# Synthetic traffic generation based on Measurement-driven modeling of large Wireless Local Area Networks

Manolis Ploumidis[a,b*] ploumid@ics.forth.gr
Elias Raftopoulos[a,b†] eraftop@ics.forth.gr
Maria Papadopouli[a,b] mgp@ics.forth.gr

August 10, 2007

## 1 Introduction

Wireless local area networks (WLANs) are increasingly being deployed to address the growing demand for wireless access. Contrary to traditional wired-network topologies that reflect the physical hardwired connection of routers, wireless network topologies are more dynamic and have a stochastic element due to the radio propagation conditions, the user mobility and client-AP association process. To improve current best-effort services offered by WLANs, several mechanisms, such as, capacity planning, link adaptation and load balancing, have been proposed. For their performance analysis, models of the network and user activity are critical. Unfortunately, current simulators use simplistic traffic and mobility models [8].

Eager to provide more accurate models for wireless user activity we performed empirical-based analysis and modeling studies in a production wireless network [5, 4]. Several questions stimulated this research effort: How do users arrive at APs? How do they roam across APs? How are flows generated at APs? What are their temporal dynamics? What are the right structures to model the user-initiated activity in a wireless network?

One of the most intriguing aspect of modeling the wireless demand is its multi-level spatio-temporal nature, namely the different spatial scales (e.g., infrastructure-wide, AP-level or client-level) and time granularities (e.g., packet-level, flow-level and session-level) inherent in the task. We proposed to capture the user-initiated activity through flows, sessions, i.e., episodes of continues wireless access in the infrastructure, and disconnections. Furthermore, we have presented a methodology for modeling the demand and specifically, the client associations at APs, sessions, and flows [1]. According to our methodology, session arrivals are modeled through a time-varying Poisson process while flow interarrivals within sessions follow a lognormal distribution. Number of flows per session and flow sizes can be accurately modeled by a BiParetto distribution. Finally, the AP-preference distribution can be accurately described by a lognormal distribution. The main contribution of this work is the implementation of a synthetic traffic generator and a testbed to illustrate the use of our models. Finally, we evaluate the validity of the synthetic trace using a well known protocol as reference metric.

## 2 Background

In our study we perform a multi-level spatio-temporal modeling, at different spatial scales, such as, infrastructure-wide, AP- and client-level, and time granularities such as packet-level, flow-level, and session-level. Our approach distinguishes two important dimensions in wireless network modeling, namely the user demand, i.e.,

*a. Department of Computer Science, University of Crete
†b. Institute of Computer Science, Foundation for Research and Technology-Hellas

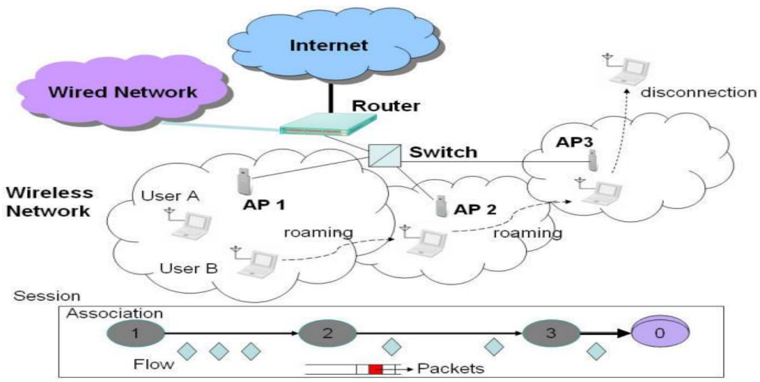


Figure 1: UNC wireless infrastructure.

user-initiated activity through flows and sessions, and the topology, i.e., network, infrastructure, and radio propagation dependencies. This enables us to "superimpose" models for the demand on a given topology, and focus on the right level of detail for the performance analysis or simulation study (e.g., AP-level, system-wide, client-level). This methodology "masks" network-related dependencies that may not be relevant to a range of systems and makes the wireless networks amenable to statistical analysis and modeling. Key structures of our modeling are the WLAN client associations and the traffic flows.

As Figure 1 shows, a wireless session can be viewed as a sequence of association and disassociation events with the APs of the infrastructure [4]. On the other hand, network flows are collections of packets sharing the same TCP "4-tuple". Sessions are statistically well-behaved, and, most significantly, robust to network dependencies. In our approach, sessions represent the high level unit of wireless network traffic load, including all flows initiated by clients of the wireless infrastructure with other Internet-side hosts. On the other hand, network flows provide a finer level of modeling the packet-level workload in a network-independent manner [7, 6]. Flows consist an important structure above the packet-level for network traffic analysis and closed-loop traffic generation.

We make use of parametric models to describe both session and flow-level traffic variables at different spatial scales. When compared with empirical models, they provide better insight to the underlying patterns of the modeled quantities. We therefore propose statistical distributions for the session and flow level-traffic variables [1].

## 2.1 Contributions

The main contribution of this work is the implementation of a synthetic traffic generator using the session and flow related models [2]. Besides this, we implemented a testbed to evaluate our synthetic traffic generator and compared our synthetic traces with real ones and traces produced using current simplistic models. Our parsimonious description of the workload in a WLAN, can be directly employed in simulation and testbed experimentation studies to generate more realistic user behaviors, while it allows for a better insight to the problem than empirical models. The network load can be simulated at both the client association and flow levels. Sessions have a well-behaved arrival process, which can be accurately described using a time-varying Poisson process [5]. In addition, our AP preference distribution can be used as a first rough approximation for distributing sessions throughout the wireless infrastructure in a manner that is representative of real workloads. Our proposed models and their parameters are summarized in Figure 2. Simulation scenarios can first produce a time series for the session arrival process, and then sample the distributions of the number of flows and their interarrivals to generate the within-session flow arrival time series. Flow sizes are also generated based on the proposed distribution.

| Modeled variable | Model | Probability Density Function (PDF) | Parameters |
|---|---|---|---|
| Session arrival | Time-varying Poisson with rate | N: #Number of Sessions between $t_1$ and $t_2$ $\lambda - \int_{t1}^{t2} \lambda(t)dt, \Pr(N - n) - \frac{e^{-\lambda}\lambda^n}{n!}, n - 0,1,\ldots$ | Hourly rate: 44(min), 1132(max), 294(median) |
| AP of first association/session | Lognormal | $p(x) = \frac{1}{\sqrt{2\pi x\sigma}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right]$ | $\mu = 4.08, \sigma = 1.44$ |
| Flow interarrival/session | Lognormal | Same as above | $\mu = -1.36, \sigma = 2.78$ |
| Flow number/session | BiPareto | $p(x) = k^{\beta}(1+c)^{\beta-\alpha} x^{-(\alpha+1)} (x+kc)^{\alpha-\beta-1} (\beta x + \alpha kc), x \geq k$ | $\alpha = 0.06, \beta = 1.72, c = 284.79, k = 1$ |
| Flow size | BiPareto | Same as above | $\alpha = 0.00, \beta = 0.91, c = 5.20, k = 179$ |

Figure 2: Parametric models for the session and flow related variables.

## 2.2 Simulation testbed

Our proposed testbed illustrates the generation, validation and manipulation of synthetic traces in order to provide a more robust and flexible foundation for simulation studies. We have implemented a matlab module that produces synthetic traces based on our parametric statistical models. The user can adjust:

- the duration and scale of the trace
- the client arrivals volume
- the number of flows generated per user initiated session
- the corresponding flow sizes

Our traffic generator ensures that the synthetic trace follows the respected models and provides data that are easy to manipulate, process, and incorporate in any simulation study. Moreover, a validation mechanism is introduced to evaluate the efficiency of our proposed synthetic traffic generator [2]. We make use of the Ns-2 simulator to replay different kind of traces. Figure 3 represents the Ns-2 scenario used to validate our models. Three wireless clients initiate TCP flows through the AP at which they are associated exchanging packets with one wired host.

The synthetic data produced are compared both against the UNC original trace, and naive models, which are widely used by the research community for simulation purposes. The metric used to evaluate the synthetic trace when compared to the real one, is TCP performance over the wireless channel as in [3], which is an independent parameter to the original modeled variables and consists a realistic wireless setting. We focus on TCP throughput and delay measured at the corresponding simulation runs. We employ Ns-2 for evaluation purposes, using three different types of simulation data, namely:

- the synthetic dataset, produced based on out two-tier simulation approach
- the real trace, captured at UNC consisting of packet headers and SNMP data
- naive models, widely used for benchmarks by the community

We show that simulated data based on our proposed models match well the original data as opposed to naive models which fail to capture the spatio-temporal characteristics of wireless user demand. Interestingly, our

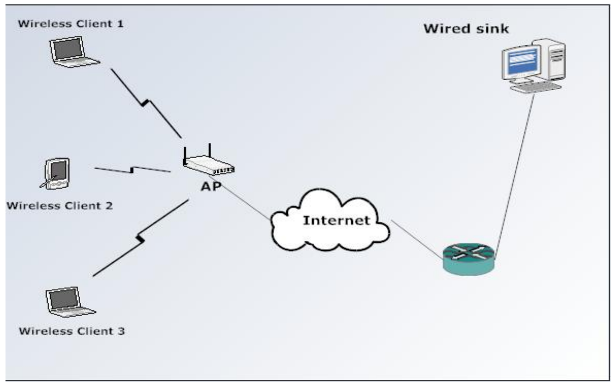


Figure 3: Ns2 Wireless scenario.

synthetic traces capture traffic demand characteristics in individual hotspot APs, implying that they can be used for modeling traffic workload over finer levels of spatial aggregation.

The equipment needed perform the demo is a laptop equipped with matlab and Ns-2. No set-up time or Internet connectivity is required.

## 2.3 Data Repository

Both UNC and FORTH have state-of-art testbeds for monitoring large scale wireless networks and collecting extensive wireless traces, such as, SYSLOG, SNMP, TCP flow, and signal-strength-based data. A data repository with these traces, monitoring tools, and models is the result of a joint effort of the Mobile Computing Group in the Department of Computer Science at UNC and the Institute of Computer Science at FORTH. Such traces enable comparative analysis modeling and validation studies using different wireless networking environments. The repository has been made publicly available at http://www.ics.forth.gr

# References

[1] F. Campos, M. Karaliopoulos, M. Papadopouli, and H. Shen. Spatio-temporal modeling of traffic workload in a campus wlan. In *Second annual international WIrless internet CONference ( WICON06 )*, Boston, USA, 2006.

[2] M. Karaliopoulos, M. Papadopouli, E. Raftopoulos, and H. Shen. On scalable measurement-driven modelling of traffic demand in large wlans. In *IEEE Workshop on Local and Metropolitan Area Networks*, Princeton NJ, USA, 2007.

[3] X. Meng, S. Wong, Y. Yuan, and S. Lu. Characterizing flows in large wireless data networks. In *ACM International Conference on Mobile Computing and Networking (MobiCom)*, Philadelphia, USA, 2004.

[4] 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.

[5] 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.

[6] V. Paxson. Empirically-derived analytic models of wide-area TCP connections. *IEEE/ACM Transactions on Networking*, August 1994.

[7] V. Paxson and S. Floyd. Wide-area traffic: the failure of Poisson modeling. In *SIGCOMM Symposium on Communications Architectures and Protocols*, pages 257–268, London, United Kingdom, August 1994. ACM.

[8] S.Borst. User-level performance of channel-aware scheduling in wireless data networks. In *Proceedings of the IEEE Conference on Computer Communications (Infocom)*, San Francisco, California, USA, April 2003.