Modeling Roaming in Large-scale Wireless Networks using Real Measurements

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

a. Department of Computer Science, University of North Carolina at Chapel Hill, USA.
b. Institute of Computer Science, Foundation for Research and Technology-Hellas, Greece.
c. Department of Computer Science, University of Crete, Greece.
Emails: {maria,mkaralio}@cs.unc.edu, moudatso@ics.forth.gr

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. We employed graph theory, modeled the roaming activity as a graph and measured its degree of connectivity. 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 degree of connectivity of the graph.

1 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. Wireless LANs have more vulnerabilities and stricter bandwidth and latency constraints than their wired counterparts. For the design and evaluation of capacity planning, resource reservation, device adaptation, and load balancing, 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.

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 on a client basis [15, 14, 5], 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 analyzes its degree of connectivity, which is modeled well by the Negative Binomial distribution. 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 2, we briefly describe the wireless network infrastructure, data acquisition, and testbed. Section 3 models the roaming activity and Section 4 presents the related work. Finally, Section 5 summarizes the main results and future work plans.

*This work was partially supported by the IBM Corporation under an IBM Faculty Award 2005 grant.

Table 1. Number of clients and APs during the tracing periods

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

2 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 the 729-acre campus and a number of off-campus administrative offices. 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. In our earlier work [5], 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. While there is a small percentage of PDAs, the majority of the devices are laptops. In addition to each AP's unique IP address, we maintain information about the buildings where APs are located in and their 2D coordinates.

We acquire and analyze syslog data from three different monitoring periods covering the interval from October 2004 to April 2005. Table 1 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.

3 Modeling roaming activity as graph

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 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 2. 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. In [14], we found that 24% of the clients were mobile in terms of inter-building AP transitions. 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.

3.1 Degree of connectivity

Figure 1 illustrates the quantile-quantile (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 qq plot is a graphical

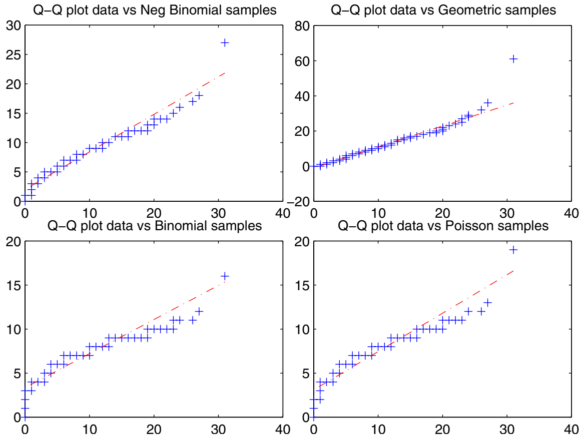

Figure 1. Q-Q plots of AP outdegree against four discrete distributions (week 1)

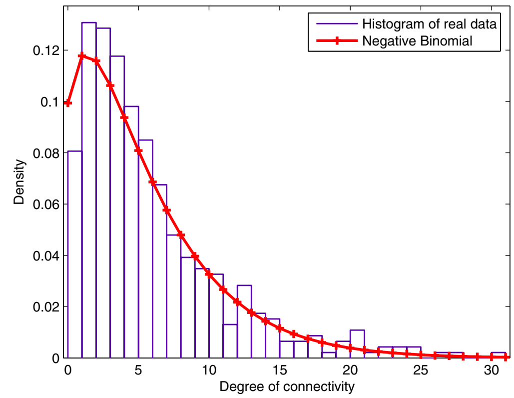

Figure 2. Histogram of data and Negative Binomial theoretical mass function

technique for determining if two data sets come from populations with a common distribution. The graphical tests suggest that both the Geometric and 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 [16] 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 2 reports the ML estimates for the Negative Binomial distribution parameters. Faloutsos *et al.* [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. Compared to the Internet topology, a campus WLAN exhibits a flatter connectivity structure. In the log-log scale plots of the outdegree frequency versus the

Table 2. ML estimates for the Negative Binomial parameters

Edges	Week 1	Week 2	Week 3
Incoming	1.51, 0.21	1.78, 0.26	1.83, 0.25
Outgoing	1.58, 0.22	1.79, 0.26	1.73, 0.23

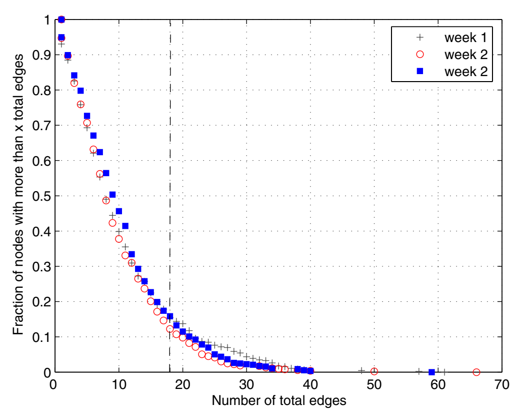

Figure 3. Degree of connectivity

outdegree (Figure 3 [13]), the correlation coefficient ranges from 0.88 to 0.91.

Figure 3 shows the CCDF of the degree of connectivity for the three tracing periods (namely, weeks 1, 2, and 3). The stochastic order of nodes with small or medium degree of connectivity is different from the one of nodes with high degree of connectivity over the three weeks. Up to a certain degree of connectivity value (18), the last week has the highest degree in the stochastic order among the three weeks. Beyond this value (indicated by a black vertical line in Figure 3), the stochastic order changes. 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). Furthermore, there is a 40% increase in the number 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 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 3 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 crosspoints 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 due to either reduced user mobility or placement of new APs. Since the 20% of crosspoints of 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, 56% of the common APs 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.

Figure 4 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 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. Arround the periphery of the campus, we can distinguish new isolated nodes with low degree of connectivity. While in more dense areas, the newly placed APs and their neighbors exhibit a relatively high degree of connectivity.

3.2 Degree of connectivity vs. distance

We expected to find a relation between the number of transitions between two APs and their distance. In [13], a scatterplot illustrates the number of transitions between APs versus their Euclidean distance. In this analysis, we only consider pairs of APs with at least one transition. The Pearson product-moment correlation coefficient R of these two variables ranges from -0.04 to -0.09 for the the three tracing periods. The low values of R imply that the correlation under discussion has non-linear characteristics, although they are still high enough to reject the null hypothesis of no dependence between the two data sets at 1% significance level. 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. These values ranged from -0.29 (week 3) to -0.45 (week 1). In all cases the null hypothesis that the two variables are independent is rejected at significance levels far beyond 1%. Further

Table 3. Evolution of crosspoints in the graph during the tracing periods

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

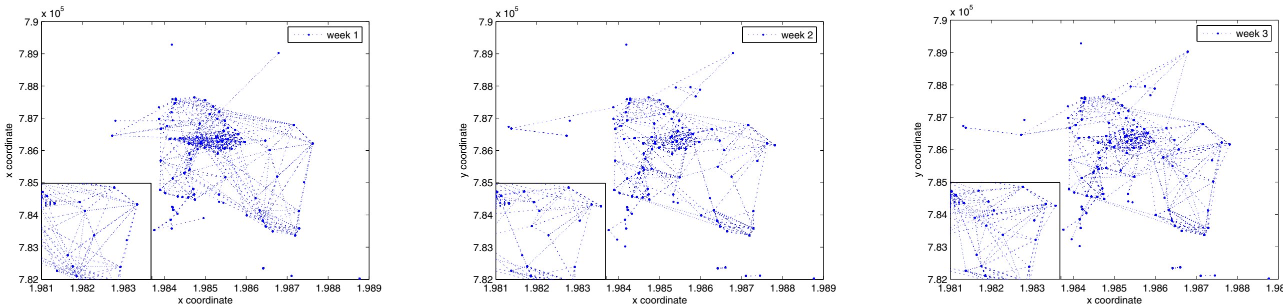

Figure 4. Transition graph in three week periods

supporting evidence for the existence of negative correlation between the distance of two APs and the number of transitions measured from the one to the other, came from the test of independence in contingency tables [17]. 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.

4 Related Work

Albert and Barabasi [1] focus on the statistical mechanics of network topology and dynamics, the main models and analytical tools for 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. Faloutsos *et al.* [7] prove that the AS and router level internet topology can be described with power laws. Lun Li *et al.* [12] 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. Borrel *et al.* [4] form 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 [5], [14], [15], the studies by Henderson *et al.* [8], Kotz and Essien [10], Balachandran *et al.* [2] and Balazinska and Castro [3] by focusing more closely on the roaming activity in an aggregate level and the AP-topological properties. In [15], 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. Minkyong *et al.* [9] cluster APs based on their peak hours and analyze the distribution of arrivals for each cluster, using the aggregate client arrivals and departures at APs. Our earlier work [5] 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 transitions between 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 while roaming in the wireless infrastructure. Furthermore, a class of bipareto distributions can be employed to model the duration of visits at APs and continuous wireless access [14].

5 Conclusions and future work

We modeled the roaming activity as a graph and measured its properties and evolution in the spatial and temporal domain. For example, the degree of connectivity can be modeled using a Negative Binomial distribution. 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. We evaluated the impact of newly added APs on the degree of connectivity.

A natural extension of this paper is the detection and modeling of the weak spots in a wireless network, through complementary graphs where an edge represents unsuccessful roaming transitions. Such graphs could be employed as diagnostic tools by revealing problems, such as misconfigured or misplaced APs. It would be interesting to validate and contrast such results with tools based on signal strength information. The acquisition of signal strength measurements in large-scale, uncontrolled environments is challenging and the use of cross-layer information in larger time scales can be helpful. Finally, we plan to analyze traces from different wireless environments, contrast their corresponding graphs, and evaluate the impact of the network size, AP density, and access pattern on the graph.

References

[1] R. Albert and A.-L. Barabasi. Statistical mechanics of complex networks. In *Rev. of Modern Physics*, volume 74, 2002.

[2] 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*, California, USA, 2002.

[3] 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 (MobiSys)*, San Francisco, USA, May 2003.

[4] V. Borrel, M. D. de Amorim, and S. Fdida. On natural mobility models. In *LNCS, International Workshop on Autonomic Communication (WAC)*, Athens, Greece, October 2005.

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

[6] R. B. D'Agostino and M. A. Stephens. *Goodness-of-fit techniques*. Statistics: Textbooks and Monographs, New York: Dekker, 1986, edited by D'Agostino, Ralph B.; Stephens, Michael A., 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, September 1999.

[8] T. Henderson, D. Kotz, and I. Abyzov. The changing usage of a mature campuswide wireless network. In *ACM International Conference on Mobile Computing and Networking (MobiCom)*, Philadelphia, September 2004.

[9] M. Kim and D. Kotz. Modeling users' mobility among wifi access points. In *WiTMeMo '05*, Berkeley, CA, USA, June 2005. USENIX Association.

[10] D. Kotz and K. Essien. Analysis of a campus-wide wireless network. Technical Report TR2002-432, Dept. of Computer Science, Dartmouth College, September 2002.

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

[12] L. Li, D. Alderson, W. Willinger, and J. Doyle. A first principles approach to understanding the internet's router-level topology. In *ACM SIGCOMM*, Philadelpia, USA, 2004.

[13] M. Papadopouli, M. Moudatsos, and M. Karaliopoulos. Modeling roaming in large-scale wireless networks using real measurements. Technical Report 369, ICS-FORTH, Greece, January 2006.

[14] 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, April 2005.

[15] 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, Greece, September 2005.

[16] A. Pettitt and M. A. Stephens. The kolmogorov-smirnov goodness-of-fit test for discrete and grouped data. In *Technometrics*, volume 19, pages 205–210, May 1977.

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