Distribution of the Frequency of Connections in Academic WLAN Networks

Enrica Zola
Department of Telematics
Universitat Politècnica de Catalunya (UPC)
Barcelona, Spain
enrica@entel.upc.edu

Francisco Barcelo-Arroyo
Department of Telematics
Universitat Politècnica de Catalunya (UPC)
Barcelona, Spain
barcelo@entel.upc.edu

***Abstract*—Understanding the trends in the use of wireless local area networks (WLANs) (i.e., how much, when and where traffic is present) is an important issue for modeling the network and for optimizing the allocated resources. Interesting results can be extracted by analyzing traces from real scenarios. In previous works, the authors studied three buildings belonging to two campuses in Barcelona (Spain) and its surroundings. Similar common trends were observed in the three buildings, despite the different amount of users, purpose of the building, geographical location and size of the campus. In this work, the fidelity of the users in accessing the WLAN in different days is analyzed in depth. The population accessing the networks is mostly composed of infrequent users: less than half of the devices access the WLAN more than four days during the studied period. Special insight is given to the underlying distribution. It is shown that, in contrast to previous studies in the same environment, the distribution of the frequency of reconnections to the WLAN is not uniform. The main difference among different buildings is the fidelity of users: users on a small campus are more likely to reappear on different days than on a large campus, where the population is more heterogeneous. The results of this analysis provide general tools for characterizing campus-wide WLAN and a better understanding of usage and performance issues in a mature wireless network in Europe.**

Keywords-WLAN traces; syslog; user behavior; frequency of connections.

I. INTRODUCTION

Among other advantages, wireless communications provide flexibility in the deployment of the network and allow users to move around while connected. Users are increasingly interested in taking advantage of the flexibility of wireless technology, and a boom in its implementation to local area networks has been seen. There are many applications of wireless local area networks (WLAN) to universities, from the classical communications (including email and web browsing) to special e-learning applications. Accordingly, universities have pioneered the development of infrastructures to provide connectivity. An example is the Wireless Andrew at the Carnegie Mellon University campus [1], an enterprise-wide broadband wireless network developed in 1993.

In the last decade, research has been carried out to understand the use of WLAN (typically using standard IEEE 802.11) networks in different scenarios, like campus-wide universities [2]-[9], corporate networks [10], and metropolitan area networks [11]. All these works differ in the way in which data are collected; *tcpdump*, the Simple Network Management Protocol (SNMP) and *syslog* are the most common tools. *Tcpdump* is normally used to sniff the traffic and analyze the applications run by users and the amount of data managed in the network [2], [4], [5], while SNMP is applied to periodically poll the access points (APs) of the network [2], [10] and to obtain information regarding the authenticated/associated users and their approximate location (i.e., the coverage range of the AP to which a user is associated). Information on the authenticated/associated devices at each AP can also be obtained through *syslog* [4], [7], a standard for forwarding log messages in an IP (Internet Protocol) network. As found in [12], although polling-based trace collection is suitable for usage statistics, it is not very suitable for deriving the association patterns of users because they tend to overlook details of association changes due to the polling interval. Differences may also be found in the duration of the period analyzed; some span a period of a week [3], in which case weekly cycles cannot be observed, others one month [10], and others three to five months [2], [4], [5], [7], in which case monthly patterns can also be observed. Each work can also involve different numbers of users. In previous studies, number of users ranged from 74 users in the earliest study [2] to 8 thousand in wider environments [4], [5], [7], [10].

In this work, we process data recently collected at the Technical University of Catalonia (UPC), Spain, from March to May 2009. UPC is composed of eight campuses in Barcelona and its surroundings. Wireless local area network traffic is analyzed at each building separately to deal with homogeneous data and not to mix behaviors from populations with different features. Two libraries were selected. The first one, BRGF, is located on a large campus in Barcelona, while the second one, EDSE, is on a campus in Castelldefels. Results from another building (EETAC) in Castelldefels, a faculty building with classrooms and studying rooms, are also provided. To complete the work presented in [7], this paper analyzes the frequency in the connections to the WLAN. A fitting distribution is found which can be used to describe how often the same user will appear in the network during a given period of time. Differences are found with respect to previous works in a wide university [4], while the pattern is more similar to that found in a corporate network [10].

The paper is organized as follows. Section 2 provides a summary of the related literature. Section 3 provides details of the methodology followed to collect the data from the WLAN and of the environment where the analysis was conducted. Main findings on users' activity are given in Section 4. The frequency of user's connections is examined in Section 5. Section 6 concludes the paper.

II. RELATED WORK

Several studies have been performed to analyze network traffic and user behavior in different WLAN environments. The first study was performed at the Computer Science Department of Stanford University in 1999 [2]. In this work, the authors prepared a testbed with 74 laptops. Users moved freely inside a building and the authors analyzed their movements and traffic during three months. Data were collected using three different techniques (i.e., *tcpdump* traces, SNMP polling and authentication logs). During the day, they observed higher activity in the afternoon, while during the week they found lower activity in the weekends. They found out that most users did not move much within the building; however, a few users were highly mobile. Regarding the total number of days that users are active (present in the network) during the traced period, they found that while some users rarely connect their laptops to the network (17 users do so on 5 days or less), others connect their laptops frequently (14 users are active at least 37 days during the traced period).

The use of the WLAN at the Saskatchewan University Campus was presented in [3]. The campus is composed of 40 buildings covering public spaces (i.e., lounges, libraries, coffee shops, etc.), classrooms, laboratories and offices. A traffic trace was collected during one week in January 2003 using EtherPeek, a software package that allows researchers to record MAC (i.e., Medium Access Control) addresses and traffic load information. MAC addresses were matched with the authentication logs obtained from each of the 18 APs of the campus. In total, 134 unique users connected to the network during the week under study. Individual users visited at most 8 different APs. Regarding the number of authentications, they observed that more than half of the total number of users authenticated more than 50 times during the week. According to the method used for collecting traces, they stated that authentications cannot be literally interpreted as distinct sessions of network usage.

Generalization of the results presented in these works is difficult due to the low number of users observed (74, 134 and 195, respectively). Moreover, the fact that users knew about the tracing study may have perturbed the user behavior in [2], while in [3] authors are aware that no effort was made to ensure that the week they analyzed was representative of overall usage patterns.

Kotz and Essien analyzed the wireless network at Dartmouth College during three months [4] and extracted information about users' mobility, card and building activity, traffic load and protocols. The main results are that network activity shows clear patterns: almost half of the users were active on a typical day, and about one third of those were mobile. The weekly pattern showed a typical student's pattern of activity, with lower activity on Fridays and Saturdays, and a pick up on Sundays. Users varied in the number of days that they used their cards, from only once to every day in the 77-day trace (many users lived on campus and could be always on-line). The distribution of the frequency in the user activity is roughly uniform between one and 77 days, with a median of 28 days. In 2004, they revisited the WLAN [5] and found that, despite a drastic increase in traffic, users were mainly non-mobile.

Similar user patterns were found in a corporate network from July 20 to August 17, 2002 [10]. Despite mobility results report higher mobility than on university campuses, users still spend a large fraction of time at only one location. The results regarding the daily and weekly patterns are similar to those observed on university campuses. The number of days that users connected to the WLAN varies greatly: only 12% to 25% of users are present more than 18 out of the 20 work days, whereas 22% to 38% of users appear only during one or two days (i.e., outside visitors mostly from other sites that the company has in the same metropolitan area). The presence of visitors and the absence of employees were uniformly distributed. In terms of the fraction of days that users access the network, the distribution is similar to a single building on a university campus [2]. Compared with a whole campus [4], more users appeared only one or two days (visitors) and fewer users appeared more than 2/3 of the days. The authors observed that the higher uniformity of a campus wide distribution might be related to the fact that the study tracked many users for prolonged periods of time (i.e., students living on campus) and not only when they went to work in specific buildings.

Mc Nett and Voelker analyzed the mobility patterns of users of wireless handheld PDAs in a campus wireless network using trace belonging to 11 weeks of wireless network activity. They also observed the frequency of connections in their traces: 50% of the users initiated more than 77 sessions over the trace period. This means that the median user initiated an average of one session per day over the trace period. Still, 20% of the users initiate roughly three sessions per day, and 10% initiated roughly four connections per day. To understand user activity on a day to day basis, they analyzed the number of days the users actually turned on their PDAs. Half the users turned on their devices less than 21 days during the trace. This is lower than the median number of days from [4]. Moreover, the distribution is not uniform as observed in [4]; 8% of the users only used their PDAs one day during their trace period. The number of users for each number of active days tends to drop from there. 20% of the users used their PDAs 60% of the total days; 10% used them more than 75% of the total days; and there were a few users who used their PDAs nearly every day.

The trends regarding the frequency of connections observed at the three buildings of the UPC are presented in the following sections. Not only the users' behavior will be related to the environment and compared with previous similar works. A deeper insight will be given to the underlying distribution, in order to model the fidelity of the user in the WLAN network.

III. TRACE COLLECTION AND SCENARIO

This paper analyzes data collected during 3 months from March 2009. Users were not informed that the study was performed. The only sensitive information gathered were the MAC and IP addresses of network cards connected to the network and the name assigned to each AP; these data are not shown.

The data were collected with syslog [14], which is a standard client/server protocol for forwarding log messages in an IP network: the client sends a small textual message to the syslog server through User Datagram Protocol and/or Transport Control Protocol connections. The access points were configured to send syslog messages to a central server whenever they received authentication, deauthentication, association or disassociation IEEE 802.11 messages. Each message contains the AP name, the MAC address of the card, the time stamp at which the AP received the message to 1-second precision, and the type of message. From the MAC address, it is possible to relate a given address to a given device; however, the same user may have multiple cards, or the same device can be used by different persons. In the rest of the paper, we will use the term “device” to refer to a given MAC address.

According to the IEEE 802.11 standard, after authentication, a device chooses the best AP among a list of nearby APs and associates with it. When a device no longer needs to use the network, it disassociates with its current AP. Disassociation can be due to the device moving into another cell (i.e., handover), to authentication problems, or to the device leaving the network. The aim of this study is to analyze users’ frequency of connections to the WLAN and extract the underlying distribution. For this purpose, only IEEE 802.11 association frames are considered. A device can only associate to an AP to which it has authenticated before, so authentication messages can be ignored. In order to catch users’ frequency, the number of days an association response frame is received by a given device is calculated. The same for each device observed in the trace (see last row in Table I)

A. *Library in Barcelona (BRGF)*

The campus in Barcelona houses the faculties of Information and Communication Technologies and of Civil Engineering, with a total of 4,339 students. On this campus, one can also find research centers, the library, student associations, and the UPC Foundation, which is dedicated to lifelong learning and professional retraining. The central library (BRGF) is housed in a five-story building of 6,300 m^2, four above-ground floors and one basement. People access the library from the ground floor, where the loan service is housed. The first and second floors house the library collection, which is divided according to the specific subjects taught on the campus; students spend their time there to work on their subjects and to study. The third floor provides specific documentation for PhD students and researchers. In the basement, there are two studying rooms and two rooms for foreign languages teaching. BRGF provides room for about 700 users and provides 63 desktop PCs and 9 laptops for library users. BRGF is open from 9 am to 9 pm from Monday to Friday, and during the exam period, the opening hours are extended until 2:30 am and during weekends.

The WLAN infrastructure at BRGF is composed of eight APs on the first four floors and four APs in the basement, which provide good coverage all over the building. A total of 5,917 devices associated to the wireless network during the whole trace period. Table I summarizes the main parameters of the three buildings.

B. *Library in Castelldefels (EDSE)*

The Castelldefels campus houses the Castelldefels School of Technology (EETAC), the School of Agricultural Engineering of Barcelona and a library; it is located 30 km from Barcelona. The library is housed in a three-story building. It is open from 8 am to 9 pm from Monday to Friday. Three APs are located on the ground floor (edse002 to edse004) where the loan service is located, four on the first floor (edse101 to edse104) and two on the second floor (edse201 and edse202), where the library collection and the studying rooms are located.

The infrastructure provides good coverage inside the building. In this analysis, the APs in the basement are also considered (edses01 to edses04) because students go to the bar and connect to the WLAN from there as well. A total of 1,419 devices associated to the wireless network during the whole trace period.

C. *Classrooms (EETAC)*

EETAC is a Higher Education School specializing in technical and scientific courses in Aeronautics and Telecommunications. It is housed in the Castelldefels campus. The total number of students attending classes at EETAC is about 1,500 persons. The building houses 20 classrooms, 25 laboratories, professors’ offices and 2 studying rooms, all distributed on 3 floors. It is open from 8 am to 9 pm from Monday to Friday.

The WLAN infrastructure at EETAC is composed of twelve access points, which provide good coverage throughout the building. A total of 1,417 devices associated to the wireless network during the whole trace period.

IV. COMPARISON OF THE ACTIVITY

In our previous study [7], the association pattern was observed during the three-month period. Table I reports the main parameters of the three scenarios. The number of APs providing the WLAN infrastructure is the same in the three buildings, despite the EETAC building is more than three time bigger than EDSE building. The number of students in the Barcelona campus is 3 times that in the Castelldefels campus, while the number of detected devices at BRGF is higher than the number of students due to the proximity to other university (e.g., students from outside the UPC are using the WLAN). Since the Castelldefels campus is far away from other universities, the number of detected devices nearly equals the number of students.

Details on the user behavior at each building are given in [7]. A decrease in the usage of the WLAN was observed at each building during Easter holidays (one week in April), during weekends at the two buildings in Castelldefels, during one week in April at EETAC building due to exam period. In order to deal with homogeneous data, only the working days in March and May are taken into account in the rest of the paper (no holidays, nor weekends). April is skipped due to inconstant activity.

V. FREQUENCY OF CONNECTIONS

The empirical cumulative distribution (CDF) of devices according to the number of days per period that they appear as active is shown in Figure 1. At BRGF, more than 50% of the devices connect just once or twice in two months. This proportion of infrequent users decreases for EDSE (35%) and EETAC (24%). This is explained by the proximity between the library at BRGF and other campuses that cause a higher proportion of occasional visits; in a smaller campus, on the other hand, users tend to re-appear more often. 10% of the users at each building appear: more than 12 days at BRGF; more than 21 days at EDSE; and more than 24 days at EETAC. Again, more similar trends are observed in the Castelldefels campus. The number of very frequent users (i.e., those who appear more than 30 days over the whole period) is 1% at BRGF and 4% at both buildings in Castelldefels. The building housing classrooms (EETAC) displays the highest fidelity among the three buildings (lower number of infrequent users, and higher number of very frequent users).

TABLE I. PARAMETERS OF THE THREE SCENARIOS.

	BRGF	EDSE	EETAC
Nº of APs	12	13	12
Sq. meters	6,300	3,000	10,000
Students in the campus	4,439	1,524	1,524 (same as EDSE)
Detected devices (whole period)	5917	1417	1419

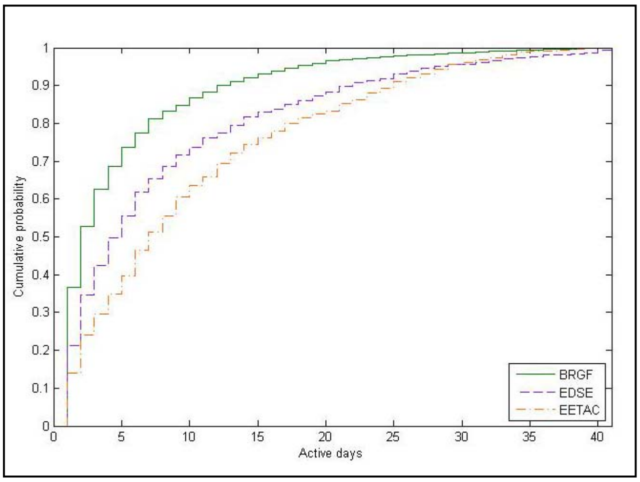

Figure 1. Fraction of active devices per period.

Table II resumes the trends observed in previous similar works. Results in [4] and [6] show very different behaviors if compared with our trace, despite all refer to traces taken from university campuses. In these other works, the percentage of users who connect once or twice during the whole period is lower than in our traces; half of the users connect more often; 10% of the users connect more than 91% of the days in [4] and more than 75% in [6], which are higher than in our trace. The distribution in [4] is roughly uniform between one and 77 days. On the other hand, the figure observed in a corporate network [10] is more similar to the one observed in our trace. EDSE can be compared to the small building (SBldg) in [10], while EETAC to the large building (LBldg). This similarity with a corporate building is due to the fact that, at UPC, students do not spend the night inside the campus (which is instead typical in the campuses in the US) but go home at night and return the next day they have classes; this behavior is similar to that of a worker. BRGF shows quite different trends, with the highest percentage of infrequent users (connected up to 2 days) and the lowest percentage of constant users.

The probability mass function of devices that are active a given number of days is shown in black in Figures 2 to 4 for BRGF, EDSE and EETAC building, respectively. Three theoretical distributions have been proposed and tested through the Chi-Square goodness of fit (gof) test: Zipf, geometric and negative binomial. The latter is not represented in the figures since it gives the worst matching. Moments' matching has been applied to estimate the parameters for each distribution, as shown in Table III. Again the lowest mean belongs to the BRGF due to its proximity to other similar campuses.

TABLE II. MOMENT ESTIMATORS FOR EACH CANDIDATE DISTRIBUTION.

Ref.	Environment	Buildings	% of users up to 2 days:	50% of users connect up to [% of total days]:	10% of the users connect more than [% of total days]:
This work	Campus	BRGF EDSE EETAC	53 35 24	5 10 15	31 51 59
[4]	Campus		10	36	91
[6]	Campus		11	25	75
[10]	Corporate network	Large Medium Small	24 22 38	12 27 30	67 63 60

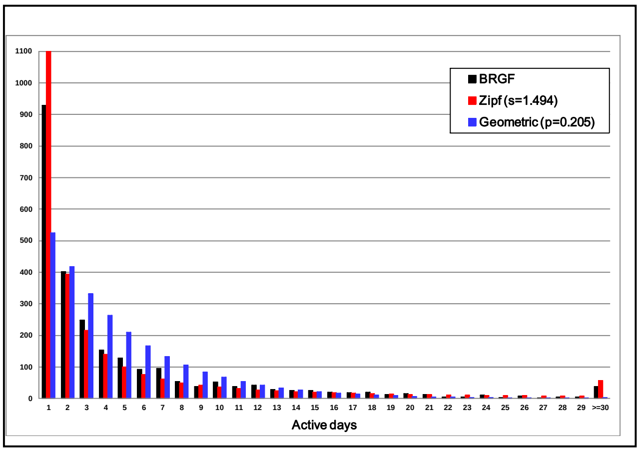

Figure 2. Probability mass function of the active devices for a given number of days (BRGF).

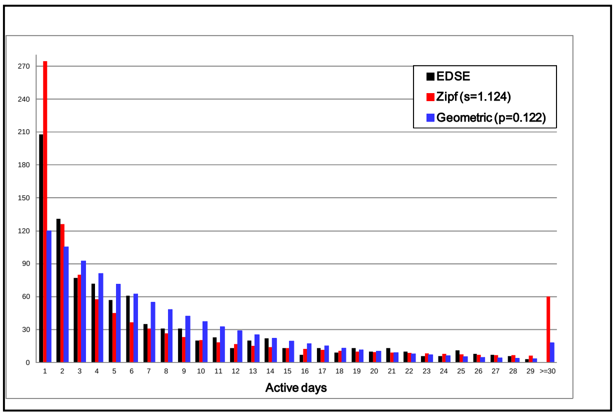

Figure 3. Probability mass function of the active devices for a given number of days (EDSE).

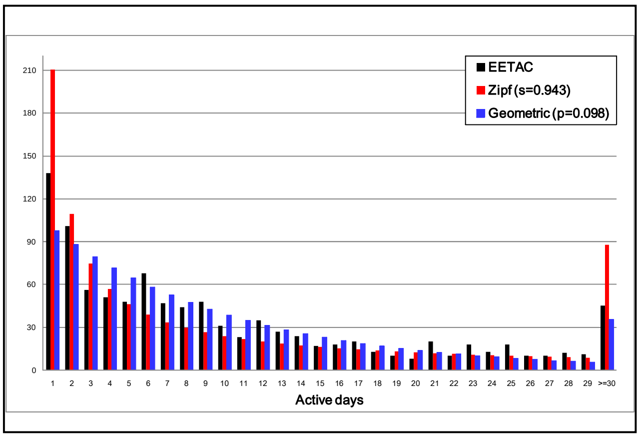

Figure 4. Probability mass function of the active devices for a given number of days (EETAC).

The Chi-Square gof test with 5% significance was performed for each building and distribution. Given the null hypothesis that the fitting distribution is F, this test compares the number O_i of observed elements of the empirical distribution in category C_i (for i = 1,2,…,k), with the expected number E_i of elements of F in category C_i. As a measure of comparison the test uses [15]:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}. \qquad (1)$$

χ^2 is known as the Chi-Square test statistic: if it is greater or equal to the critical value χ^2_α, then the null hypothesis that the fitting distribution is F must be rejected at level α. The critical value depends on the level of significance α and on the sample size n (see Table III), and can be computed as:

$$\chi^2_\alpha = n\left\{1 - \frac{2}{9n} + z_\alpha \sqrt{\frac{2}{9n}}\right\}^3, \qquad (1)$$

where z_α can be found in [15]. In this paper, we use $z_{0.05}$ = 1.6449. The test statistic χ^2 and the critical value χ^2_α are displayed in Table IV.

For all the buildings, the Zipf distribution provides a good fit. According to the Zipf law's, the frequency of active days is inversely proportional to its rank in the frequency table. Thus the most frequent number of active days will occur approximately twice as often as the second most frequent number of active days, three times as often as the third most frequent number of active days, etc. For the EETAC building, the best fit is provided by the geometric distribution. Despite the similarities shown in Table II, the distribution for the two buildings at the Castelldefels campus is not the same. Instead, the same distribution characterizes the frequency in the connection to the WLAN in the libraries (BRGF and EDSE), while a different behavior is shown in a building housing classes.

TABLE III. MOMENT ESTIMATORS FOR EACH CANDIDATE DISTRIBUTION.

Distribution	BRGF	EDSE	EETAC
Sample size	2524	980	994
Mean	4.88	8.18	10.18
std	6.27	9.10	9.12
Zipf	1.494	1.124	0.943
Geometric	0.205	0.122	0.098
Neg. binomial	p=0.876	p=0.901	p=0.878
	r=0.692	r=0.897	r=1.420

TABLE IV. CHI-SQUARE TEST STATISTIC (X^2) FOR EACH THEORETICAL DISTRIBUTION.

Distribution	BRGF	EDSE	EETAC
Zipf	**119.29**	**65.64**	146.00
Geometric	980.19	160.02	**91.39**
Neg. binomial	3092.56	4102.36	2035.40
Critical value (χ^2_α)	2641.99	1053.94	1068.46

 ISBN: 978-1-61208-186-1

VI. Conclusions

The frequency of the connections at three different buildings in two different campuses of the UPC in Barcelona (Spain) has been investigated. It has been shown that the trends are more similar to those observed in a corporate network [10] than to those reported from other universities in the USA [4][6]. This is mainly due to the fact that the students in Barcelona do not live inside the campus, so their pattern is more similar to that of workers. The distribution of the active days has been analyzed through the Chi-Square goodness of fit test: the same distribution (Zipf) can characterize the behavior at the libraries of both campuses, while the geometric distribution better fit the behavior in the building with classrooms. The results presented in this paper provide general tools for characterizing campus-wide WLAN and a better understanding of usage and performance issues in a mature wireless network in Europe. These findings may be useful both for those researchers interested in simulations under realistic scenarios and for optimal planning of a WLAN infrastructure in similar environments.

Acknowledgment

The authors would like to thank UPCnet for providing WLAN traces. This work was supported by the Spanish Government and ERDF through CICYT projects TEC2009-08198.

References

[1] A. Hills, “Wireless Andrew [mobile computing for university campus],” IEEE Spectrum, vol. 36, no. 6, June 1999, pp. 49-53, doi: http://dx.doi.org/10.1109/6.769269.

[2] D. Tang, and M. Baker, “Analysis of a Local-Area Wireless Network,” Proc. of the 6th Annual International Conference on Mobile Computing and Networking (MobiCom’00), ACM, New York, August 6-11, 2000, pp. 1-10, doi: http://dx.doi.org/10.1145/345910.345912.

[3] D. Schwab, and R. Bunt, “Characterising the Use of a Campus Wireless Network,” Proc. of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societes (Infocom’04), March 7-11, 2004, pp. 862-870, vol. 2, doi: http://dx.doi.org/10.1109/INFCOM.2004.1356974.

[4] D. Kotz, and K. Essien, “Analysis of a Campus-Wide Wireless Network,” Wireless Networks, vol. 11, no. 1-2, January 2005, pp. 115-133, doi: http://dx.doi.org/10.1007/s11276-004-4750-0.

[5] T. Henderson, D. Kotz, and I. Abyzov, “The Changing Usage of a Mature Campus-wide Wireless Network. Computer Networks,” The International Journal of Computer and Telecommunications Networking, vol. 52, no. 14, October 2008, pp. 2690-2712, doi: http://dx.doi.org/10.1016/j.comnet.2008.05.003.

[6] M. McNett, and G.M. Voelker, “Access and Mobility of Wireless PDA Users,” ACM Sigmobile Mobile Computing and Communications Review, vol. 9, no. 2, April 2004, pp. 40-55, doi: http://dx.doi.org/10.1145/1072989.1072995.

[7] E. Zola, and F. Barcelo-Arroyo, “A Comparative Analysis of the User Behavior in Academic WiFi Networks,” Proc. of the 6th Performance Monitoring, Measurement and Evaluation of Heterogeneous Wireless and Wired Networks Workshop (ACM PM2HW2N 2011), October 31-November 4, 2011.

[8] E. Zola, F. Barcelo-Arroyo, and M. Lopez-Ramirez, “User Behaviour in a WLAN Campus: a Real Case Study,” Proc. of the third Ercim Workshop on eMobility, May 27-28, 2009, pp. 67-77.

[9] R. Hutchins, and E.W. Zegura, “Measurements of a Campus Wireless Network,” Proc. of the IEEE International Conference on Communications (ICC’02), April 28 – May 2, 2002, pp. 3161-3167, vol. 5, doi: http://dx.doi.org/10.1109/ICC.2002.997419.

[10] M. Balazinska, and P. Castro, “Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network,” Proc. of the 1st International Conference on Mobile Systems. Applications and Services (MobiSys’03), ACM, New York, May 5-8, 2003, pp. 303-316, doi: http://dx.doi.org/10.1145/1066116.1066127.

[11] D. Tang, and M. Baker, “Analysis of a Metropolitan-Area Wireless Network,” Wireless Networks, vol. 8, no. 2-3, March-May 2002, pp. 107-120, doi: http://dx.doi.org/10.1023/A:1013739407600.

[12] W.J. Hsu, and A. Helmy, “On Modeling User Associations in Wireless LAN Traces on University Campuses,” Proc. of the 4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, April 3-6, 2006, pp. 1-9, doi: http://dx.doi.org/10.1109/WIOPT.2006.1666494.

[13] A. Balachandran, G.M. Voelker, P. Bahl, and P.V. Rangan, “Characterizing User Behavior and Network Performance in a Public Wireless LAN,” Proc. of the International Conference on Measurements and Modeling of Computer Systems (ACM SIGMETRICS’02), June 15-19, 2002, pp. 195-205. doi: http://dx.doi.org/10.1145/511334.511359.

[14] C. Lonvick, “The BSD Syslog Protocol,” RFC 3164 (rfc3164). IETF. 2001.

[15] A.O. Allen, “Probability, Statistics, and Queueing Theory with Computer Science Applications,” Academic Press Professional, Inc., San Diego, CA, USA, 1990, isbn: 0-12-051051-0.