# On scalable measurement-driven modeling of traffic demand in large WLANs

Merkouris Karaliopoulos [a] Maria Papadopouli [b,c] Elias Raftopoulos [b,c] Haipeng Shen [d]

a. Computer Engineering and Networks Laboratory, ETH Zurich, Switzerland.
b. Institute of Computer Science, Foundation for Research and Technology - Hellas, Greece.
c. Department of Computer Science, University of Crete, Heraklion, Greece.
d. Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, USA.

*Abstract*— **Models of traffic demand are fundamental inputs to the design and engineering of data networks. In this paper we address this requirement in the context of large-scale wireless infrastructures using real measurement data from the University of North Carolina (UNC) wireless campus network. Our modeling effort focuses on capturing the demand variation in both the spatial and temporal domain in a way that scales well with the size of the wireless network. The network traffic dynamics are studied over two different week-long monitoring periods at various levels of spatial aggregation, from individual buildings to the whole network. We model traffic workload in terms of *wireless sessions* and *network flows* and find several modeling elements that are reusable in both temporal and spatial dimensions. The same set of parametric distributions for the session- and flow-related traffic variables capture the network traffic demand in both monitoring periods. Even more interestingly, these same distributions can characterize traffic dynamics at finer spatial scales, such as a single building or a group of buildings. We use our models to generate synthetic traffic and compare with trace data. The comparison clearly illustrates the trade-off between model scalability and reusability, on the one hand, and accuracy in capturing local-scale traffic dynamics on the other. Our main contribution is a novel behavioral approach for traffic demand modeling in large wireless networks that features high flexibility in the exploitation of the spatial and temporal resolution available in data traces.**

## I. Introduction

The modeling of traffic workload in large-scale wireless networks is the main focus of this paper. Although this task has been addressed in numerous research studies in the context of wired networks [1]–[4], there have been significantly fewer contributions for wireless networks.

One reason for this is that only recently, traces from large-scale wireless infrastructures with statistically significant network usage have been made available. Furthermore, wireless network measurements are more complex than those in wired networks. Depending on how detailed view of the demand is required at the spatial dimension, e.g., at an access point (AP), at all APs located in a building or set of buildings, and the architecture of the network (single link-level subnet or multiple subnets), one needs to capture traffic at multiple physical locations. Since IEEE 802.11 MAC-layer frame sniffers are not commonly available, researchers often have to build custom equipment or resort to expensive commercial platforms for capturing the over-the-air traffic with the required level of detail. Finally, the transient characteristics of the radio propagation and user mobility make the analysis of the traffic demand dynamics challenging. It comes as no surprise that the majority of the measurement studies, *e.g.*, [5]–[7], make high-level observations about traffic dynamics in both the temporal and spatial domains without getting into the detail that modeling requires.

The *scalability* and *reusability*, properties particularly desirable in modeling, complicate the problem further. Previous modeling studies have either attempted to model traffic demand over hourly intervals at the level of individual APs [8] or studied the problem at system-level deriving models for the aggregate network-wide traffic demand [9]. Clearly, both approaches have their strong and weak points. The second approach results in datasets that are amenable to statistical analysis and provides a concise summary of the traffic demand at system-level. However, it fails to capture its variation at finer spatial detail that may be required for the evaluation of network functions with focus on the AP-level (*e.g.*, load balancing). While working at the AP-level achieves that, it fails in other respects. For example, it does not scale for large wireless infrastructures and data do not always lend to statistical analysis. Moreover, the modeling results are highly sensitive to the specific AP layout of a particular network and short-term variations of the radio propagation conditions.

The above challenges have motivated this study. We took advantage of the large wireless infrastructure of University of North Carolina (UNC) to obtain large amounts of measurement data. Our datasets provide considerable insights to the traffic load dynamics across the network and allow us to derive models of adequate detail for the traffic demand variation in space and time.

Our methodological choices attempt to strike a good trade-off between the two extreme approaches in traffic modeling, that were outlined earlier, namely, AP-level and infrastructure-wide modeling. As in [9], we model traffic workload in terms of *wireless sessions* and *network flows*. Only now, we focus on the spatial dimension of its variation, using buildings as basic entities for traffic demand modeling. Major features of user activity, such as, the traffic and roaming patterns, are studied at the building level. Then, we apply heuristics to group buildings with similar traffic characteristics and achieve the scalability objective in our modeling.

Considerable effort is devoted to the validation of our modeling methodology. We assess the reusability of system-wide models to smaller spatial scales and accuracy-scalability trade off, by synthesizing traffic according to our models and comparing it with the original trace data. Moreover, the availability of datasets from two different monitoring periods, spaced one year apart, enables us to identify time-persistent elements in modelling.

Our contributions are summarized as follows:

- A hierarchical framework for modeling traffic workload both system-wide and at finer spatial scales (*i.e.*, at building level and over groups of buildings). We find that the same set of parametric distributions describe our session- and flow-related traffic variables at various spatial scales and over two different monitoring periods.
- A novel methodology for scalable modeling of the spatial variation of traffic demand in large wireless networks open to heuristics but also more formal data analysis techniques (e.g., clustering).
- Validation of our modeling approach illustrating the fundamental trade-off between model accuracy and scalability.
- A set of analysis tools that have been made publicly available to the research community to enable further comparative studies [10].

The next section briefly describes the UNC wireless network infrastructure and the collected traces. Section III outlines the principal building blocks of our modeling approach, while Section IV zooms into the spatial variation of traffic demand. The accuracy-scalability trade-off is the subject of Section V. We review related work in Section VI and summarize our main findings in Section VII.

## II. Data collection and processing

Two types of data are used in this study, packet header traces and SNMP data, drawn from the wireless network of the UNC campus. We analyzed data acquired from two separate eight-day monitoring periods; the first dataset corresponds to the period April 13-20, 2005, whereas the second one covers the interval Apr 28-May 5, 2006.

The UNC campus wireless network comprised 488 APs by April 2005 and 741 APs one year later. The network APs are spread over more than 220 in-campus buildings, including student residence halls, academic buildings, sport halls, and libraries, and a few off-campus administrative offices, providing wireless access to 26,000 students, 3,000 faculty and 9,000 staff members.

SNMP data were collected from all the network APs every five minutes. We implemented a custom SNMP-polling system relying on a non-blocking SNMP library. APs are polled independently, so that delays incurring during the processing of SNMP polls by the slower APs do not affect the other APs. Packet header traces were collected with a high-precision monitoring card (Endace 4.3GE). The card was installed in a high-end FreeBSD server and captured all packets traversing the link between UNC and the Internet in both directions. The monitoring period was 178.2 hours in 2005 and 192 hours in 2006, yielding 175GB and 365GB of packet headers respectively. The sharp increase in the collected amount of packet headers is primarily due to the significant growth of the network infrastructure between the tracing periods.

The reservation of a separate set of IP addresses for WLAN clients enabled us to filter for the wireless traffic and correlate the SNMP data drawn from the APs with the packet header traces [11]. This allowed us to infer sessions and connections, which are central to our modeling approach, as explained in Section III.

## III. Modeling methodology

### A. Wireless sessions and network flows

Starting point for this work is our infrastructure-wide traffic demand modeling [9] and the hierarchical modeling approach that organizes the client activity into two levels, namely, the wireless session and flow. A session of a client groups this client's consecutive associations to the APs of the infrastructure into episodes of continuous activity, i.e., during which the client is connected continuously to the infrastructure and produces traffic. Sessions account for the traffic non-stationarity in time and are modeled by a time-varying Poisson process. On the other hand, flows, such as TCP connections and UDP conversations, are well-separated collections of packets between a pair of Internet hosts, *i.e.*, packets that share the same transport-layer "5-tuple". The well-established advantage of flow-level modeling when compared with packet-level modeling is its higher

TABLE I

SUMMARY OF MODELS FOR NETWORK-WIDE TRAFFIC DEMAND VARIABLES

| Modeled variable | Model | Probability Density Function (PDF) | Parameters 2005 | Parameters 2006 |
|---|---|---|---|---|
| Session arrival | Time-varying Poisson($\lambda(t)$) | $N$: # of sessions between $t_1$ and $t_2$ $\lambda = \int_{t_1}^{t_2} \lambda(t)dt$, $Pr(N=n) = \frac{e^{-\lambda}\lambda^n}{n!}$, $n = 0, 1, \ldots$ | Hourly rate: 44 (min), 1132 (max), 294 (med.) | Hourly rate: 75 (min), 1171 (max), 460 (med.) |
| Flow interarrival/session | Lognormal | $p(x) = \frac{1}{\sqrt{2\pi}x\sigma} \exp\left[-\frac{(\ln x-\mu)2}{2\sigma 2}\right]$ | $\mu = -1.37$, $\sigma = 2.79$ | $\mu = -1.49$, $\sigma = 2.92$ |
| Flow number/session | BiPareto | $p(x) = k^{\beta}(1+c)^{\beta-\alpha}x^{-(\alpha+1)}(x+kc)^{\alpha-\beta-1}$ $(\beta x + \alpha kc)$ , $x \geq k$ | $\alpha = 0.06, \beta = 1.72,$ $c = 284.79, k = 1$ | $\alpha = 0.09, \beta = 1.49,$ $c = 585.4, k = 1$ |
| Flow size | BiPareto | Same as above | $\alpha = 0.00, \beta = 0.91,$ $c = 5.20, k = 179$ | $\alpha = 0.00, \beta = 1.03,$ $c = 18.41, k = 152$ |

TABLE II

SUMMARY OF UNC CAMPUS BUILDING TYPES

| Building type | Academic | Administrative | Athletic | Business | Clinical | Library | Residential | Social |
|---|---|---|---|---|---|---|---|---|
| Number | 51 | 25 | 17 | 8 | 18 | 4 | 117 | 10 |

independence from the specific network topology and measurement conditions [1]. The flow-related session attributes we model are the in-session number of flows, in-session flow interarrivals and flow sizes. Notably, we found out that the same statistical distributions, though with different parameter sets, model our traffic variables in both monitoring periods. The in-session number of flows and the flow sizes are well modeled by the biPareto distribution, whereas the Lognormal distribution is the best fit for in-session flow interarrivals out of a set of common distributions, including Weibull, Gamma, and Pareto. A time-varying Poisson process with constant rate over intervals of an hour captures the non-stationarity of session arrivals. The results for the 2005 monitoring period are detailed in [9], whereas those for 2006 in [11]. Table I summarizes the distributions and their parameters for the two periods.

### *B. Buildings rather than APs*

One of the main advantages of working at the system-level is the availability of statistically significant data for all modeled variables. This is not always the case with individual APs; in fact, only a limited number of hot-spot APs are amenable to modeling. A second major concern with AP-level modeling is scalability. The number of distributions that have to be derived and simulated grows linearly with the number of APs, which is not desirable when studying large-scale wireless infrastructures.

Buildings are in the epicenter of our approach. The UNC campus includes approximately 250 buildings, which can be grouped according to their main/exclusive usage into eight building types listed in Table II. We view buildings as more reliable entities for modeling the spatial variation of traffic demand. In fact, an analogy between flows-packets and buildings-APs can be drawn. Much as packet-level dynamics are subject to network topology and instantaneous conditions, AP-level user activity is sensitive to radio propagation dynamics and environmental settings. One good example is the "ping-pong" effect, where a stationary user may be alternately associated with two, or even more, APs due to short-term radio signal propagation variations.

The following sections investigate the appropriate level of modeling the traffic at the spatial dimension that yields the best trade-off between accuracy and scalability. Generally, for each traffic variable listed in Table I, the spatial detail of modeling could be the building, building type, or the entire network, in the extreme case.

## IV. SPATIAL CHARACTERISTICS OF TRAFFIC DEMAND

The type of building, the population of clients that access the network, the patterns of usage, and the environment are a non-exhaustive list of factors that contribute to the spatial and temporal variation of traffic demand. In this section, we show how the modeled traffic variables of Table I vary across various time (hour, day, week) and spatial scales (building, building-type).

### *A. Variation of session-arrival rate within day/week*

Figure 1 plots the hourly session arrivals over the whole 2006 trace duration (192 hours) for some representative campus buildings. Although the absolute numbers of session arrivals and their exact variation are specific to each building, these profiles exhibit clear patterns that are, to a large extent, intuitive and closely related to the building type and usage. For example:

- Administrative and business buildings show strongly similar daily and weekly patterns in their profiles. The activity window is quite narrow during weekdays (6-8 hours long), in agreement with the working hours, whereas the activity during the weekend is almost zero.
- Residential buildings show distinctly different patterns. The number of session arrivals is more uniformly distributed across the week and hours within the day. The activity is also significant during the

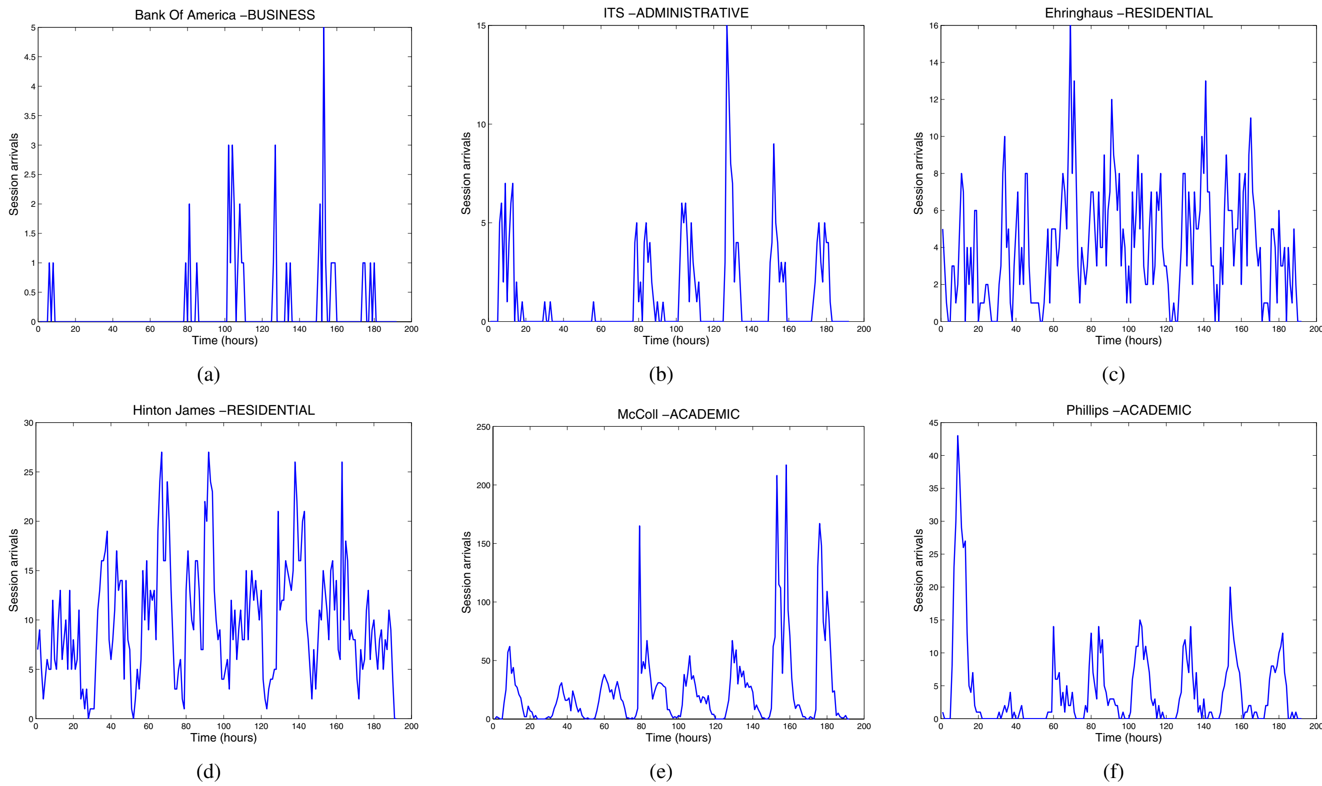



Fig. 1. Hourly session arrival rates for representative UNC campus buildings over the Apr-May 2006 monitoring period

TABLE III

BIPARETO DISTRIBUTION PARAMETERS FOR THE IN-SESSION NUMBER OF FLOWS IN DIFFERENT BUILDING TYPES

| Building type | Academic | Administrative | Athletic | Business | Clinical | Library | Residential | Social |
|---|---|---|---|---|---|---|---|---|
| BiPareto Parameters ($\alpha$, $\beta$, c, k) | (0.11, 2.15, 702.99, 1) | (0.15, 1.73, 523.72, 1) | (0.15, 1.65, 1033.84, 1) | (0.16, 2.39, 1008.76, 1) | (0.13, 2.6, 819.5, 1) | (0.06, 2.33, 862.24, 1) | (0.08, 1.34, 961.71, 1) | (0.09, 2.24, 571.21, 1) |

evening hours, often resulting in a daily or weekly peak.

- Academic buildings lie somewhere in between these two patterns. The daily window of activity is clearly broader than administrative and business buildings, since they host WLAN clients for longer time intervals during the day. Weekends see fewer session arrivals and shorter windows of activity when compared with residential buildings, but traffic is non-negligible.

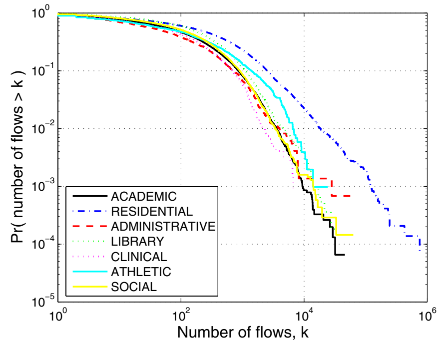
### B. Variation of session-level flow-related variables

The variation of traffic demand is also evident in the session-level variables. Their empirical distribution functions at the building-type level reflect this variation, as shown in Figure 2. Figure 2(a) shows the broad variation of the per building-type distribution tails of the in-session number of flows. The number of flows related to residential buildings sessions has a strikingly heavier tail, largely related to the more active Web browsing behavior of residential users. The plots also suggest that the BiPareto distribution can be applied for modeling the per building-type in-session number of flows. Table III lists the parameter sets for the different building types.

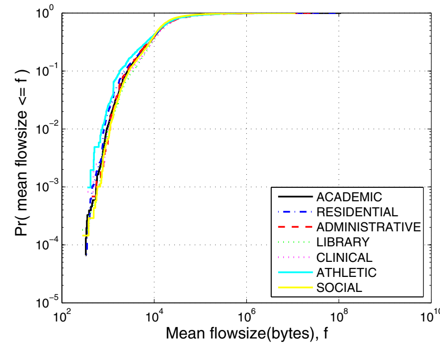
More similar are mean flow sizes across different building types. Figure 2(b) suggests that some building types cluster together, such as $\{Library, Residential\}$ and $\{Academic, Administrative, Athletic, Social\}$ with flow sizes in Clinical buildings having a more distinct behavior, yet closer to the second group of building types.

The behavior of flow interarrivals across different building types is captured in Figure 2(c). Again, the plots of mean in-session flow interarrivals suggest that the variables could be potentially modeled by the same type of distribution for all building types, though with different parameters.

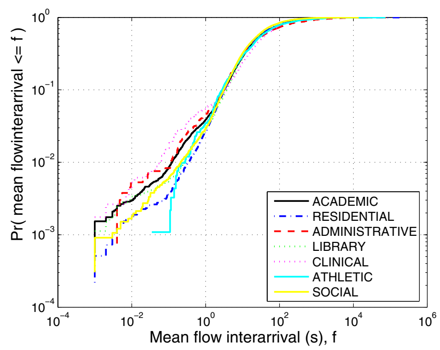
Less wide is the differentiation of client sessions with respect to mobility, at least when this is viewed at the building type level. *Building-roaming* sessions, during



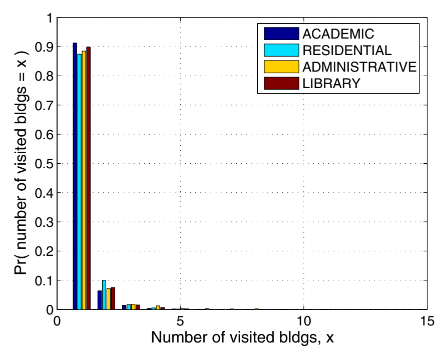

(a)




(b)




(c)




(d)

Fig. 2. Behavior of modeled session attributes across different types of campus buildings

which a WLAN client visits more than one building, account for less than 10% of the overall sessions. Figure 2(d) plots the per building type *building path length distribution*, expressing the probability that a session initiated at buildings of a certain type visits a given number of buildings. The plot clearly shows that building-roaming flows are a small percentage of the overall client sessions. Their overwhelming majority starts and finishes within the same building and this holds for all types of campus buildings.

Overall, the plots in Figure 2 show clearly that the modeled traffic variables exhibit strong variation in the spatial dimension. Although the building type is an intuitive, heuristical attribute for grouping buildings, it provides a base for unified treatment of the spatial dimension of the modeling task. The actual utility of this base is evaluated in Section V.

## V. Validation

### A. Methodology

This section compares the capability of models to capture traffic demand dynamics as we zoom in/out of the trace data and consider different levels of detail in our modeling. We have implemented in ns-2 a traffic generator that synthesizes traffic according to our models. We have configured the generator with the parameter sets that were estimated for different spatio-temporal slices of the whole network trace and compared the synthetic traffic with the real trace data. The comparison was made with respect to building-level traffic variables that were *not* explicitly addressed by our models. Such variables are the *flow arrival count process* and the *flow interarrival time-series* for the building under study. We examined first-order and second-order statistics of the flow interarrival process and hourly flow arrival counts.

The time-varying Poisson session arrivals are always modeled after the hourly building-specific data. For the flow-related variables, there are three alternatives regarding the (sub)set of trace data we consider in modeling. In increasing order of spatial aggregation, data may be specific to a single building, one of the eight building-types in Table II or, as comparison reference, the overall network (whole trace). At the time scale, the default is to use data from the entire trace. As a special case, we also extracted models by taking into account traces of a certain day. We employ these scenarios alternately in our simulations to illustrate our main findings. Each scenario poses different requirements in terms of variables that have to be modeled. Table IV summarizes these scenarios and the number of individual models (sampling distributions) that have to be implemented in the simulator when the entire wireless network is to be

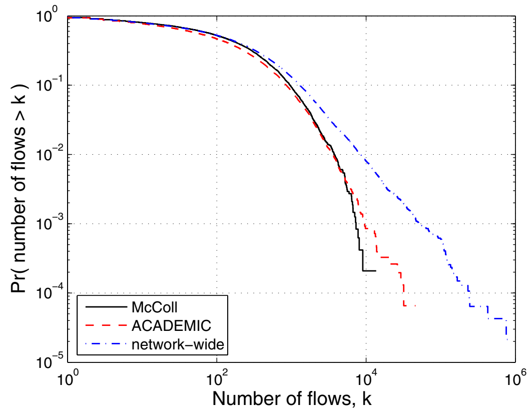



Fig. 3. Number of flows per session: ccdf under different building-grouping alternatives

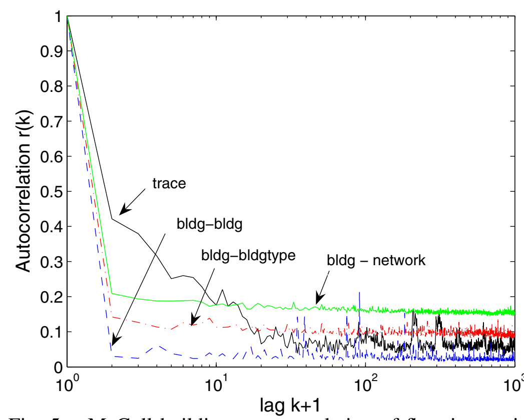



Fig. 5. McColl building: autocorrelation of flow interarrivals

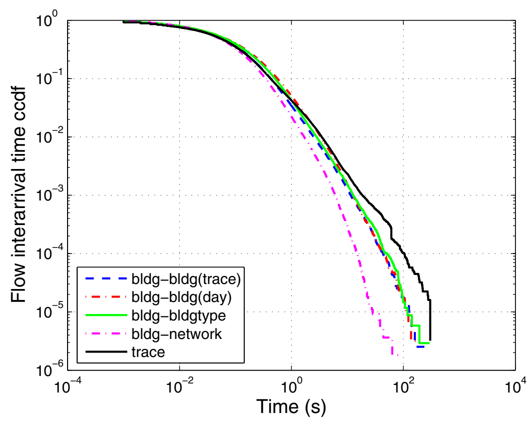



Fig. 4. McColl building: cumulative empirical distribution function of flow interarrivals

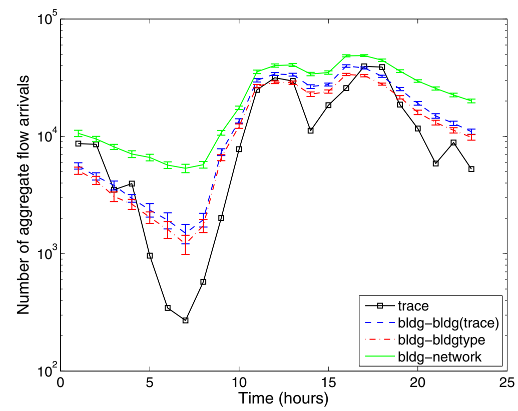



Fig. 6. McColl building: hourly flow arrival counts

simulated. Regarding the numbers in the third column, note that, irrespective of the modeling scenario, four variables always need to be modeled, namely, the session arrival process, and the three flow-related variables.

Given the heavy-tailed session durations, our simulation times are in the order of days rather than hours. We have implemented the thinning process described in [12] to simulate the time-varying Poisson process for the session arrivals.

Due to space constraints, we present results for two buildings, one academic (McColl) and one residential (Hinton James). These are two of the busiest campus buildings and represent the two main building types. At the same time, this evaluation reflects the performance of heuristics for achieving a good trade-off between model accuracy and scalability.

### B. *Accuracy considerations*

A first view of the "noise" that averaging introduces is given in Figure 3. The plot compares the cumulative empirical distribution function of the in-session number of flows for the McColl building with those estimated for all academic buildings and network-wide. The deviation between the curves increases with the degree of spatial aggregation of data. The way these discrepancies affect the flow-related metrics described in Section V-A is summarized in Figures 4-6 for the McColl and in Figures 7-9 for the Hinton James (HJ) building, respectively.

The synthetic traffic generator tracks most closely the trace when the in-session flow number and flow interarrivals are modelled separately for each one of the three days of simulation time. Notably, the results we get when we use a building-specific set of distributions derived from the whole trace, are not much better than when the respective distributions are extracted from the set of academic buildings. Figure 6 shows clearly that the aggregation in the time-dimension (whole trace rather than day) may cancel out the benefit of getting higher spatial resolution (building rather than building type) out of the trace data. At the same time, staying at the building-type level gives us comparable precision with that obtained when zooming into the building-specific data. Likewise, Figures 4 and 7 suggest that reuse of the network-wide distributions for modeling traffic demand at finer spatial scales is an even worse alternative. Despite the simplicity related to it, the averaging results

TABLE IV

MODELING ALTERNATIVES-SCENARIOS FOR SIMULATION VALIDATION

| Modeling scenario | Description | Sampling distributions |
|---|---|---|
| bldg-bldg(trace/day) | Session arrivals are modeled after bldg-specific data over the whole trace duration, flow-related variables after bldg-specific data for the whole trace duration (one day of the trace) | $4 \cdot N/(1+3 \cdot 8) \cdot N$<br>$N$ : number of bldgs |
| bldg-bldgtype | Session arrivals are modeled after bldg-specific data and flow-related variables over data aggregated at bldg type level | $N + 3 \cdot M$,<br>$M$ : number of bldg types |
| bldg-network | Bldg-specific data for session arrivals, network-wide distributions for the flow-related variables | $N + 3$ |

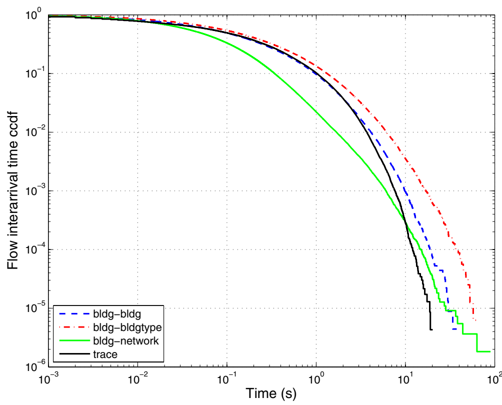


Fig. 7. Hinton James building: cumulative empirical distribution function of flow interarrivals

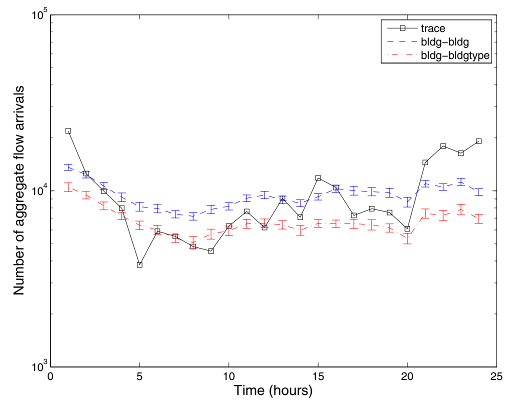


Fig. 8. Hinton James building: hourly flow arrivals

in even higher loss of information.

The differences between the autocorrelation processes in Figures 5 and 9 are minimal. The bldg-bldgtype curve is not much worse than the one corresponding to the bldg-bldg scenario. The deviation of all curves from the original trace points to the existence of a non-negligible second-order structure that is not taken into account in our models.

Finally, inferior in absolute terms is the match for the hourly flow counts, which is the most demanding and sensitive metric. Figures 6 and 8 plot the averages over 40 simulation runs along with their 95% confidence intervals. In this case, further improvement would be obtained by modeling the flow-related variables over shorter time periods than over the full monitoring period or a day. In fact, the standard practice is to focus the modeling attention on short time windows where the building activity experiences its peak (busy hour). In any case, the aggregation along the building type performs only marginally worse than the bldg-bldg scenario.

### C. Scalability considerations

The required number of sampling distributions for modeling each campus building under the bldg-bldg scenario would be $4 \cdot N$, where $N$ is the number of campus buildings. Repeating this for each single day of the trace would increase this number by a factor of $D$ the number of days.

When all buildings of the same type are modeled after a common set of distributions for flow-related variables, their number is reduced down to $N + 3 \cdot M$, where $M$ is the number of building types. Smaller values of $M$ can make the difference more dramatic and vice-versa. In general, $M$ acts as a tuning knob that can trade computing requirements with model accuracy and determines the complexity of the simulator.

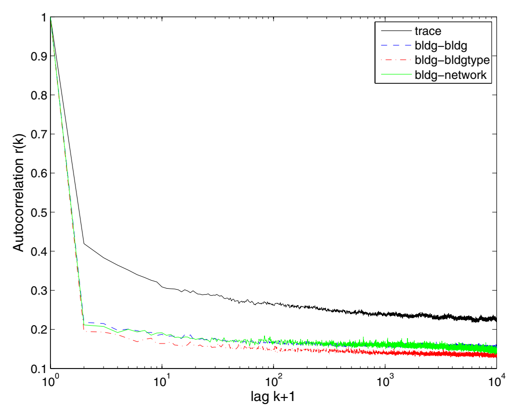


Fig. 9. Hinton James building: autocorrelation of flow interarrivals

## VI. RELATED WORK

Measurement-based studies of WLAN traffic demand have a much shorter history than those of the wired counterparts (e.g., [2]–[4], [13]). High-level observations about the temporal and spatial variation of the traffic demand appear in a number of papers [5]–[7], [14], which have drawn measurement data from different types of wireless infrastructures, namely, campus WLANs [5], [14], [15], enterprise WLANs [6], conference hotspots [7]). Traffic load diurnal/weekly periodicities have been noted in [5], [6], [16]. Almost all studies describe traffic demand variations amongst the monitored APs, with [5], [6], [17] also describing building and building-type dependencies.

Meng *et al.* were the first to model the WLAN traffic at higher detail, focusing on flow arrivals at 15 APs of the wireless infrastructure at Dartmouth campus in one-hour intervals [8]. They proposed a Weibull distribution, and captured the non-stationarity of traffic in the variation of its scale parameter, which is estimated via Weibull regression. Furthermore, they modeled flow sizes with a lognormal distribution. They found that a small percentage of the flows is roaming, *i.e.*, accessing data from more than one AP, and model the number of AP visits within an session with a geometrical distribution. They also observed strong similarity in the flow arrival processes at neighboring APs.

Our earlier work [9] focused on traffic demand at system-level. Contrary to [8], we captured the non-stationarity of traffic workload at the session- rather than the flow-level via a time-varying Poisson process for session arrivals. We believe that this hierarchical approach provides better insight to the underlying causes of the *temporal* variations of the workload. Moreover, we employed a larger set of data acquired from a larger number of APs, which allows us to see a significantly higher *spatial* variation of traffic load. Our work in this paper attempts to combine the precision of this first approach with the scalability properties that models should have when destined for use in large wireless infrastructures. Our work has shown that buildings and building types can be promising levels of spatial demand aggregation, in that the introduced averaging of the information in the spatial dimension does not sacrifice significantly the capacity of models to track the traffic load dynamics.

## VII. Conclusions

Our paper addresses the problem of traffic demand modeling in large wireless networks, emphasizing on the spatial dimension of the traffic load variation at various scales of spatial aggregation in the UNC wireless campus network.

In earlier work [9], we proposed a hierarchical modeling framework for aggregate network-wide traffic demand drawing on wireless sessions and network flows. This paper derives two notable results related to it. Firstly, we find out that the statistical distributions proposed for network-wide traffic demand, *i.e.*, time varying Poisson process for session arrivals, biPareto for in-session flow numbers and flow sizes, and Lognormal for in-session flow interarrivals, are valid over two different monitoring periods, spaced one year apart. Secondly, the same distributions apply when modeling traffic at finer spatial scales, such as individual buildings or groups of buildings with similar usage. Given the second result, we promote buildings as the primary entities for traffic demand modeling in the spatial dimension. Modeling at building level circumvents several problems emerging when working at AP-level, namely non-amenability to statistical processing, higher sensitivity of monitored traffic variables to the short-term propagation conditions, and lack of scalability.

We elaborate on this aspect and consider grouping buildings according to their use for modeling the wireless session-level flow-related traffic variables. The resulting accuracy-scalability trade-off is evaluated here with respect to characteristic traffic variables. The impact of this trade-off upon certain network functions will be ultimately determined via system-level simulations. Our current work is towards this direction.

## Acknowledgment

This work was partially supported by the General Secretariat for Research and Technology and by European Commission with a Marie Curie IRG grant.

## References

[1] V. Paxson and S. Floyd, "Wide area traffic: the failure of poisson modeling," *IEEE/ACM ToN*, vol. 3, no. 3, pp. 226–244, 1995.

[2] A. Feldmann, "Characteristics of TCP connection arrivals," in *in Self-Similar Network Traffic And Performance Evaluation (K. Park and W. Willinger, eds.).* John Wiley & Sons, 2000.

[3] W. S. Cleveland, D. Lin, and D. X. Sun, "IP packet generation: statistical models for TCP start times based on connection-rate superposition," in *Proc. of ACM Sigmetrics*, Santa Clara, CA, United States, June 2000, pp. 166–177.

[4] J. Cao, W. S. Cleveland, D. Lin, and D. X. Sun, "On the nonstationarity of internet traffic," in *Proc. of ACM Sigmetrics*, Cambridge, MA, United States, June 2001, pp. 102–112.

[5] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campuswide wireless network," in *Proc. of ACM MobiCom*, Philadelphia, PA, United States, September 2004.

[6] M. Balazinska and P. Castro, "Characterizing mobility and network usage in a corporate wireless local-area network," in *Proc. of MobiSys*, San Francisco, CA, United States, May 2003.

[7] A. Balachandran, G. Voelker, P. Bahl, and V. Rangan, "Characterizing user behavior and network performance in a public wireless LAN," in *Proc. of ACM Sigmetrics*, CA, June 2002.

[8] X. G. Meng, S. H. Y. Wong, Y. Yuan, and S. Lu, "Characterizing flows in large wireless data networks," in *Proc. of ACM MobiCom*, New York, NY, United States, 2004, pp. 174–186.

[9] F. Hernandez-Campos, M. Karaliopoulos, M. Papadopouli, and H. Shen, "Spatio-temporal modeling of traffic workload in a campus WLAN," in *Second Annual International Wireless Internet Conference*, Boston, MA, USA, 2006.

[10] UNC-FORTH Archive of Wireless Traces, Models, and Tools. [Online]. Available: http://www.cs.unc.edu/Research/mobile/datatraces.htm

[11] M. Karaliopoulos, E. Raftopoulos, M. Papadopouli, and H. Shen, "On scalable measurement-driven modeling of traffic demand in large WLANs," ICS-FORTH, Heraklion, Greece, Tech. Rep. 383, July 2006. [Online]. Available: http://www.ics.forth.gr/ftp/tech-reports/2006/2006.TR383_WLAN_Scalable_Traffic_Modeling.pdf

[12] P. Lewis and G. Shedler, "Simulation of nonhomogeneous poisson process by thinning," *Naval Research Logistics Quarterly*, vol. 26, pp. 403–413, 1979.

[13] V. Paxson and S. Floyd, "Wide-area traffic: the failure of Poisson modeling," in *Proc. of ACM Sigcomm*, London, United Kingdom, August 1994, pp. 257–268.

[14] F. Hernandez-Campos and M. Papadopouli, "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, September 2005.

[15] F. Chinchilla, M. Lindsey, and M. Papadopouli, "Analysis of wireless information locality and association patterns in a c ampus," in *Proc. of IEEE Infocom*, Hong Kong, March 2004.

[16] M. Papadopouli, H. Shen, E. Raftopoulos, M. Ploumidis, and F. 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.

[17] 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, September 2005.