

# Modeling client arrivals at access points in wireless campus-wide networks

Maria Papadopouli<sup>a,c</sup>   Haipeng Shen<sup>b</sup>   Manolis Spanakis<sup>c</sup>

**Abstract**—Our goal is to model the arrival of wireless clients at the access points (APs) in a production 802.11 infrastructure. Such models are critical for benchmarks, simulation studies, design of capacity planning and resource allocation, and the administration and support of wireless infrastructures. Our contributions include a novel methodology for modeling the arrival processes of clients at APs and the use of a powerful visualization tool for finding detailed interior features and quantile plots with simulation envelope for goodness-of-fit test. Time-varying Poisson processes can model well the arrival processes of clients at APs. We validate these results by modeling the visit arrivals at different time intervals and APs. Furthermore, we propose a clustering of the APs based on their visit arrival and functionality of the area in which these APs are located.

## I. INTRODUCTION

Wireless networks are increasingly being deployed and expanded in airports, universities, corporations, hospitals, residential, and other public areas to provide wireless Internet access. While there is a rich literature characterizing traffic in wired networks, there are only a few studies available that examined and modeled wireless traffic and access patterns in a production network. In general, most of the wireless measurement studies provide general statistics on the overall usage of a wireless infrastructure and client mobility without detailed analysis or modeling.

One of our research directions is the systematic analysis, characterization, and modeling of the wireless infrastructures and access patterns and the implications of these results on the design of protocols and monitoring tools to support a production network. We investigate the traffic load characteristics (e.g., bytes, number of packets, associations, distinct clients, type of clients), their dependencies and interplay in various time-scales, from both the perspective of a client and an access point (AP).

Modeling how wireless clients arrive at different APs, how long they stay at them, and the amount of data they access can be beneficial in capacity planning, administration and deployment of wireless infrastructures, protocol design for wireless applications and services, and their performance analysis. Currently, most of the simulation studies on wireless networks and protocols consider simplistic association and mobility patterns for the wireless users [23], [12]. There are a few studies on the mobility and association patterns in cellular

networks [21], [14], [5]. In this study, we focus on modeling the arrivals of clients at APs. Such models can be used for simulation studies and nicely extend our earlier results [6], [18]. Specifically, in [6], we modeled the associations of a 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 [18].

A better understanding of the arrival rate of clients at APs can also assist in forecasting the traffic demand at APs. Short-term (e.g., a few minutes) forecasting can be employed in the design of more energy-efficient clients and resource reservation and load balancing (among APs) mechanisms. Long-term forecasting is essential for capacity planning and understanding the evolution of the wireless traffic and networks. In our earlier studies [11], [17], we analyzed the traffic load and some forecasting algorithms. We observed some intriguing phenomena in the traffic load of APs regarding the uploading user behaviour, its relation with the number of associated clients, and its diurnal and spatial locality patterns [11], [17]. Furthermore, due to the burstiness in the traffic load at hotspot APs and the heterogeneity of this large, diverse networking environment, the standard forecasting methods exhibit relatively large error. These observations motivate us to further analyze the client associations and arrival patterns. By incorporating information about the client and flow arrivals at APs, we may improve the traffic load predictions.

The main contributions of this work are the following: (a) a novel methodology for modeling the arrival processes of clients at wireless APs, (b) the use of a very powerful visualization tool (the SiZer map) for finding detailed interior features and quantile plots with simulation envelope for goodness-of-fit test, (c) models of the arrival processes of clients at APs as a *time-varying Poisson process* with different arrival-rate function to model their arrival at an AP. Furthermore, we investigate the impact of the type of building (i.e., its functionality) in which the AP is located at the arrival rate and cluster these visit arrival models based on the building type.

Section II describes our infrastructure and data acquisition process. In Section III, we focus on analysis and modeling of the interarrivals. Section IV discusses previous related research. In Section V, we summarize our main results and

a. Department of Computer Science, University of North Carolina at Chapel Hill, US. (maria@cs.unc.edu)

b. Department of Statistics & Operations Research, University of North Carolina at Chapel Hill, US. (haipeng@email.unc.edu)

c. Institute of Computer Science, Foundation for Research and Technology-Hellas, Greece. (spanakis@ics.forth.gr)

discuss future work.

## II. WIRELESS INFRASTRUCTURE AND DATA ACQUISITION

The data comes from the large campus wireless networks deployed at UNC [1]. UNC's network provides coverage for 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 able to communicate using the campus wireless network. A total of 488 APs were part of the campus network for our study. These 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 ran Cisco IOS while the 340s ran VxWorks. The majority of APs on campus were configured to send trace data via *syslog* messages to a *syslog* server in our department. The messages sent by the APs are detailed in [6].

In addition to each AP's unique IP address, we maintain information about the building the AP is located in, its *type*, and its coordinates. The possible building types are academic, administrative, athletic, business, clinical, dining, library, residential, greek, student stores, health affairs, playing fields, performing arts, and theatre.

### A. Visit definition

The *syslog* messages are ordered based on their timestamp (i.e., the time when they were received from the monitor in our department). The *session generator* [18] parses these *syslog* ("info-type") messages [2] sequentially for each client, interprets each event with respect to the Cisco documentation [2], and creates state information for each client. Specifically, for each client, it generates a sequence of visits at APs and its transitions from one AP to another or to its disconnection from the infrastructure. We define the *visit* of a client (e.g., client *C*) to an AP (e.g., AP *i*) at a certain time *t* as the sequence of the *continuous* associations of the client to that AP. The first association of that sequence occurs at time *t* and indicates the start of the visit and the arrival of the client *C* to AP *i*. Prior to *t*, client *C* was associated with a *different* AP or disconnected. The end of this visit occurs when the client associates with a different AP or gets disconnected from the wireless infrastructure. A client may visit the same AP more than once. In this paper, we model the *visit arrivals* of clients at a given AP. We use the terms *visit arrivals* and *client arrivals* interchangeably. More details about the session generation can be found in [18].

### B. Hotspot APs

The overloaded APs according to their hourly and also overall amount of bytes accessed from them (i.e., hotspot APs) represent an interesting set of APs in the production network. In our earlier studies, we analyzed their traffic characteristics and employed traffic forecasting algorithms on them. These studies motivate us further to characterize the daily arrival process of client visits to these hotspots. From the set of the 488 APs we have identified 16 hotspot APs. We define an AP as hotspot when it belongs in the intersection of two sets, namely the top 5% APs based on total maximum traffic and

the top 5% APs based on maximum hourly traffic. For our analysis, we excluded the hotspot APs with unknown building type and location. The distribution of the selected hotspot APs per building type is as follows: academic (8 APs), library (3 APs), residential (3 APs), and theater (2 APs)<sup>1</sup>.

## III. ANALYSIS

We review the definition of a time-varying Poisson process and construct a test for such a process in Section III-A. The test is then employed in Section III-B to show that the visit arrival process at a particular AP seems to be a time-varying Poisson process.

### A. Time-varying Poisson Processes

1) *Background*: Suppose  $\{\Lambda(t) : t \geq 0\}$  is a stochastic point process, which counts the number of events (or arrivals) in  $[0, t]$ . Sometimes,  $\{\Lambda(t)\}$  is referred as the arrival process of the events of interest. For example, in the current paper,  $\{\Lambda(t)\}$  is the arrival process of client visits at a particular AP.  $\{\Lambda(t)\}$  is a Poisson process if it has the following two properties:

- 1) The number of arrivals in disjoint intervals are independent;
- 2) For some finite  $\lambda > 0$ ,  

$$P(\Lambda(t) = j) = e^{-\lambda t} (\lambda t)^j / j!, j = 0, 1, \dots$$

Thus, for each *t*,  $\Lambda(t)$  is a Poisson random variable with mean  $\lambda t$ , the product of the arrival rate  $\lambda$  and the interval length *t*. Note that a Poisson process is a renewal process where the inter-arrival times are independent exponential[20]. It is well-known that such a process results from the following behavior: there exist many potential, statistically identical arrivals; there is a very small yet non-negligible probability for each of them arriving at any given time; and arrivals happen independently of each other.

One closely related variation is a time-varying Poisson process, where the arrival rate is a function of time *t*, say,  $\lambda(t)$ . Such a process is the result of time-varying probabilities of event arrivals, and it is completely characterized by its arrival rate function. Smooth variation of  $\lambda(t)$  is familiar in both theory and practice in a wide variety of contexts, and seems reasonable for modelling client visits to an AP. One example of such an arrival rate is illustrated in Fig. 1(a) using a SiZer map, which is described in [19].

2) *Construction of a Statistical Test*: We would like to construct a test for the null hypothesis that an arrival process is a time-varying Poisson process, with a slowly varying arrival rate. To begin with, we break up the interval of a day into relatively short blocks of time. For convenience, blocks of equal length, *L*, are used, resulting in a total of *I* blocks; though this equality assumption can be relaxed. For the analysis in Section III-B, we use  $L = 0.1$  hour. Let  $T_{ij}$  denote the *j*th ordered arrival time in the *i*th block,  $i = 1, \dots, I$ . Thus

<sup>1</sup>We had only limited information about the exact functionality of the rooms in which the hotspots were located. For example, APs in academic buildings could be found in classrooms, offices for advising students, lounges, and meeting rooms. In residential buildings (dorms or greek houses), we found hotspot APs in lounges and labs.

$T_{i1} \leq \dots \leq T_{iJ(i)}$ , where  $J(i)$  denotes the total number of arrivals in the  $i$ th block. Define  $T_{i0} = 0$  and for  $j = 1, \dots, J(i)$ ,

$$R_{ij} = (J(i) + 1 - j) \left[ -\log \left( \frac{L - T_{ij}}{L - T_{i,j-1}} \right) \right]. \quad (1)$$

Under the null hypothesis that the arrival rate is constant within each time interval, the  $\{R_{ij}\}$  will be independent standard exponential variables as we now discuss.

Let  $U_{ij}$  denote the  $j$ th (unordered) arrival time in the  $i$ th block. Then the assumed constant Poisson arrival rate within this block implies that, conditionally on  $J(i)$ , the unordered arrival times are independent and uniformly distributed between 0 and  $L$ . Denote  $V_{ij} = \frac{L - U_{ij}}{L - U_{i,j-1}}$ , and it follows that  $V_{ij}$  are independent standard exponential. Note that  $T_{ij} = U_{i(j)}$ , thus

$$V_{i(j)} = \log \left( \frac{L}{L - U_{i(j)}} \right) = \log \left( \frac{L}{L - T_{ij}} \right).$$

As one can see,  $R_{ij} = (J(i) + 1 - j) (V_{i(j)} - V_{i(j-1)})$ . Then, the exponentiality of  $R_{ij}$  results from the following lemma.

**Lemma:** Suppose  $X_1, \dots, X_n$  are independent standard exponential, then  $Y_i = (n - i + 1)[X_{(i)} - X_{(i-1)}]$ ,  $i = 2, \dots, n$ , are independent standard exponential.

Any customary test for the exponential distribution can then be applied to  $R_{ij}$  for testing the null hypothesis. For convenience, we use the familiar Kolmogorov-Smirnov test[8]. This nonparametric test is based on the maximum deviation between the empirical cumulative distribution function (CDF) of the data and the hypothesized theoretical CDF.<sup>2</sup>

### B. The Visit Arrival Process at AP222

For illustration purpose, we show below the analysis of the arrival process of client visits at the hotspot AP222, which is located in an academic building. We have also validated the analysis using APs of other building types.

1) *Exploratory Data Analysis:* Fig. 1(a) is a SiZer analysis of the visit start times at AP222. SiZer is a powerful statistical visualization tool for detecting significant features within the data, which is described in the [19]. The top panel displays a number of thin blue curves, which are estimated smooth curves for the underlying visit arrival rate function,  $\lambda(t)$ . These are local linear density estimates obtained from the observations, some of which are displayed as jittered green dots on the top. The blue curves correspond to different levels of smoothing, which are controlled by the smoothing parameter, the "bandwidth" (window size for averaging)[9]. The bottom panel plots the SiZer map of the data. The horizontal locations in the SiZer map are the same as in the top panel, and the vertical locations correspond to the same logarithmically equally spaced bandwidths used for the blue curves in the top panel. As explained in the [19], blue (or red) colors in the SiZer map indicate that the blue curves are significantly increasing (or decreasing), while purple colors correspond to no significant trend.

As one can see, the arrival rate appears to be time-varying at all scales. For coarse scales (or large bandwidths), there is an overall daily increasing (or blue) trend; for medium scales, the rate function decreases first between early morning and 14:00, and starts to increase until 18:00 before decreasing again. More features appear for fine scales. There are several alternating blue and red stripes between 14:00 and midnight, which correspond to the bumps in the blue curves. This suggests that the arrival rate has several ups and downs during this period. To better look at the features, we focus on the hour between 17:30 and 18:30, which has the largest arrival rate and consists of 2143 visits. We first calculate the inter-arrival times of every two consecutive sorted visits, and Fig. 1(b) shows the corresponding SiZer map. Interestingly, the green dots appear as evenly spaced vertical stripes, which cause evenly spaced bumps in the blue curves and alternating blue and red strips in the SiZer map at the corresponding locations at fine scales. This phenomenon suggests that the inter-arrival times only take on a set of discrete values. A closer look reveals that the space between two neighboring green stripes is exactly 1/3600, which means that the phenomenon is an artifact caused by the rounding of visit arrival times to nearest whole seconds. To compensate for this rounding effect, we "unround" the data by adding independent uniform (-0.5, 0.5) noise to each visit start time before calculating the inter-arrival times. This unrounding does not change the appearance of Fig. 1(a); however, it does change the SiZer map of the inter-arrival times dramatically, which is shown in Fig. 1(c).

The distribution of the inter-arrival times is analyzed in Fig. 2(a)-(b). Note that this distribution is exponential if the arrival process is Poisson. Fig. 1(a) already shows that this is not the case here, which is confirmed by Fig. 2(a), an exponential quantile plot for the inter-arrival times. This is a graphical method for assessing the goodness of fit of the exponential distribution to the data. The red curve is the main quantile plot, which plots the *actual* data (based on our traces) quantiles versus the corresponding *theoretical* exponential quantiles. The parameter for the theoretical distribution is estimated using maximum likelihood. When the theoretical distribution is a good fit, the red curve should follow the diagonal green line closely. To account for possible sampling variability, a blue envelope of 100 overlaid curves is constructed. Each blue curve is a similar quantile plot, where the "data" are simulated from the theoretical exponential distribution. This blue envelope provides a simple visual accounting for the sampling variability. When the theoretical exponential distribution fits the inter-arrival times well, the red curve should lie mostly within the envelope. The observed substantial departure of the red curve from the blue envelope in Fig. 2(a) strongly suggests that the inter-arrival times are not exponentially distributed.

Fig. 2(b) is a Weibull quantile plot for the inter-arrival times. The two parameters of the Weibull distribution are fitted by matching the 90th and 99th percentiles of the data and the theoretical distribution. The plot seems to suggest that the inter-arrival times are approximately Weibull. In addition, the inter-arrival times in our data are not independent as shown by the corresponding auto-correlation plot in [19]. All the auto-correlations are significantly positive. The strong

<sup>2</sup>One can also test the exponentiality using a graphical tool, such as an exponential quantile plot as the one in Fig. 2(c).

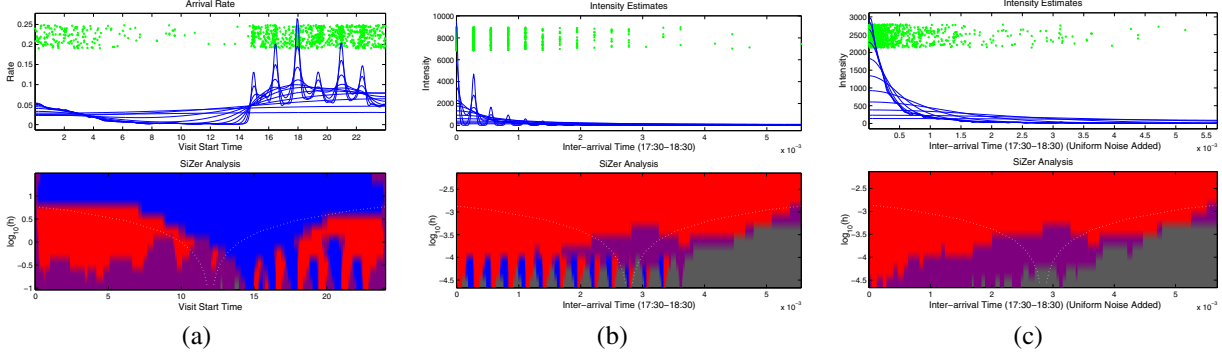


Fig. 1. Visit Arrivals (AP222): (a) SiZer Map of Visit Start Times; (b) SiZer Map of Visit Inter-arrival Times; (c) SiZer Map of Visit Inter-arrival Times with Uniform Noise Added.

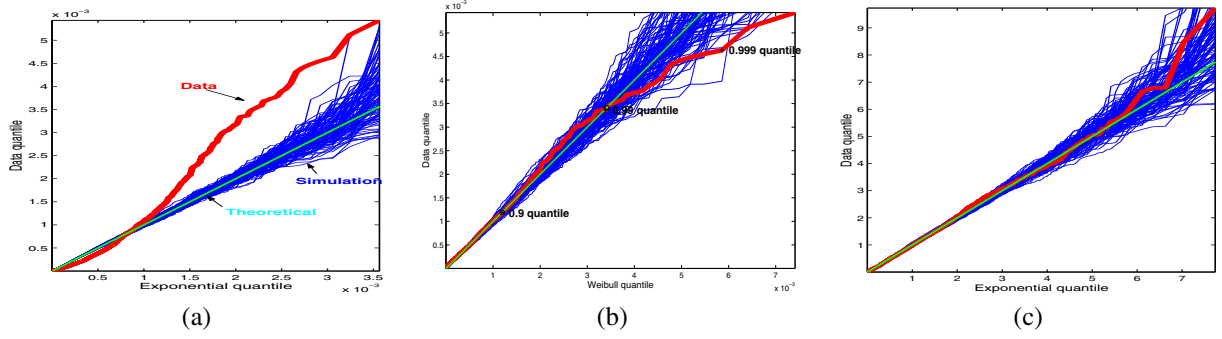


Fig. 2. Visit Inter-arrival Times between 17:30 and 18:30: (a) Exponential Quantile Plot, (b) Weibull Quantile Plot; (c) Exponential Quantile Plot for  $R_{ij}$  between 17:30 and 18:30 (AP 222).

auto-correlation of the inter-arrival times suggests that we can not model the visit arrival process as a renewal process with independent Weibull inter-arrival times. A more appropriate model is to combine Weibull inter-arrival times with a suitable dependence structure as suggested by [7]. Generating or simulating such a dependent process is much more complicated, because one has to specify the dependence structure reasonably. Below, we propose to use time-varying Poisson processes as an alternative model, which has a nice practical interpretation and is easier to simulate.

2) *A Time-varying Poisson Process Model for The Visit Arrivals:* In this section, we use the test proposed in Section III-A and an exponential quantile plot to show that the arrival process of client visits at AP222 is a Poisson process with a time-varying arrival rate. The analysis is carried out in detail for the process between 17:30 and 18:30 only. We break the hour into ten 6-minute intervals, and calculate the  $R_{ij}$  according to (1) by setting  $L = 0.1$  hour. The corresponding Kolmogorov-Smirnov test statistic is 0.0188, and has a p-value of 0.15 with 2143 observations, which means that the null hypothesis can not be rejected. Fig. 2(c) shows the exponential quantile plot for the  $R_{ij}$ , which clearly suggests that they are exponential. The maximum likelihood estimate for the exponential parameter is 1.0024, which is very close to 1 and agrees with the claim that the  $R_{ij}$  are standard exponential. The corresponding auto-correlation plot suggests that the  $R_{ij}$ s are approximately independent [19]. Thus, the null hypothesis that the visit arrival process within the hour is time-varying Poisson is validated both mathematically and graphically.

There are well developed methods for simulating time-varying Poisson processes, for example, the *thinning* method described in [13], [22]. Along with models for visit durations, we can generate synthetic traces. We also looked at a few other hours at AP 222, and the result is consistent with the results reported in Section III-B. Furthermore, we repeated the analysis on three other hotspot APs, namely AP 405(library), AP 442 (theatre), and AP 460 (residential building), and got similar results. Together with AP 222, we have one AP from each building type.

### C. Clustering of APs based on their building types

How does the functionality of an area affect the arrival rate of client visits at APs located in that area? We are interested in exploring the spatial and functionality locality on the visit arrivals and if general statistics on the arrivals can reveal a clustering. In general, we have limited information about the exact activities, schedules, and usage characteristics in the areas around the APs. Furthermore, there are areas used for diverse activities (e.g., floors in an academic building with meeting rooms, offices, classrooms, and lounges). In this analysis, we also focused on hotspot APs. We did observe clusters of hotspot APs with similar arrival patterns according to their functionality (e.g., lounges vs. labs in residential buildings vs. classrooms). In Fig. 3(a), each line corresponds to a non-academic AP and indicates aggregate hourly percentage of visit arrivals at that AP. Different colors and line types are used to differentiate the APs according to their building type. We can distinguish three clusters of APs, one in libraries, the

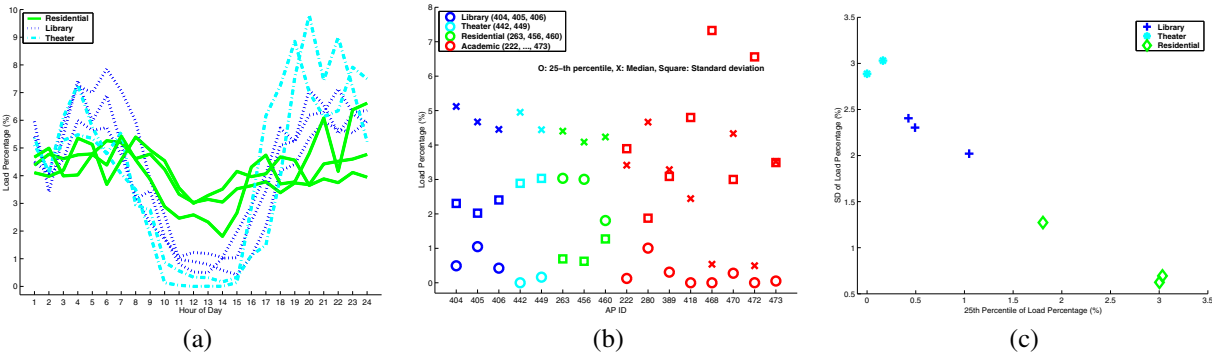


Fig. 3. Clustering of APs based on building type: (a) Hourly percentage for non-Academic APs; (b) For each AP, its 25<sup>th</sup>-percentile, median and standard deviation; (c) For each AP excluding the academic ones, 25<sup>th</sup>-percentile vs. the standard deviation.

second one in lounges of residential buildings, and the third one in meeting rooms and lounges in theaters (recreational centers). For each AP, we also calculated the 25<sup>th</sup>-percentile, median, and standard deviation. Fig. 3(b) plots the three summary statistics for each AP, trying to find some similarities. There is a great similarity among APs in meeting rooms and lounges of theater/recreational building types (APs 442, 449). APs in lounges or labs in dorm/residential areas (like AP 263 and 456) have similar patterns which differ from the ones in greek buildings/residential (AP 460). Similarly, APs located in classrooms (such as AP 389, 470, and 473) have similar patterns which actually differ from the visit arrival pattern in the classroom where lectures for a middle school take place (AP 418) or the area with offices for advising students (AP 280) (Fig. 3(b)). Also, note the similarity in the arrival patterns at the three library APs. Fig. 3(c) plots the 25<sup>th</sup> percentile vs. the standard deviation for each non-academic AP and the separation among library, theater, and residential buildings is more clear.

#### IV. RELATED WORK

Balazinska and Castro [4] used SNMP to characterize a much larger wireless network in three IBM buildings (177 APs). The study examined the maximum number of simultaneous users per AP (mostly between 5 and 15), total load and throughput distributions. However, they only polled the APs at every 5 minutes to get the current association information and do not study visit arrivals.

Balachandran *et al.* [3] considered a three-day conference setting with four APs and modeled user arrivals under minute resolution level. They proposed a two-stage Markov-Modulated Poisson Process (MMPP), which looks like a Poisson process with two arrival rates (one for Peak period and one for non-Peak period). But they do not show the performance of this model, such as the distribution of inter-arrival, goodness-of-fit. Note that this is a special conference setting and the user population is more homogeneous than ours in a very large and diverse campus network. In our environment, we believe that our model works better. For example, you can observe that the arrival rate for AP 222 changes quite dramatically between 14:00 and midnight (Fig. 1(a)). As the analysis shows, it is even unreasonable to assume the arrival rate is constant between 17:30 and 18:30, not to mention the whole ON period.

Note that the two-stage MMPP needs to specify two stages, ON and OFF. By looking at the SiZer map, it suggests to treat 14:00 to midnight as the ON period while midnight to 14:00 as the OFF period. Kotz *et al.* [10] 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).

A subset of the Dartmouth syslog messages and tcpdump traces was revisited by Meng *et al.* [15] for flow modeling purposes. They proposed a two-tier (Weibull regression) model for the arrival of flows at APs and a Weibull model for flow residing times, and they also observed high spatial similarity within the same building. Note the difference between static flow (modeled in the [15]) and visit arrival (modeled here); the first indicates the start of a TPC flow of *any client* at that AP (with all the packets of that flow accessed via that AP), whereas the second indicates the start of a client's visit at that AP. During a visit, each client (associated with that AP) may generate multiple stationary TCP flows. They make a compelling case against Poisson modeling of wireless flows (at least for busy APs). Actually, the visit inter-arrival times in our setting could be also modelled by a Weibull fit, which is similar to theirs. However, we discovered that the inter-arrival times are strongly positively correlated, which makes the Weibull finding less attractive because it does not model the correlation structure. Furthermore, the Weibull distribution is caused by the non-stationary arrival rate within the hour, which also explains why the inter-arrival times are correlated. However, they do not model the correlation between the Weibull inter-arrival times. That means, we need to find the distribution of the number of flows within a visit and the distribution of inter-arrival times between flows within a visit. Our finding of the Poisson process for the visits suggests that the arrival process for the flows should be a clustered Poisson process, which could have correlated Weibull inter-arrival times. Similar observations have been made in the wired traffic [16]. For example, a user web-browsing session vs. the cluster of flows generated by the embedded links in this session. It is part of future work to investigate this.

## V. CONCLUSIONS AND FUTURE WORK

One of our contributions is a novel methodology for modeling the arrival processes of clients at wireless APs. We found that time-varying Poisson processes can model well the arrival processes of clients at APs and validated these results by modeling the visit arrivals at different time intervals and APs. Furthermore, we discovered that we can cluster the APs based on their visit arrival and functionality of the area in which these APs are located. We intend to study the spatial correlations of APs (e.g., how the arrival rates at different APs vary with their geographic distance) and classify the APs further based on additional parameters (e.g., traffic characteristics, number of associations, and distinct clients). As mentioned in Section I, we will apply the forecasting algorithms to more homogeneous clusters of APs and improve them by incorporating additional information about their APs.

Our finding of the Poisson process for the visits suggests that the arrival process for the flows should be a clustered Poisson process, which could have correlated Weibull inter-arrival times. One of the next steps is to investigate the flow arrivals per client and provide clusters of clients based on their flow arrival models. Such models can be beneficial in employing resource allocation, load balancing, and detecting abnormal client access behaviour.

This research is part of a comparative analysis study on wireless access patterns in various environments, such as a medical center, research institute, campus, and public wireless network. We are in the process of validating our models with traces from other large campus-wide wireless infrastructures (e.g., Dartmouth University) and applying our analysis on the 3-day conference setting [3] to contrast the models (time-varying Poisson vs. MMPP). Understanding and forecasting the access patterns at APs can have a dominant impact on the operation of wireless APs and this study sets a direction for exploring further these issues.

## VI. ACKNOWLEDGMENTS

This work was partially supported by the IBM Corporation under an IBM Faculty Award 2004/2005 grant.

## REFERENCES

- [1] America's most connected campuses. <http://forbes.com/home/lists/2004/10/20/04conncampand.html>.
- [2] Appendix C Event Log Messages. In *Cisco Aironet Access Point Software Configuration Guide*.
- [3] Anand Balachandran, Geoffrey Voelker, Paramvir Bahl, and Venkat 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.
- [4] Magdalena Balazinska and Paul 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.
- [5] Amiya Bhattacharya and Sajal K. Das. LeZi-update: an information-theoretic approach to track mobile users in PCS networks. In *Proceedings of the Annual ACM/IEEE International Conference on Mobile Computing and Networking*, pages 1–12, Seattle, Washington, USA, August 1999.
- [6] Francisco Chinchilla, Mark Lindsey, and Maria Papadopoulou. Analysis of wireless information locality and association patterns in a campus. In *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, Hong Kong, March 2004.
- [7] W. S. Cleveland, D. Lin, and D. X. Sun. Ip packet generation: statistical models for tcp start times based on connection-rate superposition. *Performance Evaluation Review: Proc. ACM Sigmetrics 2000*, 28:166–177, 2000.
- [8] Ralph B. D'Agostino and Michael A. Stephens. *Goodness-of-Fit Techniques*. Marcel Dekker, 1986.
- [9] J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London, 1996.
- [10] Tristan Henderson, David Kotz, and Ilya Abyzov. The changing usage of a mature campuswide wireless network. In *ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, Philadelphia, September 2004.
- [11] Felix Hernandez-Campos and Maria Papadopoulou. A comparative measurement study of the workload of wireless access points in campus networks. In *16th Annual IEEE International Symposium on Personal Indoor and Mobile Radio Communications*, Berlin, Germany, 2005.
- [12] Amit Jardosh, Elizabeth M. BeldingRoyer, Kevin C. Almeroth, and Subhash Suri. Towards realistic mobility models for mobile ad hoc networks. In *ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, San Diego, September 2003.
- [13] P.A.W. Lewis and G.S. Shedler. Simulation of nonhomogeneous poisson process by thinning. *Naval Research Logistics Quarterly*, 26:403–413, 1979.
- [14] T. Liu, P. Bahl, and I. Chlamtac. Mobility modeling, location tracking, and trajectory prediction in wireless atm networks. *IEEE Journal on Special Areas in Communications, Special Issue on Wireless Access Broadband Networks*, 16(6):922–936, August 1998.
- [15] Xiaoqiao Meng, Starsky Wong, Yuan Yuan, and Songwu Lu. Characterizing fbws in large wireless data networks. In *ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, pages 174–186, Philadelphia, 2004.
- [16] C. Nuzman, I. Saniee, W. Sweldens, and A. Weiss. A compound model for tcp connection arrivals for lan and wan applications. *Computer Networks*, 40:319–337, 2002.
- [17] Maria Papadopoulou, Haipeng Shen, Elias Raftopoulos, Manolis Ploumidis, and Felix Hernandez-Campos. Short-term traffic forecasting in a campus-wide wireless network. In *16th Annual IEEE International Symposium on Personal Indoor and Mobile Radio Communications*, Berlin, Germany, 2005.
- [18] Maria Papadopoulou, Haipeng Shen, and Manolis Spanakis. Characterizing the duration and association patterns of wireless access in a campus. In *11th European Wireless Conference*, Nicosia, Cyprus, 2005.
- [19] Maria Papadopoulou, Haipeng Shen, and Manolis Spanakis. Modeling client arrivals at access points in wireless campus-wide networks. Technical Report 357, FORTH-ICS, Crete, Greece, May 2005.
- [20] S. T. Ross. *Stochastic Processes*. Jon Wiley & Sons, New York, 1995.
- [21] Xiaodong Wang and Zigang Yang. Joint mobility tracking and handoff in cellular networks via sequential monte carlo filtering. In *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, New York, New York, June 2002.
- [22] K. P. White. Simulating a nonstationary poisson process using bivariate thinning: the case of typical weekday arrivals at a consumer electronics store. *Proceedings of the 31st conference on Winter simulation: Simulation—a bridge to the future*, 1:458–461, 1999.
- [23] JungKeun Yoon, Mingyan Liu, and Brian Noble. Random waypoint considered harmful. In *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, San Francisco, CA, April 2003.