On Modeling User Associations in Wireless LAN Traces on University Campuses

Wei-jen Hsu and Ahmed Helmy

Department of Electrical Engineering, University of Southern California
Email: {weijenhs, helmy}@usc.edu

***Abstract*— In this paper we analyze wireless LAN (WLAN) traces collected from four different sources, including three university campus WLANs and one corporate WLAN to compare the similarities and differences of user association behavior to access points (APs) in these environments. This study provides extensive comparison of multiple WLAN traces, and outlines a basis for creating models for user association patterns in WLANs. We propose a set of important metrics for modeling the association patterns of wireless LAN (WLAN) users. Specifically, we look into (a) Activeness of users, (b) Macro-level mobility, (c) Micro-level mobility and (d) Repetitive association patterns. We find that (1) A significant portion of users are offline for non-neligible fraction of time (on average, the online time fraction is between 87.68% and 14.12% for the traces we studied). (2) Users visit only a small subset of APs (on average less than 5%, and the maximum is less than 35%), and (3) Users show periodic pattern of visiting the same APs in some traces. The findings along these aspects show similar trends among the traces, with differences in details due to both underlying user population/environments and methodologies of trace collection.**

I. Introduction

Recently, wireless networks have been deployed ubiquitously in various environments, especially in university campuses and corporations, and gained popularity rapidly. With more users switching to wireless networks, the importance of understanding user behavior in such environments is becoming clearer. From the vast amount of wireless LAN (WLAN) traces available to the research community (e.g. from [1], [2], [3], [12]), one can obtain important and fundamental knowledge about its users. Among the vast space for potential investigation, we focus on the following question: How do we realistically model user[1] behavior and usage in campus WLANs? More specifically, if we are interested in modeling the association patterns of individual users to the access points (APs), what characteristics are important to observe from the traces? And, how do users in different environments differ (or not) on these aspects? We seek to answer these questions by an extensive study of WLAN traces.

In this paper we gain further understanding of realistic user behavior (e.g. usage and mobility) utilizing the most extensive wireless LAN traces collected to date from three university campuses (USC, Dartmouth, UCSD) and one corporate network. Such understanding is important for several reasons: First, analysis of user behavior and network usage patterns enables accurate assessment of wireless network utilization and aids in developing better management techniques and capacity planning decisions. Second, trace analysis is also a necessary first step towards developing realistic models of usage patterns and mobility models that are crucial for the design, simulation and evaluation of wireless networking protocols. Third, as new technologies evolve (e.g. variants of 802.11 WLANs, or ad hoc networks), fundamental understanding of user behavior becomes essential for the successful deployment of such emerging technologies.

Several studies have been previously conducted on analysis of WLAN traces [1], [2], [3], and we borrow from these traces and studies. These studies are quite helpful, but each of them was conducted on a single campus with a different focus, and hence it becomes unclear whether their findings generalize beyond the studied campus. In our study, we go beyond previous work to compare user behavior across different traces, and try to observe the general trends and quantify the detailed differences among them. We look into the aspects that we consider important to model user behavior in WLANs, and reason about the commonalities and differences of these aspects between campuses. For the metrics we study, we find that in general most of the campuses follow similar trends, such as (1) Most nodes display on-off usage pattern. They are turned off for non-negligible amount of time, and switched between online and offline states often. This fact is largely overlooked by previous researches on modeling WLAN users although it is an omnipresent phenomenon from all traces. (2) Most nodes visit only a small portion of the access points (APs) on campus. Therefore, preference in user association is another important aspect to model users of WLANs. The above findings may be intuitive, but it is surprising to observe that the on-off pattern of users change significantly as the popularity of WLAN increases through years, but the ratio of vsited APs hardly changes. In other words, there are varying and invariant user characteristics as one technology gains its popularity. (3) In most cases we identify repetitive patterns in user association over various time frames (e.g., days, weeks). Users re-appear at the same AP it previously associated with with higher probabilities after time gaps of integer multiples of days. We propose *network similarity index (NSI)* as a quantitative metric to capture such repetitive pattern. These findings point to unrealistic assumptions often made in user modeling (for

[1]In this paper we use the terms *user*, *node*, and *mobile node (MN)* interchangeably. We assume that one MAC address in the trace corresponds to a unique device (MN), and a MN is always tied to the same user.

both usage model and mobility model) and simulation, as the findings from traces are significantly different from the general assumptions (e.g., always-on users with no preferences in their association patterns). In addition to that, in the future the findings can be used as guidelines for realistic modeling of user behavior in WLANs.

As one expects, the details of these user behavior metrics depend on the underlying campus environment and user device type; we will comment on the findings throughout the paper. In addition to that, in this work we also compare two different trace collection methodologies, polling-based (e.g. SNMP) and event-based (e.g. syslog). We show the difference between these two trace collection methods by generating an *emulated* SNMP trace in post-processing from higher resolution syslog trace, and compare the differences among the two traces. Sometimes, major differences can be attributed to different trace collection methods used. This brings to being the need for a standard methodology for trace collection to make data from different environments comparable.

The major contributions of this paper are the following: First, by using WLAN traces from four different sources, comparing the results and highlighting both similarities and differences, it is the largest scale trace-based study in the literature as we are aware of. Although some of the findings in the paper match with simple intuition of user behaviors, by extensive investigation we are able to further quantify and show the minor differences in detail systematically, and reason about the cause of those differences (e.g. methodology of trace collection, user population, network environment, time of trace collection, etc.) Second, by proposing metrics for describing individual MN behaviors, we propose a basis on which AP association models for individual MNs can be established. We also find several facts that indicate that conventional, randomly generated synthetic mobility models (such as random waypoint, random walk, etc.) are not adequate for a heterogeneous environment such as university campuses and corporations.

The paper is organized as follows: In section II we discuss the related works. The studied wireless network environments and trace-collection related issues are discussed in section III. We introduce various metrics to describe individual user behavior in-depth in section IV. Finally, we point out future work directions and conclude the paper in section V.

II. Related work

Influenced by the gaining popularity of wireless LANs in recent years, there is an increasing interest on studying usage of wireless LANs. Several previous works [1],[2],[3] have provided extensive study on wireless network usage statistics and made their traces available to the research community. Our work is built upon these understandings and traces, and we carry out analyses that go beyond the previous work.

With these traces available, more recent research works focus on modeling user behaviors in wireless LANs. In [8] the authors propose models to describe traffic flows generated by WLAN users. In [4] the authors focus on modeling association session lengths, and in [5] the focus is on the arrival patterns of users at APs. These papers focus on one or several aspects of modeling user behavior. In our paper, we propose a general framework which is applied to capture fundamental aspects of user association behaviors in the WLAN traces (e.g. Initiation and termination of association, preferences in association, handoff, and repetitive periodic patterns in association) that can be used to build models for WLAN users. We also compare similarities and differences of the above metrics across different environments. Along the same line of modeling user association behaviors, in [6] the authors propose to cluster APs based on the time of peak user arrivals. They focus more on modeling the changes of associated users at the APs by modeling the arrival and departure processes. We focus more on modeling associations of individual users in this paper. In [9] the authors have a similar goal as ours and propose a mobility model based on association session length distribution and AP preferences. However, there are also other important metrics that are not included, such as user on-off behavior and repetitive patterns. We add these metrics to provide a more complete description for user behaviors in wireless networks. By studying multiple traces from different environments collected at different times, we are able to establish that most traces display similar trends, but the details differ due to difference in user population, environment, time, and methodologies of trace collection.

Although many wireless networking protocols have been developed over the past decade, the majority have been designed independently of the context in which they may be deployed and are usually evaluated using artificial (i.e., synthetic, sometimes unrealistic) models. Refer to [10] and [11] for excellent summary of synthetic mobility models used in the literature. We believe that the design and evaluation of the next generation wireless networks should go hand-in-hand with deep, insightful understanding of the realistic environments in which they will be deployed and used. Although WLANs studied in this paper do not provide raw node mobility models directly, they represent a model that combines coarse-grained (i.e. per-AP granularity) nodal mobility, usage of network, and on-off patterns from university campuses. In that sense, one may envision that all-encompassing models may be built by studying the traces. Understanding of such realistic scenarios sheds lights on sometimes falsely taken assumptions in user modeling, and quantifies the detailed behaviors of users so that future models can incorporate them. For example, MNs in synthetic models are always on and homogeneous in their behavior. Both of these characteristics are not observed in our study of real WLAN traces.

In the work we also try to observe periodical association behavior of users in WLAN traces. This investigation is along the same line with the findings by Chen at el. in [7]. In their work, the authors seek to identify the repetitive sequences of user associations and find that weekly patterns are the most prominent ones. In our investigation, we search for the time gap after which a user is more likely to re-appear at the same

TABLE I

STATISTICS OF STUDIED TRACES

Trace source	Unique users	Unique APs	Unique buildings	Trace duration	User type	Environment	Trace collection method	Analyzed part in this paper	Users in analyzed part	Labels used in graphs
MIT[1]	1,366	173	3	Jul. 20 ’02 to Aug. 17 ’02	Generic	3 Engineer buildings	Polling	Whole trace	1,366	MIT-cons MIT-rel
Dartmouth[3]	10,296	623	188	Apr. ’01 to Jun. ’04	Generic	Whole campus	Event-based	Jul. 2003	2,518	Dart-03
								Apr. 2004	5,582	Dart-04 Dart-rel Dart-cons
					PDA only				25	Dart-PDA
					VoIP device				63	Dart-VoIP
UCSD[2]	275	518	N/A	Sep. 22 ’02 to Dec. 8 ’02	PDA only	Whole campus	Polling	Sep. 22 ’02 to Oct. 21 ’02	275	UCSD
USC[12]	4,548	79 ports	73	Dec 03-Now (trap) Apr 20 05-Now (detail)	Generic	Whole campus	Event-based	Apr. 20, ’05 to May. 19 ’05	4,528	USC

AP. Under this definition, we find that users are more likely to re-appear at the same locations they visited before at integer multiples of days, and the strongest tendency of re-appearance occurs at 1 day, followed by 7 days.

III. TARGET ENVIRONMENT AND TRACE COLLECTION METHODS

In this study we mainly focus on wireless traces collected from university campus and corporation networks. We obtain wireless traces from four different sources, including totally over 12,000 distinct users and over 1,300 APs. To our best knowledge this is the most extensive study of user behavior in wireless networks so far. Among the traces, the USC trace is collected specifically for the purpose of our studies, while Dartmouth [3], UCSD [2], and MIT [1] traces were collected by other research groups. We summarize the important characteristics of these traces in Table I and explain the major issues below.

These four traces are chosen to represent different environments. We study the differences and similarities of user behavior in these traces, and try to attribute them to the underlying differences in the corresponding environments as appropriate. In order to make the results comparable between traces, we only analyze selected one-month periods from the longer Dartmouth and UCSD traces. For the UCSD trace, we use the first month from it, as the user activity decreased during their study due to loss of interest in participation and some minor problems in trace collection[2]. We select two one-month periods from the longest Dartmouth trace: July 2003 (*Dart-03*, during the summer vacation) and April 2004 (*Dart-04*, during the spring quarter). All these traces, except UCSD trace, collect measurements of generic wireless network users, including but not limited to laptops, PDAs, and VoIP devices [3]. UCSD trace is from a specific study about PDA users. To further compare the association behaviors of smaller, handheld devices (e.g., PDA, VoIP devices) to generic wireless devices in the same environment, we also separate the PDA (*Dart-PDA*) and VoIP device (*Dart-VoIP*) users from Dartmouth trace during April 2004, and study their behavior specifically. However, according to the device type information provided in Dartmouth trace archive, there are only 25 PDA users and 63 VoIP device users during this time period. The results we get from these small sample sizes may need to be verified by studies in larger scale. All the traces, except the MIT trace, are collected from the entire campus wireless network. The MIT trace is collected from three engineering buildings in a corporation network, hence its user population is not as diverse as the other traces, and the geographic scope of trace collection is smaller. USC trace is the only one that has coarser, per switch port location granularity, while the others have per AP location granularity. Each switch port in USC trace has several APs connected to it. The geographic coverage of a switch port approximately corresponds to a building on campus.

The methods of collecting wireless network traces can be categorized into two major categories: (i) Polling-based methods which record the association of MNs at periodic time intervals, using SNMP (MIT trace[1]) or association tracking software on the MNs (UCSD trace[2]), and (ii) Event-based methods which record MN online/offline events using logging server (e.g. syslog) [3], [12]. For Dartmouth trace we use the derived association history trace from their syslog trace, and for USC trace the logs are collected from the switch. It is generally accepted that event-based approach provides more accurate records of MN behavior in the network. However, there is no in-depth study to quantify the differences between these two approaches. In order to further understand the effects of different methods of trace collection on the user behavior metrics obtained, we also create an *emulated polling trace* as follows: For an event-based trace, we observe the trace at regular time intervals and emulate what would be recorded if the trace were taken by polling-based method. We then process the emulated polling trace as we do to a normal polling-based trace, and compare the findings with the original event-based trace. We use April 2004 Dartmouth trace (*Dart-04*) to carry out this experiment, obtaining *Dart-cons* and *Dart-rel* traces based on the conservative and relaxed assumptions detailed below.

For traces collected using polling-based approach, we obtain only ”sample points” of MN association at regular time intervals in the trace, and duration of association must be derived. Here, an important assumption is made about the association duration for each observed association sample. We test two different assumptions in this respect: (a) A conservative (*MIT-cons, Dart-cons*) approach, in which a MN is assumed to remain associated with the AP only until the next expected polling (recording) epoch, unless indicated otherwise by new samples in the trace. This approach reflects what is observed from the trace faithfully, but may have the drawback that inaccuracy in polling intervals or lost SNMP records (since SNMP uses UDP as transport layer protocol) will lead to the conclusion that the MN is switching between online and offline status while it has been always on. (b) A more relaxed approach (*MIT-rel, Dart-rel*), in which a MN is assumed associated with the AP for four polling intervals after it is observed with the AP, unless indicated otherwise by the trace. This approach is more robust to imperfections (e.g. packet losses) in the trace collection process, however, it may erroneously increase the duration of association with APs after a MN is in fact offline. The polling interval for MIT trace is 5 minutes, and we use the same polling interval to obtain the samples for emulated polling traces from Dartmouth traces (described in the last paragraph). Hence the conservative and relaxed approaches assume a MN remains associated with an AP for 5 and 20 minutes, given a sample indicating the MN is with the AP, respectively. For UCSD the polling interval is 20 seconds. We use only the relaxed approach to UCSD trace.

IV. Analysis of individual user behavior

In this section we propose metrics to describe and compare behaviors of individual users in the studied environments. These metrics correspond to different aspects of MN association behaviors in a WLAN. We shall use Fig. 1 to illustrate. One could see the association pattern of a MN as a sequence of associated APs (shown by shades in Fig. 1), potentially with time segments during which the MN is *offline* (e.g. not associated with any AP) between associations. We look into four major categories to understand user behavior as follows:
(a) *Activeness of users*: This category captures the tendency of a user to be online (i.e. How actively the user shows up in WLAN). In general wireless network users are not always on, but show up in the trace intermittently, as opposed to always-on nodes assumed in the synthetic models. (i.e. What proportion of time is "shaded" in Fig. 1?)
(b) *Macro-level mobility of users*: This category captures how widely a MN moves in the network in the long run (i.e., for the whole trace duration), and how its online time is distributed among the APs. The intention is to capture overall long-run statistics and preference of a MN visiting APs. (i.e. How are the shades distributed in Fig. 1? Do we need many different intensities of shades for each user as it associates with many APs? Can we find few "dominant" APs for each MN?)
(c) *Micro-level mobility of users*: This category captures how MNs move in the network while it remains associated with some AP (i.e. handoff). The intention here is to capture the mobility of a MN while *using* the wireless network, a different objective from macro-level mobility. (i.e. How often does the MN change associations without leaving the network?)

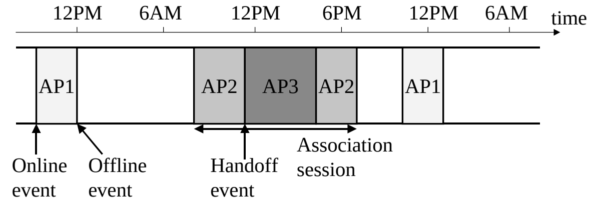


Fig. 1. Illustration of a MN's association pattern with respect to time of the day. Note the on-off, handoff, and repetitive pattern of association.

(d) *Repetitive association pattern of users*: This category captures the user association behavior with respect to time of the day. We expect users to show repetitive structure in association patterns during similar times of different days, as their mobility patterns are dictated by their daily schedule. This idea is shown in Fig. 1: The user appears at AP1 during late evenings in subsequent days. We propose the *network similarity index (NSI)* to quantitize repetitive pattern in association.

Before presenting the analyses using the above metrics we first introduce some terminologies (refer to Fig. 1). **Online event** is defined as a MN starting a new association to any AP from offline status. **Offline event** is defined as a MN disassociating itself from the current AP and changing to offline state. **Handoff event** is defined as a MN changing its association from one AP to another with no offline time in between. **Association session** is defined as the duration between an online event to the next offline event. There can be many handoff events within one association session. **Total online time** is the sum of the lengths of association sessions (i.e. Sum of the "shaded" intervals in Fig. 1), and **existence time** is the time difference between a MN's first online event and its last offline event in the studied trace. Existence time is a conservative measure of the time duration for which the MN is a potential user of the network. Before a MN first shows up and after it last disassociates, we assume that it is not part of the network.

A. Activeness of the users

Activeness of users is the first aspect we look into in attempt to compare the different traces. Activeness of users can be captured by either total online time fraction of a MN or the number of association sessions generated by a MN. We choose to define the *online time fraction* as the ratio between MN's total online time to its existence time. Note that, following this definition, MNs that associate with the APs for only one session have online time fraction of 1.0. This definition tend to over-estimate user activeness for one-time users.

We plot the CCDF of online time fraction of users in various traces in Fig. 2. From Fig. 2 we observe that in all traces only a small portion of users are always on even though by definition the user activeness is already over-estimated, except for the Dart-04 trace. The average online time fraction is 87.68% for Dart-04 trace, and between 36.44% (Dart-03) and 14.12% (UCSD) for other traces. The standard deviation for online time fraction is large, varying from 0.24 to 0.36

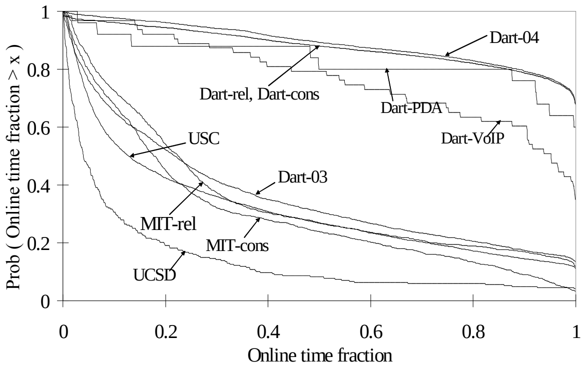


Fig. 2. CCDF of online time fraction

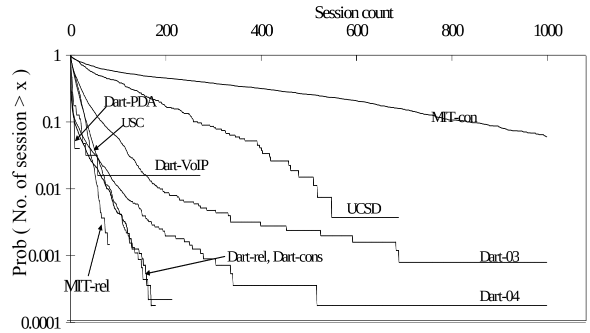


Fig. 3. CCDF of number of association sessions by users

for all traces. These observations argue strongly that **users have on-off usage patterns, where some of the users are heavy users (with high on time) while many are light users**. The distributions of the *on/off* times seem to depend heavily on the environment (i.e., campus). UCSD trace, which focused only on PDA users, is the least active one among all traces. The other traces (MIT, USC, Dart-03) are not very different in online time fraction distribution. The activeness of MNs increase significantly from 2003 to 2004 in Dartmouth trace, which agrees with the findings in [3]. By comparing the curves of Dart-04, Dart-rel, and Dart-cons, we observe that **online time fraction is consistent for the same trace under different trace collection (or trace reconstruction) methods**. Comparison between Dart-04 with Dart-PDA and Dart-VoIP shows that during the same trace period, the handheld devices are less active than the average of the total population. However, handheld devices in Dartmouth trace are much more active than the UCSD trace, but the reason is not clear at this point and warrants further investigation.

We also check whether the significantly higher online time fraction in Dart-04 trace is caused by users with only one short association session (hence its online time fraction over-estimated by our definition). It turns out that the high online time fraction in Dart-04 trace is caused by significant increase of always-on users. In Dart-04 trace, there are 27.5% of users that initiate only one association session which lasts for the duration of 30 days, the whole trace period. The same number for Dart-03 trace is less than 0.04%! There are two possible reasons for the very different behavior in the two time periods. (1) July 2003 was during summer vacation, hence the activity was significantly lower, or (2) The way people use WLAN has been changed between these two trace period at Dartmouth College. Users in Dart-04 trace tend to use wireless LAN as a replacement for wired network, and keep their device associated with WLAN, instead of establishing the connection only when it is needed. If the later speculation is true, as we see this paradigm shift from using WLAN as temporary connection to always-on, permanent connection, it is possible that the online time fraction will also increase significantly for other deployments.

We further compare the CCDF of number of association sessions generated by users in these traces in Fig. 3. We observe that the **PDA users in UCSD trace generate more association sessions than users in other traces** (except MIT-con trace, explained below), which include generic wireless network devices (mainly laptop users) during comparable trace duration. This fact, together with the less online time fraction in Fig. 2, indicates that UCSD PDA users are more likely to use the devices for shorter but more frequent sessions. However, this observation does not apply to handheld devices in Dartmouth. Both PDAs and VoIP devices initiate less sessions than the general devices in Dartmouth. From the figure we also observe that count of association sessions is sensitive to the trace collection method. The emulated polling traces (Dart-rel and Dart-cons) show very different distributions form the original Dart-04 trace, since **traces collected by polling at regular intervals will overlook association sessions shorter than the polling interval**. Comparing the CCDF curve of Dart-04 to Dart-rel or Dart-cons in Fig. 3, we see that the emulated polling traces observe only one fifth of sessions for the MN with most sessions (200 versus 1000). Another technical difficulty here is to adequately translate a record seen in polling-based traces to the duration of association, as we find the curves of MIT-cons and MIT-rel drastically different. A closer investigation into the MIT trace reveals that although SNMP polling intervals are typically 5 minutes, sometimes records of MN association are obtained at longer intervals, leading to bogus terminations and re-initiation of association sessions if the conservative assumption is used and hence the high association session counts shown by curve MIT-cons.

B. Macro-level mobility of users

In this section we capture the long-term mobility of users by obtaining the overall statistics of AP association history during the whole trace period. We investigate the number of APs a user associates with and the fraction of online time it associates with each of the APs. The purpose of this section to understand the preference of MN association at the access point level. Note that the observation could not directly translate to the preference of user visits at geographic level, as APs are not uniformly deployed on the campuses. For example, popular locations on campus may have multiple APs deployed in anticipation of high usage, hence artificially reduce the load observed in the trace. Nevertheless, we can have some idea about how widely a MN visit (in terms of number of visited

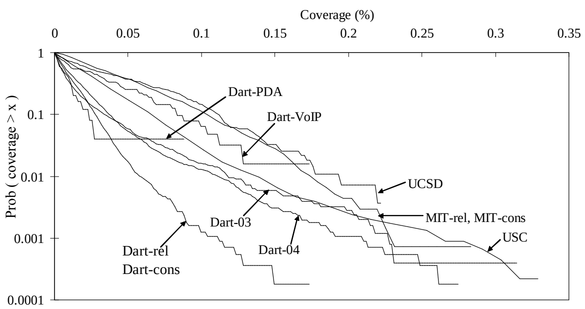


Fig. 4. CCDF of coverage of users.

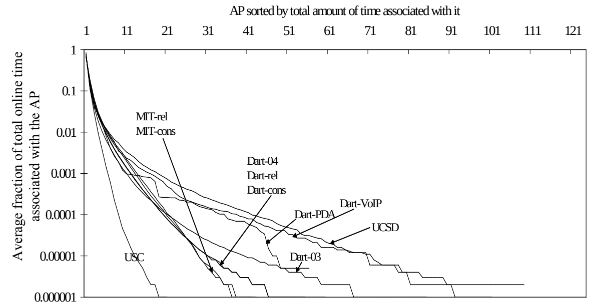


Fig. 5. Average fraction of time a MN associated with APs. For each MN, the AP list is sorted based on association time before taking average.

APs) from this section. If a user visits more APs, and stay at more APs with non-negligible fraction of its online time, it is an indication that the user visits wider range on the campus (i.e. more *mobile* in long run) than another user who visits few APs and spend most of its online time at one or two APs.

We define *coverage* of a user as the *percentage* of APs in the campus it associates with during the trace period. For USC trace we use switch ports in place of APs. The distributions of coverage of users in the traces are shown in Fig. 4. This metric captures how widely a user moves (at AP granularity) for the whole period of trace in the studied network environment.

We observe that **users have small coverage in all environments**. The average coverage is between 4.52% (UCSD) and 1.10% (Dart-cons/rel). None of them have a single user visiting more than 35% of all APs. In UCSD trace, the PDA users seem likely to visit a larger portion of campus than the generic users do in the other campus-wide traces, due to the portability of PDAs. Similar observation applies to the VoIP devices in Dartmouth trace, which is the most mobile sub-user group in the Dartmouth trace. However, PDAs in the Dartmouth trace are less mobile than general users in the period we studied. We suspect that the result may be influenced by a few extreme users (there are only 25 PDA users identified during this period, and half of them visit only 4 or less APs). MIT trace is collected from only three buildings, hence the relative coverage of users is a bit higher. **Coverage seems to remain stable with respect to time change**, although the activeness of users changes significantly (compare Dart-03 and Dart-04). **Coverage is sensitive to the trace collection method** since the polling-based method overlooks short sessions and **under-estimates** the coverage metric. However, different re-construction methods of the polling-based trace (conservative or relaxed approaches) result in the same coverage, as the metric counts the number of APs a MN associates with, not the association duration.

We further study the average percentage of online time a user spends with every AP it visits. We order the APs a user ever visits during the trace period by the user's total association time with each AP, and average across users to get the average percentage of online time a user spends with its most visited AP to least visited AP. These results are shown in Fig. 5. From the figure we observe that for all environments, the general trend is that **each user has very few APs at which it spends most of its online time**. In particular, for all the traces on average a MN spends more than 65% of its online time with *one* AP, and more than 95% of online time at as few as 5 APs. **The left-end of the curves are similar, but the tails vary**. The higher mobility of UCSD PDA users translates into a longer tail, where in addition to those few most visited APs, the users also access the wireless network at much more locations with small time fraction as compared to other traces. Similar observations apply to Dart-VoIP and Dart-PDA traces. It is interesting why Dart-PDA trace shows small coverage in Fig. 4 but high average fraction of time associated with less popular APs here. These two points, however, do not contradict each other. A closer investigation shows that although there are a small fraction of widely-visited PDAs (from Fig. 4), those who visit many APs distribute more of their online time to less popular APs. This metric is **robust** to different trace collection methods and assumptions of trace translation, as the curve for Dart-04 is close to Dart-cons or Dart-rel. Similar observations are made for the MIT trace.

C. Micro-level mobility of users

In this section we study the per-association session mobility of a user, which reflects its short-term mobility. This captures a different dimension of user mobility as compared to the previous section: How mobile the user is while *using* the network. We use handoff statistics as a measure of user mobility while using the network. However, after investigation of handoff statistics, we discover a lot of handoff events are due to ping-pong effect rather than real movements. Ping-pong effect refers to the phenomenon of excessive handoff events due to disturbance in wireless channels while the MN itself is stationary. Hence, we cannot directly link the handoff statistics to micro-level mobility of the users. Development of better filters for ping-pong effects is needed before we can really understand the micro-level mobility form the WLAN traces.

First we show the CCDF curves for the total handoff event count during the whole trace period in Fig. 6. Our first intuition is that user mobility should be dependent on the device type, and handheld devices should display higher mobility than users in other traces. This is true for Dart-VoIP trace, as the VoIP devices have the most per-user handoff count among all traces. However, PDAs in both UCSD and Dartmouth trace do not have more handoff events than other traces. For UCSD trace, this may be related to the fact that PDAs are usually

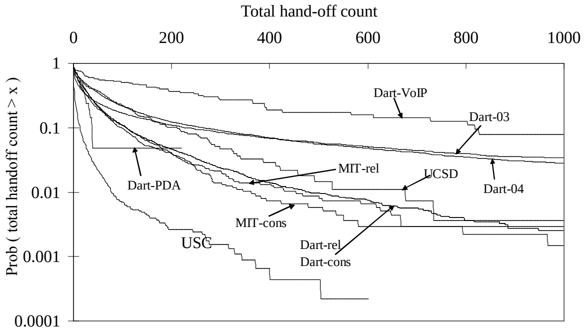


Fig. 6. CCDF of total handoff count per MN

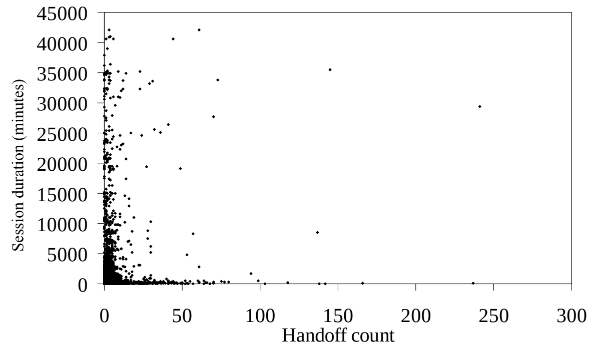


Fig. 7. Scatter plot: Session durations versus handoff count in the session. USC trace.

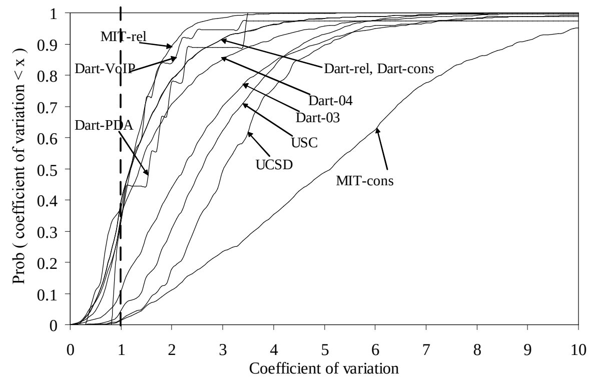


Fig. 8. CDF for coefficient of variation of handoff rate from the traces. Note that for all traces, coefficient of variation is larger than 1.0 for at least 60% of MNs with more than one session.

used for short sessions, hence experience less handoff events. For Dart-PDA trace, some of the PDAs are online for long durations, but they do not have many handoff events. The reason is not clear at this point.

From Fig. 6 we observe that the exact number of handoff count depends heavily on the network environment (e.g., the deployment of the APs, etc). In USC trace, the coarse location granularity directly leads to smaller handoff counts. On the other hand, Dartmouth traces have much more handoff events than other traces. We also observe that handoff counts in Fig. 6 are sensitive to the trace collection method, as curve for Dart-04 differs significantly from Dart-rel and Dart-cons. This is again because the polling-based method overlooks quick changes of association within polling intervals and hence many handoff events are not captured. In addition to the above, we also observe that for all the traces, handoff counts vary significantly among users - There are some users with many handoff events and some with few.

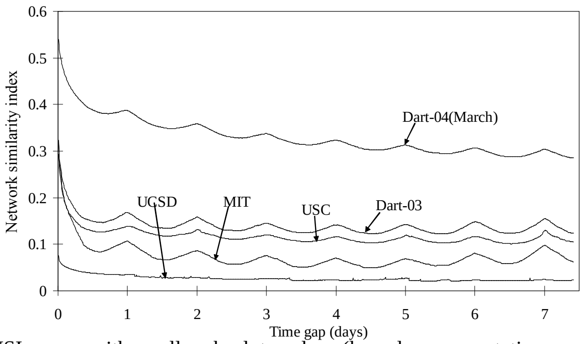
To better understand the cause of handoff events, we look into the relationship between session lengths and handoff events in the session for each trace. As an example, we show a scatter plot for session lengths (in minutes) and handoff counts for all sessions in USC trace in Fig. 7. From the graph, we see that there is no clear trend between session lengths and handoff counts. In some cases, we see extremely long sessions without any handoff events, or extremely many handoff events in a session with short duration. The correlation coefficients between session lengths and handoff counts for all the studied traces are between 0.377 and 0.030. So we can see that session length and handoff count have a weak linear correlation to each other in all traces.

We further look into the following statistics to observe whether the sessions with high handoff counts are all from a small set of extremely mobile users: For each user, we calculate the average handoff event per unit time (i.e. the *handoff rate*) for each of its sessions, and then calculate the mean and variance for its handoff rate from all the sessions the user initiates. If high mobility leading to high handoff count is an intrinsic property for some users, we should see that those users show high average and low variance in the handoff rate. We use the coefficient of variation (standard deviation divided by mean) to understand the degree of variation in the handoff rates for users. In Fig. 8, we show the CDF of the coefficient of variation of handoff rate for the studied traces. Only the users with more than one session and one handoff event are considered in the graph, since users with one session automatically result in 0 variance for its handoff rate. From the figure, we see that the handoff rate displays high variance for most of the users. In all traces, more than 60% of users have its coefficient of variation of handoff rate larger than 1.0 (i.e. Standard deviation being larger than mean). This indicates even for a given MN, handoff rate varies drastically from session to session.

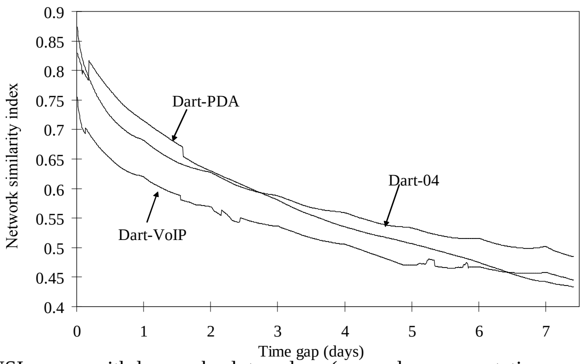
Combining the observations in the preceding paragraphs, we conclude that handoff events not only distribute unevenly between users, but also happen unevenly between the sessions for the same user. This indicates that handoff events are greatly influenced by the environmental condition when a session is established rather than the property of the MN who initiates the session. We even observe that some MNs have hundreds, sometimes even thousands, handoff events between less than 5 APs within a session. Such scenario is much more likely due to ping-pong effect rather than true user mobility. Reduction of ping-pong effect is an important issue to make better interpretation from WLAN traces and warrants further study.

D. Repetitive association pattern of users

Naturally user behavior changes with respect to time of the day and day of the week, as people follow daily and weekly schedules in their lives. In some cases, user association pattern



(a) NSI curves with smaller absolute values (less always-on, stationary users)



(b) NSI curves with larger absolute values (more always-on, stationary users)

Fig. 9. Network similarity indexes. The peaks represent intervals for which there is high similarity.

repeats itself day to day or week to week. In this section we try to quantify such repetitive pattern by defining the *network similarity index (NSI)* below. We try to find the tendency of users displaying periodical association behavior by observing NSI in the traces.

We start the definition with *location similarity index* for individual users. First we take snapshots of associated APs of the user every minute. To study the tendency of the user showing repetitive behavior after a certain time gap (e.g., every 24 hours), we consider all snapshot pairs that are separated by this time gap, and calculate the fraction of all such pairs where the user associates with the same AP in both snapshots. This is an indication of how likely this user re-appears at the same location after the chosen time gap. *Network similarity index (NSI)* for a given time gap is the average of *location similarity index* of all users for this time gap. Hence, NSI represents, *in average sense*, how likely would a node associates with the same AP after the given time gap for the trace under discussion.

In Fig. 9 we show the *NSI* for all the traces. To see the details better, we split the figure into two parts: curves with smaller absolute NSI values are shown in Fig. 9(a), and curves with bigger absolute NSI values are shown in Fig. 9(b). We will discuss the physical meaning of absolute value of NSI later in this section.

From Fig. 9(a), we see that **in most of these traces** (i.e. USC, MIT, Dart-03) **we observe noticeably higher network similarity index if the time gap is close to integer multiples of a day**. This is an indication that **users have the strongest tendency to show repetitive association pattern at the same time of each day**. It is also interesting to observe that for these traces, the **network similarity index for the gap of 7 days (i.e., a week) is the second highest**, only slightly lower than that for the gap of 1 day. This indicates **weekly repetitive pattern is also strong** in these traces. On the other hand, UCSD trace shows little repetitive pattern as there is almost no obvious spikes in its *NSI* curve. This can be attributed to its user population being PDA users. Unlike laptops, which are more related to work, PDAs are usually used in a more casual way in short, scattered durations. Hence it is expected that PDA users show less repetitiveness in their usage pattern.

However, we see that in Fig. 9(b), NSI curves for Dart-04 trace or its sub-groups of users[2] do not show strong patterns of periodicity as discussed above. We suspect that the periodical association behavior in Dart-04 trace is hidden (only minor fluctuation is visible closer to integer multiple of days) due to the increase of always-on users (cf. section IV-A). In Dart-04 trace, we have more always-on, stationary users using WLAN as replacement of wired networks. This is reflected by the higher average value of the *NSI* curves, indicating larger fraction of users always stay at the same location. This may be attributed to the fact that Dartmouth traces include users in student dormitories, which are mainly stationary users and have contributed to high location similarity indexes. We further compare the NSI curve of Dart-04 trace in Fig. 9(b) to the NSI curve of Dart-04-March (only used in this experiment) in Fig. 9(a). For Dartmouth College, the month of March contains the spring break, when some of the stationary users in dorms are absent, and we see that the periodicity of association behavior is more visible in the March trace. From the above experiment, we argue that the periodic behavior in the average NSI curve comes from non-stationary users (e.g., those who come to work or classes during day time and follow a regular schedule), not the stationary users who use WLANs as a replacement of wired LANs. This point is partly supported by the findings in [7]: Most users displaying periodicity in association have home locations at academic buildings. USC has not deployed WLAN in dormitories yet, and MIT trace is mainly focused on buildings for work. That may be the reason why periodic association behaviors are more obvious in those traces. However, more investigation has to be done before we can draw firm conclusion on this. This part is left for future study.

V. Conclusion and future work

In this paper we study the wireless network traces from four different sources collected by various methods, with focus on different environments and user populations. This work extends previous works [1], [3], [2] on analyses of WLAN traces by considering traces from multiple campuses and multiple aspects to model user behavior. We also make our

[2]Curves for Dart-cons and Dart-rel are not shown to make the graph more readable. They are not very different from Dart-04 curve

own WLAN traces collected at USC campus available at [14], together with many pointers to existing WLAN trace archives.

We propose several metrics to describe individual user behavior and capture various aspect of modeling user association pattern in WLANs. The findings along these metrics from the traces point out important common features in all studied traces. Wireless network users in university campuses and corporate network are characterized by (1)Large percentage of offline time. Except for Dart-04 trace, there are less than 20% of users that are always on, and more than 68% of users are offline more than 50% of time. Even in Dart-04 trace, there are more than 30% of users not always on. (2)Limited visited APs in the network and large proportion of online time spent at very few of its most visited APs. The coverage of users never exceeds 40% in all traces, and users spend more than 95% of their online time with as few as 5 APs, and (3)Periodic association patterns with strong daily/weekly pattern. We believe that these metrics capture important characteristics about users in wireless networks that are largely overlooked by earlier work on mobility modeling and wireless network simulation. In this paper we have quantified these characteristics to be integrated in future research of user association or mobility modeling. We plan to work on a more realistic model that uses the findings along the dimensions proposed in section IV to describe users in WLANs. The more realistic model could contribute to more accurate performance evaluation. Also, the statistics obtained using the fore-mentioned metrics can be considered as characteristics or "fingerprints" for particular environments or user population. It should be interesting to develop mechanisms to inspect these "fingerprints" and argue about similarity/dissimilarity between environments.

We also find the detailed distributions of the metrics are different due to the difference in underlying user population and sometimes due to trace collection methods. By comparing the traces collected by event-based logging method and the emulated polling-based traces for the same environment, we find that they sometimes show dissimilar results. Hence, although polling-based trace collection is suitable for usage statistics, they are not very suitable for deriving the association patterns of users, as they tend to overlook details of association changes. Also, we need better heuristics to remove ping-pong effects to make better interpretation about micro-level mobility events (i.e. handoff) from the traces.

Besides individual user behaviors, we also work on the understandings of *encounters* among users, using the same traces. We have some results on this related research topic in [13], and we will also pursue a model that describe the encounter processes between nodes in various environments.

ACKNOWLEDGMENTS

We would like to thank Brian Yamaguchi and Carl Hayter at USC ISD division for helping us collected the WLAN traces on the USC campus for the past two years. We would also like to extend our thanks to Dr. David Kotz, Dr. Tristan Henderson, Dr. Marvin McNett, and Dr. Magdalena Balazinska for helping us obtain the Dartmouth, UCSD, and MIT traces, respectively.

We are also thankful to USC HPCC and Jabed Faruque for providing high quality computing facility. In addition, we would like to thank Fan Bai, Sapon Tanachaiwiwat, and Shaocheng Wang for providing useful feedback on this work.

This work was funded in part through the NSF CAREER Award 0134650.

REFERENCES

[1] M. Balazinska and P. Castro, "Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network," In Proceedings of MobiSys 2003, pp. 303-316, May 2003.

[2] M. McNett and G. Voelker, "Access and mobility of wireless PDA users," ACM SIGMOBILE Mobile Computing and Communications Review, v.7 n.4, October 2003.

[3] T. Henderson, D. Kotz and I. Abyzov, "The Changing Usage of a Mature Campus-wide Wireless Network," in Proceedings of ACM MobiCom 2004, September 2004.

[4] M. Papadopouli, H. Shen, and M. Spanakis, "Characterizing the Duration and Association Patterns of Wireless Access in a Campus," 11th European Wireless Conference 2005, Nicosia, Cyprus, April 10-13, 2005.

[5] M. Papadopouli, H. Shen, M. Spanakis, "Modeling Client Arrivals at Access Points in Wireless Campus-wide Networks," 14th IEEE Workshop on Local and Metropolitan Area Networks, Chania, Crete, Greece, September 18-21, 2005.

[6] M. Kim and D. Kotz, "Modeling Users' Mobility among WiFi Access Points," in Proceedings of the International Workshop on Wireless Traffic Measurements and Modeling (WiTMeMo '05), Seattle, Washington, June 2005.

[7] G. Chen, H. Huang, and M. Kim, "Mining Frequent and Periodic Association Patterns," Dartmouth College Computer Science Technical Report TR2005-550, July 2005.

[8] X. Meng, S. Wong, Y. Yuan, and S. Lu, "Characterizing Flows in Large Wireless Data Networks," in Proceedings of ACM MobiCom, September 2004.

[9] C. Tuduce and T. Gross, "A Mobility Model Based on WLAN Traces and its Validation," in Proceedings of IEEE INFOCOM, March 2005.

[10] T. Camp, J. Boleng, and V. Davies, "A Survey of Mobility Models for Ad Hoc Network Research," Wireless Communication & Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications, vol. 2, no. 5, pp. 483-502, 2002.

[11] F. Bai, A. Helmy, "A Survey of Mobility Modeling and Analysis in Wireles Adhoc Networks", Book Chapter in submission to Kluwer Academic Publishers.

[12] W. Hsu and A. Helmy, "IMPACT: Investigation of Mobile-user Patterns Across University Campuses using WLAN Trace Analysis," unpublished USC technical report. Available at http://nile.usc.edu/MobiLib/Trace_analysis_TR.pdf

[13] W. Hsu and A. Helmy, "On Nodal Encounter Patterns in Wireless LAN Traces," the Second International Workshop On Wireless Network Measurement (WiNMee 2006), April 2006.

[14] MobiLib: Community-wide Library of Mobility and Wireless Networks Measurements (Investigating User Behavior in Wireless Environments). USC WLAN trace and pointers to many WLAN trace archives available at http://nile.usc.edu/MobiLib.