its visits is to compute the percentage of the visits that have duration within an interval of the median visit of that session. We define as the *similarity index* of a session, the percentage of visits that are within a certain interval of their median (such as [0.9*median, 1.1*median], where median is the median duration of the visits in that session). Fig. 7 reveals that sessions are pretty “imbalanced” with respect to the distribution of the visit duration. For example, more than 50% of the sessions have less than 10% of their visits in the 10% interval of the median duration of their session. Figs. 7 and 8 do not include the stationary sessions, since by definition, their similarity index is 100% (since they have only one visit). The difference between the “All Session” and “Mobile Sessions” is all these sessions with visits to two or more APs located at the *same* building. As expected, the larger the threshold is, the larger fraction of sessions with higher similarity index we have.

We have shown that, the more mobile a session is, the longer its duration tends to be, with shorter visit durations. Given the low similarity index presented earlier, such sessions tend to have a small percentage of long visits and a short-visit majority. As a result, a more mobile session is less transient (harder for all visits to fall below a certain threshold). This indicates that all our results are consistent with each other.

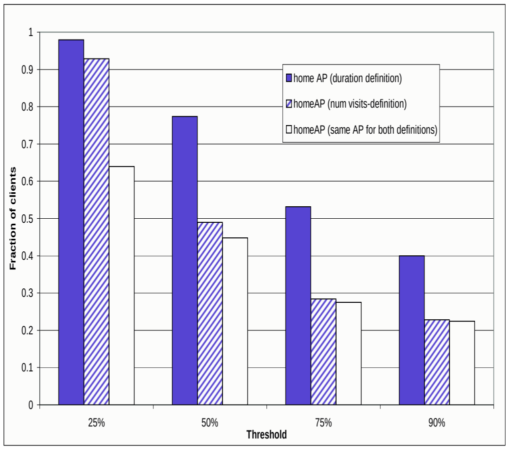


Fig. 9. Fraction of clients that have a home-AP for different thresholds according to the two definitions.

F. Locality properties of clients

Earlier, we looked at the distribution and characteristics of the visit duration in a session. However, we wanted to know in a larger time scale where clients tend to spend most of their wireless time. Also, what are the locality properties of their roaming and which are the hot spots in a campus with respect to associations.

Is there any AP that the client visits more frequently or spends most of its time? We defined the *(duration-based) homeAP* of a client to be the AP (if any) at which this client spends a large percentage of its wireless access time. Similarly, the *(number-of-visits-based) homeAP* of a client is the AP (if any) that this client visits more frequently. We use a threshold for the percentage of wireless access time and number of visits that varies from 25% to 90%.

Fig. 9 shows that the duration-based definition is more relaxed than the frequency-based. More than 50% of clients spend more than 75% of their time to a single AP whereas 30% of them visit more than 75% of the times the same AP. We also notice when a home AP according to the duration-definition exists for a client, it coincides with the home-AP of that client using the number-of-visits definition.

G. Topological characteristics of mobile sessions

What are the spatial properties of the mobile access? Which are the hotspots of the wireless infrastructure? For this, we focus on the inter-building transitions and pairs of buildings that correspond to the starting point and destination of each mobile session.

Figs. 10 show all inter-building transitions of mobile clients throughout the campus. In the leftmost plot, each point corresponds to a building that had at least one inter-building transition and there is an edge between two points, if there is at least one transition between these two buildings. Each edge has also a weight which is the total number of such transitions. The two rightmost graphs of Fig. 10 include only the edges with weights that are above the 75% and 90% of all weights. In Fig. 11 (leftmost), each point corresponds to the geographical location of a building in the campus with an AP. There is an edge between two buildings, when there is at least one mobile session with its first visit starting from the first building and its last visit ending in that (second) building. The number of such sessions gives the weight of the edge. Figs. 11 (middle) and (right) show only the buildings with edges of weights that are above the 75-th and 90-th percentile of all weights.

Figs. 10 and 11, suggest a pattern of transitions of all mobile clients. Each graph shows several separate subgraphs that correspond to clusters of wireless activity in the infrastructure. This result combined with the existence of a home AP for a large percentage of mobile clients reveals a certain spatial locality in the user access pattern. Users tend not to move outside of each of these clusters. In the rightmost map of Fig. 10, nodes without edges correspond to buildings in which many mobile clients started and ended their sessions.

The rightmost plots in Figs. 10 and 11 reveal also the hot spots of the wireless infrastructure with respect to where most popular inter-building transitions take place. Comparing these two graphs, we can identify the various hot spots of our campus wireless infrastructure. Not surprisingly, these hot spots are indeed places that a high student population is visiting frequently (a representative example is the Lenoir Hall with dining center, undergraduate library, classrooms, and stores in close proximity).

These figures summarize the aggregated inter-building behavior of clients. Currently, we analyze similar data and plots combined with additional information, such as, several session features, traffic load, off durations. Using concepts from graph theory, the goal is to capture in more detail how the wireless activity evolves and the different correlations in the temporal and spatial domain. This will allow us to provide a systematic way to indicate not only hot spots but also regions with intermittent connectivity or unusual or unexpected association patterns. It can be useful in providing better capacity planning and improved services, especially in environments with quality of service constraints.

H. PDA users

PDA users possibly correspond to a group of clients with their own distinct characteristics and we were interested in analyzing their access and mobility features. However, we can not identify the PDA devices from the

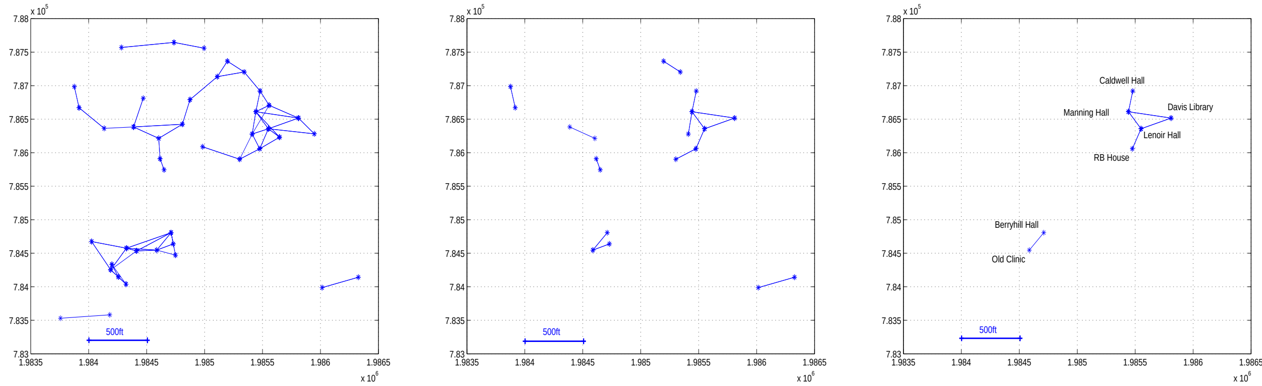


Fig. 10. There is an edge between two buildings, if there is at least one session with such inter-building transition. The middle and rightmost maps include only the edges with number of interbuilding transitions above the 75-th percentile and 90-th percentile of all, respectively.

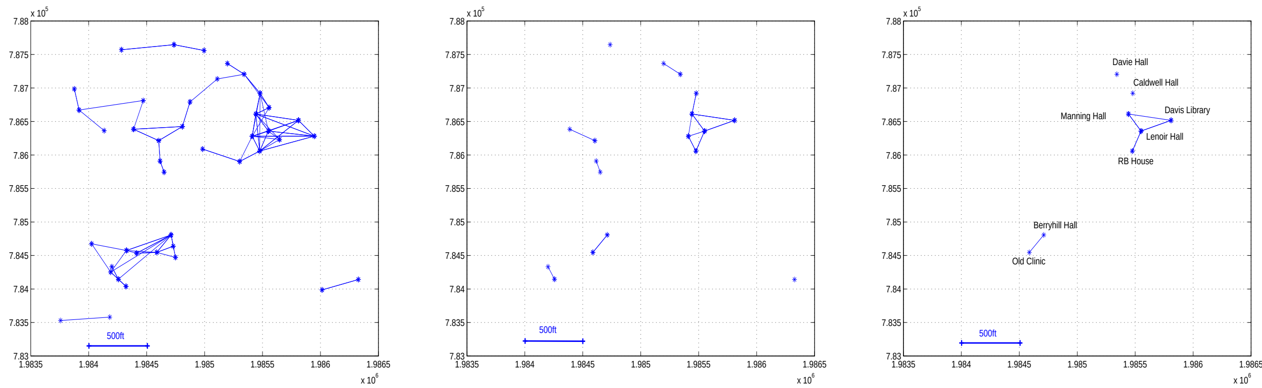


Fig. 11. There is an edge between two buildings, if there is at least one session that starts at one building and finishes at the other. The leftmost map shows all the edges between buildings with at least one session. The middle and rightmost maps include only the buildings with a number of sessions that are above the 75-th percentile and 90-th percentile of all mobile sessions, respectively.

traces. For this reason, we sent an announcement to a UNC-local PDA-related newsgroup to find PDA users that would be willing to share their "anonymized" MAC address. In that way, we could identify the corresponding clients in our syslog traces and analyze their patterns. Some of these users also mentioned their typical usage pattern (e.g., for what purpose they mainly use their PDA) and affiliation at UNC. This is only a very preliminary study, since only seven users responded and we were able to trace the five of them in the trace. There were no well-defined mobile sessions. Most of their mobile sessions violate the inter-building transitions or completion conditions. Three PDA users contribute to seven mobile sessions that accessed 37 different APs located in 16 different buildings. One distinct characteristic of their pattern is the higher mean distance-metrics compared to the mobile sessions contributed by the remaining clients (Tbl. IV). Interestingly, most of these long sessions were contributed by a client in a medical building. [2] This is in-line with our hypothesis that PDA users is a distinct group of clients.

Mean (Median)	*PDAs*	*All clients-PDAs*
LCFS	2 (1)	1.57 (1)
Unique APs	6.28 (4)	3.84 (3)
Unique bldgs	3.14 (2)	2.76 (2)
AP path length	9.85 (11)	24.84 (5)
Bldg path length	2.71 (1)	13.33 (3)
Range (ft)	3,154 (2,118)	910.78 (380.06)
Duration (s)	139,470 (3,227)	23,342 (3,731)

TABLE IV

STATISTICS ON MOBILE SESSIONS OF PDAS VS. ALL CLIENTS EXCLUDING THE ONES WITH PDA.

I. Off duration

We consider all sessions of all clients in the trace. Fig. 12 is a histogram of the off time on the log-10 scale, which shows that many off periods are very short. Actually, 14% of the off durations are no more than 5s. One possible conjecture is that some of these short off periods are due to short-period of loss-signal. We tried to correlate these small off periods with APs to see if there

[2]This study participant had mentioned using the PDA for patient information and prescriptions.

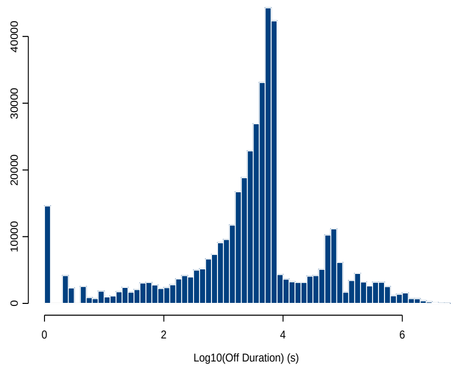


Fig. 12. Histogram of the disconnection periods (i.e., off periods between consecutive sessions of the same client). All sessions are considered.

are certain areas of "intermittent" connectivity. However this is not an easy task with only the syslog information. It becomes more interesting and accurate when we get information from the physical (such as signal strength information) and upper layers (including packet losses and disconnections of the TCP flows).

Actually 20% of these very short off durations belong to only 22 clients (out of the total 5397 clients), who all have at least 70% of their off durations no longer than 5s. This provides a possible clustering for the clients. We decided to exclude the aforementioned 22 clients. In addition, 1037 clients are *visitors* who only have at most 3 off periods, who are excluded as well. Interestingly, the 27% of our clients are visitors.

In the usual log-log CCDF plot of the off durations (not shown here), there are again two nearly linear regions, which suggests one to use a BiPareto fit still. However, the second linear region is almost vertical, which suggests that the off duration has an upper limit. In another word, the off duration is drawn from a censored distribution. The longest off period lasts for 6,013,861s (about 70 days). Due to the data collection period (77 days), there is a physical limit on how long an off duration can be. To uncover the original uncensored distribution, one can guess what proportion of the data is beyond the current maximum and take that into account. Suppose that proportion is 0.0001, the uncensored CCDF differs from the original data visibly only in the tail. This improves the BiPareto fit because the tail is more linear. Currently, we are trying to model the uncensored off duration by a BiPareto distribution.

V. Related work

This research extends our earlier study [6], the studies by Kotz *et al.* [7], [17], Balachandran *et al.* [8], Tang and Baker [9], and Balazinska and Castro [10] by focusing more closely on the association and mobility patterns of individual clients rather than on the entire population of mobile clients and in a finer time granularity, throughout the tracing period and campus region. We monitor the behavior of each wireless user with respect to its association patterns and carry out user-behavior analysis more accurately. Similarly, Balazinska and Castro study the wireless association patterns of users in a corporation considering aggregate information about their associations at every AP. They do provide a clustering of users based on their duration at one AP. However, they only poll the APs at every 5min to get the current association information and do not capture path related information for these associations. In our previous work, we had focused on predicting the AP of the next association for a client without modeling the duration of the wireless access. To the best of our knowledge, this is the only study using wireless traces that tracks the associations of each individual wireless user in such granularity and models the duration of their access and its spatial locality properties.

VI. Conclusions and future work

The main contributions of this paper are the models of visits, session, and disconnection periods. We propose a methodology for visit and session generation and their analysis with extensive data from the wireless infrastructure in a campus. We show that as the session mobility increases, the visit duration tends to decrease stochastically. This stochastic order still holds when conditioning on building types. The opposite happens in the case of the session duration. Moreover, there exist different stochastic orders among the visit durations of different building types when conditioning on the session mobility. We investigated the impact of mobility on the session and visit durations and distinguished several classes of access patterns and clients. The models of the access duration and the next-AP (from our previous work) can be used in simulations to model more accurately the wireless user access. In addition, a wireless client may use these models for smoother handoffs (e.g., enabling a more efficient pre-authentication and pre-association process). A longer-term goal is to extend the study with traffic load information and explore the bandwidth requirements for different classes of clients and access patterns. Administrators can use these models to determine the optimal density of APs and also tune the QoS parameters at each AP. For example, depending on the expected duration and traffic load of the associated clients, an AP may decide about a new association request. This is critical for the support of voice over IP, augmented reality, virtual reality applications or games. Although currently the wireless network is underutilized, one can expect that in the near future, more applications with real-time constraints will use the IEEE802.11 infrastructure to communicate.

This research is a part of a comparative analysis study on wireless access patterns in various environments, such as a medical center, research institute, campus, and a public wireless network. We have obtained traces from the wireless infrastructure of several other universities and public networks. We are in the process of applying our analysis on them. We intend to capture the different features of their access patterns, find the dominant ones, and model them. As new wireless applications and services are deployed, reshaping the wireless arena, it would be interesting to observe and analyze the evolution of the wireless access in the spatial and temporal domain.

ACKNOWLEDGMENT

The authors would like to thank Francisco Chinchilla for his aid in the session generation and measurement analysis, and Dr. Diane Pozefsky for her feedback.

REFERENCES

[1] J. Yoon, M. Liu, and B. Noble, "Random waypoint considered harmful," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, San Francisco,CA, Apr. 2003. [Online]. Available: http://mobility.eecs.umich.edu/papers/infocom03.pdf

[2] "Wireless and mobility extensions to ns-2," http://www.monarch.cs.cmu.edu/cmu-ns.html.

[3] X. Wang and Z. Yang, "Joint mobility tracking and hard handoff in cellular networks via sequential monte carlo filtering," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, New York, New York, June 2002. [Online]. Available: http://www.ieee-infocom.org/2002/papers/149.pdf

[4] T. Liu, P. Bahl, and I. Chlamtac, "Mobility modeling, location tracking, and trajectory prediction in wireless atm networks," *IEEE Journal on Special Areas in Communications, Special Issue on Wireless Access Broadband Networks*, vol. 16, no. 6, pp. 922–936, Aug. 1998.

[5] A. Bhattacharya and S. K. Das, "LeZi-update: an information-theoretic approach to track mobile users in PCS networks," in *Proceedings of the Annual ACM/IEEE International Conference on Mobile Computing and Networking*, Seattle, Washington, USA, Aug. 1999, pp. 1–12.

[6] F. Chinchilla, M. Lindsey, and M. Papadopouli, "Analysis of wireless information locality and association patterns in a campus," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, Hong Kong, Mar. 2004.

[7] D. Kotz and K. Essien, "Analysis of a campus-wide wireless network," Dept. of Computer Science, Dartmouth College, Tech. Rep. TR2002-432, September 2002. [Online]. Available: http://www.cs.dartmouth.edu/reports/abstracts/TR2002-432/

[8] A. Balachandran, G. Voelker, P. Bahl, and V. Rangan, "Characterizing user behavior and network performance in a public wireless lan," in *Proceedings of the ACM Sigmetrics Conference on Measurement and Modeling of Computer Systems*, 2002.

[9] D. Tang and M. Baker, "Analysis of a local-area wireless network," in *ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, Boston, Massachusetts, USA, Aug. 2000, pp. 1–10.

[10] M. Balazinska and P. Castro, "Characterizing mobility and network usage in a corporate wireless local-area network," in *First International Conference on Mobile Systems, Applications, and Services (iMobiSys)*, May 2003. [Online]. Available: http://nms.lcs.mit.edu/˜mbalazin/wireless/wireless-mobisys03.pdf

[11] "Cisco Aironet 350 Access Point Specifications," `http://www.cisco.com/en/US/products/hw/wireless/`.

[12] "Appendix C Event Log Messages," in *Cisco Aironet Access Point Software Configuration Guide*.

[13] *Standard 802.11*, 1999th ed., IEEE/ANSI.

[14] "Cisco aironet 350 series client adapters," in *Cisco Aironet Wireless Lan Client Adapters, Data Sheet*.

[15] G. Casella and R. L. Berger, *Statistical Inference, Second edition*. Duxbury, June 2001.

[16] C. Nuzman, I. Saniee, W. Sweldens, and A. Weiss, "A compound model for tcp connection arrivals for lan and wan applications," vol. 40, pp. 319–337, October 2002.

[17] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campuswide wireless network," in *ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, Philadelphia, Sept. 2004.