Modeling the Roaming in Large-scale Wireless Networks using Real Measurements

Maria Papadopouli[a,b]   Michael Moudatsos[b]   Merkourios Karaliopoulos[a]

Abstract— Campus wireless LANs (WLANs) are complex systems with hundreds of access points (APs) and thousands of users. To analyze the performance of wireless networking protocols, researchers need to construct simulations and testbed experiments that reproduce the characteristics of these networks. However, the generation of realistic models and benchmarks is challenging and there is only a limited set of models of roaming and access based on real measurement data. The main contribution of this paper is the modeling of the roaming activity. Specifically, we employed graph theory, modelled the roaming activity as a graph and measured its properties (e.g., diameter, degree of connectivity, connected components). For example, the negative binomial distribution models well the degree of connectivity. Furthermore, we analyzed the evolution of the roaming activity in the spatial and temporal domain and its impact on the properties of the graph.

I. INTRODUCTION

Wireless networks are increasingly being deployed and the demand for wireless access grows rapidly. It becomes particularly intriguing to study the temporal and spatial evolution of wireless networks. Unlike the wired networks that are relatively fixed, the wireless networks are highly dynamic due to the radio propagation effects and user mobility. Clients change access points (APs) due to their signal quality or mobility. Wireless LANs have more vulnerabilities and stricter bandwidth, and latency constraints than their wired counterparts. While in several cases overprovisioning in wired networks is acceptable, it can become problematic in the wireless domain due to interference, environmental, regulatory and cost reasons. Current wireless networks cannot efficiently support real-time multimedia applications and thus, mechanisms, such as capacity planning, resource reservation, device adaptation, and load balancing, need to be employed to improve their quality of service provisioning. For the design and evaluation of these mechanisms, the characterization of the roaming activity and inter-AP connectivity is critical.

Currently, there is lack of publicly available models for the topology of large-scale wireless infrastructures and their access patterns. Such models are essential for performing meaningful simulation and performance analysis studies. Contrary to traditional wired-network topologies that reflect the physical hardwired connection of routers or AS, wireless topologies are more dynamic and have a strong stochastic element due to radio propagation, user mobility and the AP association mechanisms. Graph theory has been used extensively in modeling the topology of wired networks and performance analysis of routing and flow control algorithms. We aim to characterize the roaming in a wireless infrastructure, identify the regions with high roaming activity, and derive topological models of the infrastructure. Unlike our earlier research efforts that focused on roaming per client basis [1], [2], [3], this paper investigates it in an aggregate level. We employ graph theory to model the roaming of wireless clients during a time interval as a graph. The main contribution of this paper is methodological; It models the roaming activity as a graph and analyses its properties, such as the degree of connectivity, diameter, and number of connected components. The negative binomial distribution models well the degree of connectivity. We discuss the impact of the spatial and temporal growth of the wireless infrastructure on the graph and its properties. Also, we discuss the non-linear correlation between the number of roamings between two APs and their geographic distance.

The paper is structured as follows. In Section II, we briefly describe the wireless network infrastructure, data acquisition, and testbed. Section III models the wireless network connectivity as a graph and studies its properties in the spatial and temporal dimension. Section IV presents the related work and Section V summarizes the main results and future work plans.

II. LARGE-SCALE WIRELESS NETWORK TESTBED AND DATA ACQUISITION

The IEEE 802.11 infrastructure at the University of North Carolina at Chapel Hill provides coverage for

Week	Tracing Period	Clients	Total APs
1	17-24, October 2004	8880	459
2	2-9, March 2005	9049	532
3	13-20, April 2005	9881	574

TABLE I

NUMBER OF CLIENTS AND APS DURING THE TRACING PERIODS.

the 729-acre campus and a number of off-campus administrative offices. The university has 26,000 students, 3,000 faculty members, and 9,000 staff members. Undergraduate students (16,000) are required to own laptops, which are generally capable of using the campus wireless network via clients identified by their MAC addresses. The network APs belong to three different series of the Cisco Aironet platform: the state-of-the-art 1200 Series (269 APs), the widely deployed 350 Series (188 APs) and the older 340 Series (31 APs). The 1200s and 350s come with Cisco IOS, while the 340s run VxWorks.

The campus APs were configured to send syslog messages to a syslog server in our department. An AP generates syslog messages for IEEE 802.11 MAC events, indicating when a user associates or disassociates, authenticates or deauthenticates with an AP, or roams from and to another AP [4]. In our earlier work [3], we describe in detail how clients communicate with APs, the events that allow us to log the clients' activities, and the measures taken to ensure users' privacy while acquiring and processing the traces. The results reported in this paper rely heavily on the syslog messages, as it is discussed in Section III. In addition to each AP's unique IP address, we maintain information about the buildings the APs are located in and their coordinates.

We acquire and analyze wireless data from three different monitoring periods covering the interval from October 2004 to April 2005. Table I describes the evolution of the wireless infrastructure across each period. The increase of APs and WLAN clients is significant not only between the first and second tracing period, but also within the month separating the second from the third tracing period. Such analysis allows tracking the temporal evolution of the wireless network and drawing hints for time-persistent features.

III. MODELING WIRELESS ROAMING ACROSS AN INFRASTRUCTURE

We model the roaming of clients within the campus wireless network during a tracing period T as a graph $G_T = (V_T, E_T)$. Each AP deployed in the infrastructure corresponds to a *node* of the graph. We create an edge from node i to node j, if at least one *client transition* from AP i to AP j was recorded during the tracing period
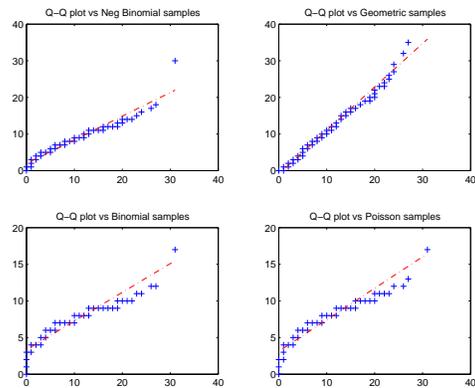

Fig. 1. Q-Q plots of AP outdegree against four discrete distributions (week 1)

T. There is a transition from AP i to AP j, if there is a roaming syslog message from AP i and a reassociation message from AP j, without any disassociation message from AP j at the same time. All these syslog messages for building a transition should come from the *same* client. These transitions are based on the syslog messages generated by APs upon IEEE 802.11 MAC-level events as described in Section II. The *weight* of an edge indicates the total number of transitions between the corresponding APs and its *distance* the geographic distance of the buildings where the corresponding APs are located. To construct the graph, we consider all clients' transitions. We call these graphs *roaming graphs*.

A transition occurs due to either actual user mobility or changes in signal strength that affect the association. Typically, a client selects to associate with the AP from which it receives stronger signal. When the signal strength drops below a threshold, the client may scan the channels and select the AP from which it receives the strongest signal. Wireless networks are highly dynamic environments and clients may experience several rapid "oscillations" between APs. Such oscillations due to radio propagation are very transient and do not reflect typical roaming activity. To distinguish them, we define the *wiggling*, as the case in which a client that was associated with an AP, gets associated *briefly* (i.e., for less than one second) with another AP, and then moves back to the first one.

A. Degree of connectivity

Figure 1 illustrates the qq-plots of the measured outdegrees for week 1 against large samples drawn from four different discrete distributions, parameterized so that their mean is equal to the measured outdegree mean. The quantile-quantile (q-q) plot is a graphical technique for determining if two data sets come from
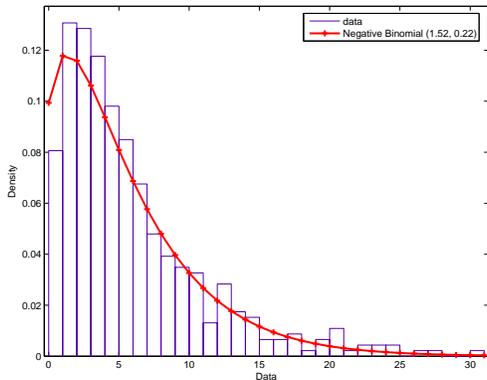
Fig. 2. Histogram of data and Negative Binomial theoretical mass function

Edges	Week 1	Week 2	Week 3
Incoming	1.516, 0.217	1.782, 0.26	1.83, 0.25
Outgoing	1.582, 0.225	1.797, 0.261	1.737, 0.239

TABLE II

ML ESTIMATES FOR THE NEGATIVE BINOMIAL PARAMETERS.

populations with a common distribution. A q-q plot is a plot of the quantiles of the first data set against the quantiles of the second data set. The graphical tests suggest that both the Geometric and the Negative Binomial distributions are possible matches for the degree of connectivity distribution, with the linear regression yielding correlation coeffiecient values in the order of 0.95-0.98 for all tracing periods and indegree and outdegree of connectivity. Nevertheless, the hypothesis that the degree of connectivity follows the Geometric distribution is rejected by the chi-square test, even at the 10% level of significance. On the contrary, the Negative Binomial distribution passes the test at 1% significance level for all traces, revealing a time-persistent feature in the infrastructure connectivity. Notably, the specific hypothesis passes also the Kolmogorov-Smirnov statistic [5] for discrete data, which normally yields more extreme values from the Pearson chi-square statistic [6]. The histogram of the degree of connectivity is plotted against the theoretical distribution with Maximum Likelihood (ML) estimates for its parameters r and p, in Figure 2, whereas Table II reports the ML estimates for the Negative Binomial distribution parameters for the three different tracing periods.

Power laws appear in many natural networks from the human respiratory system to social networks of human driving cars, or human crowds with bluetooth devices. Furthermore, the degree of connectivity of citation and hypertext networks follows a power law. Faloutsos *et al.*
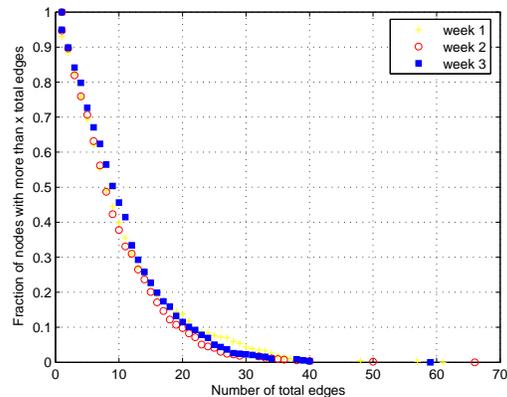

Fig. 4. Degree of connectivity.

[7] argue in favor of a power-law relationship between the degree of connectivity and its frequency for the Internet topology at the network domain connectivity level. Determining the existence of similar relationships in our campus WLAN topology at the AP connectivity level, yielded clearly less strong indications for the existence of such a relationship. Compared to the Internet topology, campus WLAN exhibits a flatter connectivity structure. In the log-log scale diagrams of Figure 3, the correlation coefficient resulting from the linear regression ranges from 0.88 to 0.91. A rule of thumb for reliable power-law relationship inference is that the linearity in the log-log diagrams should span at least 3 orders of magnitude, which is not our case [8].

Figure 4 shows the CCDF of the degree of connectivity for the three tracing periods (namely, weeks 1, 2, and 3). The stochastic order for nodes with small or medium degree of connectivity is different from the one for nodes with high degree of connectivity over the three weeks. Specifically, while the last week has larger percentages of nodes with small or medium degrees, the first week has the largest percentage of nodes with high degree. Let us define as *crosspoints* the nodes with a degree of connectivity greater than 18 and focus more on the distribution of these nodes. A crosspoint corresponds to an AP that received (or sent) a number of roaming clients from (to) a relative large number of neighboring APs. We will refer to that AP as crosspoint AP. Essentially, crosspoint APs identify regions with high roaming activity. In week 1, the percentage of crosspoints (14%) is greater than that of weeks 2 and 3 (13% and 10%, respectively). There is a 40% increase in the percentage of crosspoints from week 2 to week 3.

How does the infrastructure evolve and what is its impact on the properties of the graph? Where are the new APs being placed? New APs were added, aiming
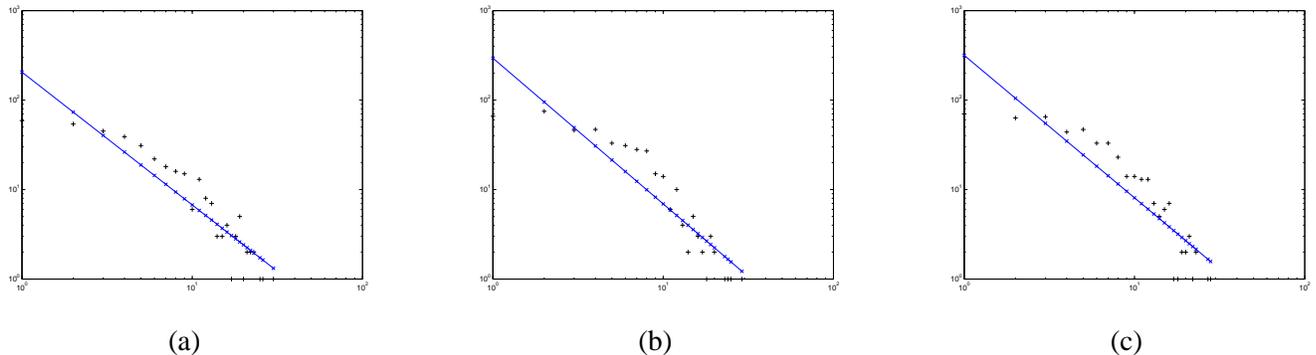
Fig. 3. Log-log plots of the outdegree frequency vs the outdegree: (a) week 1; (b) week 2; (c) week 3.

to extend or enhance the wireless coverage by offering better signal strength to more clients. There was a 16% and 8% increase in the number of APs during the second and third week, respectively. A newly added AP could be placed in either an isolated area or close to other APs. Isolated APs contribute with nodes with a small degree of connectivity. Table III summarizes the growth of the infrastructure in terms of number of APs with high roaming activity. The "Common to prior week" indicates the number of APs that were crosspoints in both current and previous tracing period, "Common in all weeks" the number of APs which were crosspoints in all three weeks, and "Common in any week" the number of crosspoints of the current week which were also crosspoint in any of the other two weeks. The total number of crosspoints falls slightly (9%) from week 1 to week 2, before it rises in week 3. This fall can be caused either due to reduced user mobility or to the placement of the new APs. Since the 20% of the crosspoints of the week 2 and 30% of week 3 were newly added APs with respect to prior weeks, we speculate that they were placed in popular areas with extensive roaming activity.

The placement of new APs in proximity to other APs can only partially alleviate the wigglings, since they may introduce new ones. This was apparent in our infrastructure. Specifically, the 56% of the APs that are common in all three tracing periods experience an increase in the number of wigglings (median rise of 2) whereas the remaining APs had a decrease (median fall of 4.5 including those with fall equal to 0, from week 1 to week 2). From week 2 to week 3, a 21% of APs had an increase in their number of wigglings by a median rise of 5 and the remaining had a median fall of 3.

A related issue is the existence of preferential attachment phenomena in the topology. To analyze this phenomena, we focus on the newly added nodes in the graph from one tracing period to the next. A topology is characterized by preferential attachment when newly
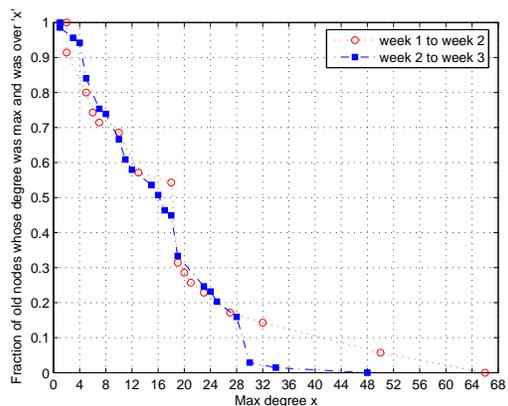

Fig. 5. Preferential attachment CCDF

added nodes are more likely to be "attached" to higher connected nodes than a lower one. There has been a debate regarding the use of the preferential attachment in the generation of the Internet topology [9], [10]. It is unclear whether or not wireless infrastructures evolve according to such processes, since depending on the placement, coverage, and channel assignment, the placement of new APs in proximity of APs with high roaming activity may either enhance the coverage or create interference. To determine whether new APs are placed close to crosspoints, we define the *preferential attachment probability* $P(d)$ as the probability the *maximum* degree of all old nodes which are connected to the new one to be d. As Figure 5 shows there is a preferential attachment phenomena in the infrastructure. Considering the transition from week 1 to 2, 55% of the newly added nodes are connected to *at least one* old crosspoint and this percentage becomes 45% for the next transition period (from week 2 to week3). Figure 6 illustrates the graph evolution through the three tracing periods and its spatial characteristics. Each point in the plot represents all APs located in the building (with the same coordinates as the point). Consequently, edges

Week	Total nodes	Crosspoints	Common to prior week	Common in all weeks	Common in any week
1	459	71	0	51	55
2	532	65	52	51	62
3	574	91	61	51	64

TABLE III

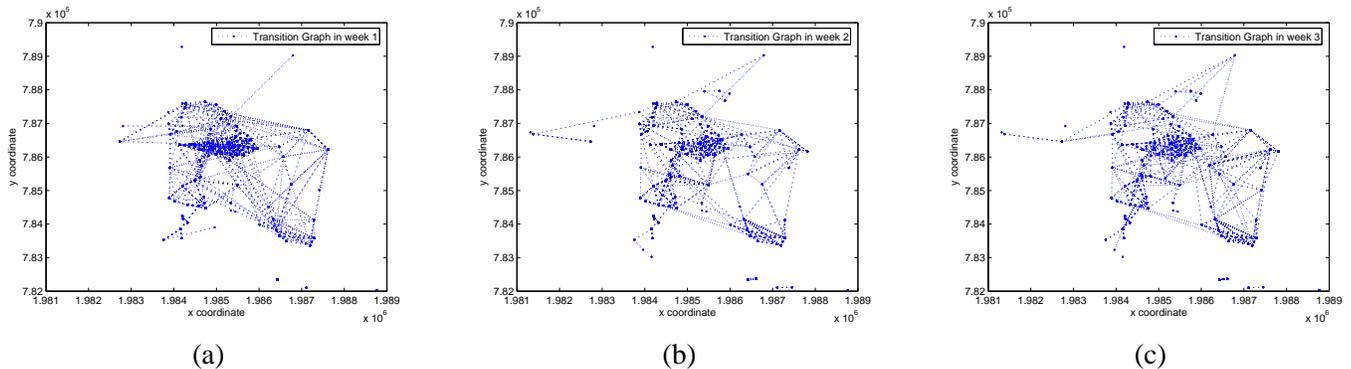EVOLUTION OF CROSSPOINTS IN THE GRAPH DURING THE TRACING PERIODS.

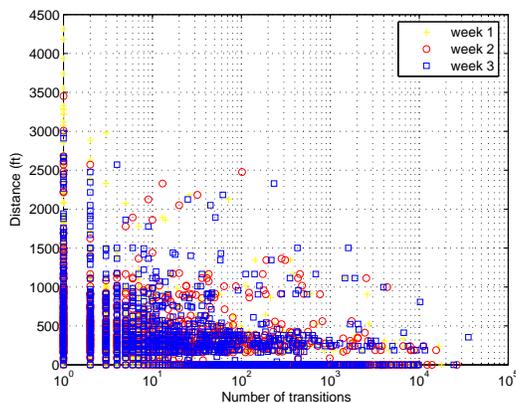Fig. 6.   Transition graph: (a) week 1; (b) week 2; (c) week 3.

Fig. 7.   Number of transitions at each distance

between APs in the same building or edges between APs in different buildings are overlapping with each other and are not distinguishable in that plot.

B. Degree of connectivity vs. APs distance

We expected to find a relation between the number of transitions between two APs and their distance. Figure 7 depicts the scatterplot of the number of transitions between AP pairs (i.e., weight of an edge) versus the Euclidean distance between the respective APs (i.e., distance of an edge). In this analysis, we only consider the pairs of APs with at least one transition. The negative slope that is prevalent in the scatterplot suggests a negative association between the two variable, i.e., as the distance between two APs increases, the number of transitions decreases. The Pearson product-moment

Week	Num. of connected components	Edge diameter
1	13	11
2	24	14
3	22	12

TABLE IV

GRAPH DIAMETER AND NUMBER OF CONNECTED COMPONENTS.

correlation coefficient R ranges from -0.04 to -0.09 for the data sets corresponding to the three different tracing periods. The low values of R, although they are still high enough to reject the null hypothesis of no dependence between the two data sets at 1% significance level, imply in the same time that the correlation under discussion has non-linear characteristics. We computed the Spearman rank correlation coefficient [11], a non-parametric statistic that does not make any assumption about the underlying distribution of the analyzed data. The values ranged from -0.29 (week 3) to -0.45 (week 1) and in all cases the null hypothesis that the two variables are independent is rejected at significance levels far beyond 1%. Further supporting evidence for the existence of negative correlation between the distance of two APs and the transitions measured from the one to the other, came from a third test, the test of independence in contingency tables [12]. The computed values of the statistics are four (week 2) to seven (week 1) times the critical values at the 1% significance level, rejecting strongly the null hypothesis for independence of the two data sets.

C. Diameter and connected components

The diameter of a graph is the maximal of the shortest paths between any pair of its nodes. In general, the number of connected components expresses the connectivity of a graph. Two nodes belong to different connected components if there is no path between them. While in the Internet topology, the existence of connected components might have little or no meaning, in roaming graphs, they reflect the spatial density and coverage of APs. The number of connected components and diameter are crude indications of the robustness of roaming (e.g., the likelihood of roaming without disconnection). Table III-B indicates that the size of the diameter does not change over time. This, with the increase in the number of connected components, implies that, apart from enhancing a building's coverage, new APs were placed to serve buildings that were not covered by any AP in the past. Furthermore, we can distinguish a giant component in all three weeks. A giant component is a subgraph containing more than 50% of the total nodes of the graph. In our graph 61%, 54% and 60% of all nodes belong to the giant component in the week 1, 2 and 3, respectively. In all three weeks, the number of buildings in this giant component was the same, which also explains why the diameter remained constant.

IV. RELATED WORK

There is rich bibliography concerning natural network and wired network topologies. In [9], Albert and Barabasi review the advances in the field of complex networks, focusing on the statistical mechanics of network topology and dynamics. After reviewing the empirical data that motivated the recent interest in networks, they discuss the main models and analytical tools, covering random graphs, small-world and scale-free networks, the emerging theory of evolving networks, and the interplay between topology and the networks robustness against failures and attacks. In [7], Faloutsos et al. prove that the AS and router level internet topology can be described with power laws. In [10], Lun Li et al. propose a complementary approach of modeling internet's router-level topology. They characterise current degree-based approaches as incomplete since graphs with the same node degree distribution can result from different graphs in terms of network engineering. From this viewpoint, they use the notion of the "first-principles approach" to identify some minimal functional requirements and physical constraints required for developing simple models of the Internet's router-level topology. They focus on a few critical technological and economic considerations, together with abstract models of user demand, that they claim provide insight into the types of network topologies that are possible. They introduce the *network performance* and *network likelihood* to contrast generated and real network topologies. Borrel et al. [13] form a mobility model based on the scale-free spatial distribution observed in real-life networks. Unlike current strictly individual or group mobility models, they introduce a new mobility model that incorporates individual behavior and group interaction based on the preferential attachment principle. Their simulations yielded topologies with scale-free spatial distribution of nodes.

This research extends our earlier studies [3], [2], [1], the studies by Kotz and Essien [14], Balachandran et al. [15] and Balazinska and Castro [16] by focusing more closely on the roaming activity in an aggregate level and the AP-topological properties. In [1], we modeled the arrival of wireless clients at the access points (APs) in a production 802.11 infrastructure as a time-varying Poisson processes. These results were validated using quantile plots with simulation envelope for goodness-of-fit tests and by modeling the visit arrivals at different time intervals and APs. Our earlier work [3] models the associations of each wireless client as a Markov-chain in which a state corresponds to an AP that the client has visited. Based on the history of the transitions between such APs, we build a Markov-chain model for each client. Even for the very mobile clients, this model can predict well the next AP that a client will visit as it is roaming the wireless infrastructure. Furthermore, a class of bipareto distributions can be employed to model the duration of the visits at APs and also the duration of a continuous wireless access [2].

Kotz et al. [15] studied the evolution of the wireless network at Dartmouth College using syslog, SNMP, and tcpdump traces. They reported the average number of active cards per active AP per day (2-3 in 2001, and 6-7 in 2003/2004) and average daily traffic per AP by category (2-3 times higher in 2003/2004; twice or thrice more inbound than outbound traffic).

V. CONCLUSIONS AND FUTURE WORK

The main contribution of this paper is the modeling of the roaming activity in a large-scale wireless infrastructure using real-life measurements. Specifically, we modeled the roaming activity as a graph and measured its properties and evolution in the spatial and temporal domain. The placement of new APs results in a decrease of the percentage of crosspoint APs. Furthermore, a large percentage of APs are placed in the vicinity of APs with high roaming patterns. The degree of connectivity can be modeled using a negative binomial distribution. There is

a rapid growth of the wireless infrastructure. We evaluated the impact of the newly added APs on the degree of connectivity, diameter, number of connected components and continuity of roaming. The diameter size appears to be unaffected by the network infrastructure growth and increase of the client population. Furthermore, a giant component in the graph persists in all three tracing periods.

A natural follow-up of this paper is the detection and description of the weak or dead spots of the topology, through complementary graphs where an edge represents the unsuccessful roaming transitions (e.g., transitions which resulted in undesirable client disassociations and flow terminations). Such graphs could employed as diagnostic tools by revealing problems, such as misconfigured or misplaced APs. It would be interesting to validate their results and contrast them with tools based on signal strength information. The acquisition of signal strength measurements in large-scale, uncontrolled environments is challenging so the use of higher level information in larger time is attractive. Furthermore, we plan to model the spatial distribution of APs and clients. We are in the process of validating our models using the wireless traces from Dartmouth and FORTH and contrast the impact of the scale in the number of APs and clients on graph.

<div align="center">REFERENCES</div>

[1] M. Papadopouli, H. Shen, and M. Spanakis, "Modeling client arrivals at access points in wireless campus-wide networks," in *14th IEEE Workshop on Local and Metropolitan Area Networks*, Chania, Crete, Greece, 2005.

[2] M. Papadopouli, H. Shen, and M. Spanakis, "Characterizing the duration and association patterns of wireless access in a campus," in *11th European Wireless Conference*, Nicosia, Cyprus, 2005.

[3] 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.

[4] "Cisco event log messages," http://www.cisco.com/univercd/cc/td/doc/product/wireless/airo1200/accsspts/ap120scg/bkscgaxc.htm/.

[5] A. Pettitt and M. A. Stephens, "The kolmogorov-smirnov goodness-of-fit test for discrete and grouped data," in *Technometrics*, vol. 19, no. 2, May 1977, pp. 205–210.

[6] R. B. D'Agostino and M. A. Stephens, *Goodness-of-Fit Techniques*. Marcel Dekker, 1986.

[7] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," in *SIGCOMM Symposium on Communications Architectures and Protocols*, Philadelphia, Sept. 1999.

[8] M. E. J. Newman, "Power laws, pareto distributions and zipf's law," in *Contemporary Physics*, vol. 46.

[9] R. Albert and A.-L. Barabasi, "Statistical mechanics of complex networks," in *Rev. of Modern Physics*, vol. 74, 2002.

[10] L. Li, D. Alderson, W. Willinger, and J. Doyle, "A first principles approach to understanding the internet's router-level topology," in *ACM SIG-COMM*, Philadelpia, 2004.

[11] E. L. Lehmann and H. J. M. D'Abrera, *Nonparametrics: Statistical Methods Based on Ranks, rev. ed.* New Jersey, NJ, USA: Prentice-Hall, 1998.

[12] S. Ross, *Introduction to probability and statistics for engineers and scientists*. Wiley Series in Prabability and mathematical statistics, 1987.

[13] V. Borrel, M. D. de Amorim, and S. Fdida, "On natural mobility models," 2005.

[14] 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.

[15] 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.

[16] 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