**EIFFEL Position Paper on**

**NETWORK MONITORING, MEASUREMENTS & MODELING**

**Editor: Maria Papadopouli (draft, 4/18/2010)**

**I. Introduction**

Heterogeneous wired and wireless networks of various technologies have been deployed to provide Internet access. Especially, the deployment of wireless networks in universities, corporations, communities, and metropolitan areas has been rapidly expanded. It is critical to understand their performance and workload, in order to develop networks that are more robust, easier to manage and scale, and more able to efficiently utilize their scarce resources. While in several cases over-provisioning in wired networks is acceptable, it can become problematic in the wireless domain. A number of mechanisms, such as capacity planning, dimensioning, resource reservation, device adaptation, and load balancing, need to be employed to support such networks. Real-life measurement studies can be particularly beneficial in the development and analysis of such mechanisms, as they can uncover deficiencies of the technologies and different phenomena of the access and workload.

As more traces from various wireless network environments become available, it is critical to develop measurement methodologies, benchmarks, and tools for searching for law-like relationships across these different traces that can be generalized to a wide range of different conditions. The availability of measurements and benchmarks can play a dramatic role in comparative performance analysis and (cross-)validation studies through repeatability. The existence of testbeds, tools, and benchmarks is of tremendous importance. Rich sets of data can impel modeling efforts to produce more realistic models, and thus, enable more meaningful performance analysis studies. A non-exhaustive list of general metrics that motivate the measurement studies, namely, *application performance*, *traffic demand*, *human mobility, network topology, fault tolerance/robustness.*

As popular applications and services from wired networks shift to the wireless arena, new applications emerge, and the use of wireless-enabled devices evolves rapidly, it would be interesting to perform comparative analysis of traces collected from various networking environments. In general the performance of an application has been characterized based on the application requirements, such as in terms of network benchmarks, data resolution and media quality, user-perception metrics (e.g., MOS or A/B), interactivity model, usage pattern, and traffic demand.

Network benchmarks, such as jitter, latency, and packet loss, have been used to quantify network performance. However, what is their impact on how a user “perceives” the performance of a certain application? It is important to understand which network performance characteristics have dominant impact on the performance of certain applications. Shifting the attention from MAC- and network-based metrics to application-based characteristics, the quantification of the user satisfaction and application requirements with more formal subjective and objective metrics/benchmarks is required. In addition, two other related issues need to be also addressed:

* distinguish the conditions that degrade substantially the performance of a given application
* investigate the predictability of these conditions

The user workloadis another important parameter that needs to be measured, whichincludes the amount of traffic (e.g., in terms of number of packets or bytes, downloading vs. uploading), packet or flow arrival process, interactivity model, application type, and usage pattern. Understanding how user behavior changes depending on the network topology, network performance, technology characteristics and human mobility is another challenging issue that has not been well-studied.

Network topologies can be described based on their connectivity and link characteristics, distribution and density of peers, degree of clustering, co-residency time, inter-contact time, duration of disconnection from the Internet, and interaction patterns. In contrast 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. Modeling wireless network topologies opens up new research directions.

Measurements and modeling of human mobility also brings more challenges. One of the problems is that complex mobility and topology models are rich sub-fields of their own expertise. There should be tools and methods for others to effectively and easily use models from these sub-fields in standard simulators. Depending on the environment/use case, the human mobility can be group or in individual, spontaneous or controlled, pedestrian or vehicular, known *a priori* or dynamic.

Finally, criteria that describe the robustness, scalability, and fault-tolerance include the number of active neighboring devices, the degree of vulnerability under the loss of valuable links or APs, and the impact of induced failures on the performance. Network conditions can be characterized by link quality criteria (e.g., packet losses, delays, SINR), the spatio-temporal distributions of traffic demand and application mix, and the distributions of regions of weak connectivity or no signal.

Section II presents a brief overview of network modeling, focusing mostly on traffic modeling. It will also discuss the main modeling objectives. Section III discusses monitoring issues, while Section IV focuses on the availability of empirical measurements for analysis and modeling.

**II. Modeling**

A large body of literature has developed concepts and techniques for modeling Internet traffic, especially in terms of statistical properties (e.g., heavy-tail, self-similarity). For example, heavy-tailed distributions appear in the sizes of files stored on web servers [1], data files transferred through the Internet [2], and files stored in general-purpose Unix filesystems, suggesting the prevalence and importance of these distributions. Self similarity characteristics exist in Internet traffic. In a pioneering work, Leland et al. showed that LAN traffic exhibits a self-similar nature [3]. Evidence of self-similarity was also found in WAN traffic [4]. In that work, Paxson and Floyd demonstrated that self-similar processes capture the statistical characteristics of the WAN packet arrival more accurately than Poisson arrival processes, which are quite limited in their burstiness, especially when multiplexed to a high degree. Self-similar traffic does not exhibit a natural length for its “bursts”. Its traffic bursts appear in various time scales [3]. The relation of the self-similarity and heavy-tailed behavior in wired LAN and WAN traffic was analyzed by Willinger et al. [5]. On the other hand, Poisson processes can be used to model the arrival of user sessions (e.g., telnet connections and ftp control connections). However, modeling packet arrivals in telnet connections by a Poisson process may result in inaccurate delay characteristics, since packet arrivals are strongly affected by network dynamics and protocol characteristics. Web traffic exhibits also self-similarity characteristics. Crovella and Bestavros showed evidence of this and attempted to explain them in terms of file system characteristics (e.g., distribution of web file size, user preference in file transfer, effects of caching), user behavior (e.g., “think time” accessing a web page), and the aggregation of many such flows in a LAN [6]. The majority of web traffic in wired networks is below 10KB while a small percentage of very large flow account for 90% of the total traffic. They employed power laws to describe web flow sizes. Similar phenomena were also observed similar phenomena in the campus-wide wireless traffic. A nice discussion of the use of power law and lognormal distributions in other fields can be found in [7]. While there is rich literature on traffic characterization in wired networks (e.g., [10-13]), there is significantly less work of the same depth for WLANs (e.g., [14-16]

A typical evolution of a technology consists of the following steps:

1. Simple simulations
2. Advanced and more realistic simulations
3. Emulations and tests in small-scale, often controlled testbeds
4. Tests in large-scale testbeds
5. Adoption and use in production networks

Thus it is common practice for a preliminary evaluation of a technology to explore its behavior under well-understood conditions and simple models. For example, most of the performance analysis studies on wireless network protocols and mechanisms employ traffic models to simulate saturation conditions (asymptotic behavior). There are only few studies that employ stochastic packet-rate models or replay real-life traces. Most of the simulators use quite simplistic models, as mobility, topology, access, and traffic models are rich sub-fields on their own. However, for more comprehensive performance analysis studies, sets of measurements are collected from production networks for statistical analysis and modeling.

In general, models should have the following properties:

* Accuracy
* Robustness
* Scalability
* Parsimony
* Reusability
* “Easy” interpretation

Often, we can provide models for various spatial or temporal scales, e.g., based on data that were captured at the level of APs vs. network-wide or at a certain location vs. metropolitan-area, and different time periods. By selecting the appropriate spatio-temporal granularities of the models, the right balance between reusability and accuracy can be addressed. Often models at a very fine spatial or temporal detail can be very accurate at the cost of a lower scalability and amenability. When models are based on a larger scale (e.g., network-wide scale), there is a gain of simplicity at the cost of a higher loss of detail.

Often in simulations, networks are “formed” using some unrealistic or even incorrect assumptions. In the case of wireless networks, often performance analysis studies make the following assumptions:

* wireless links are symmetric
* link conditions are static
* the density of devices in an area is uniform
* the traffic demand and access patterns are fixed
* the communication pairs (i.e., source and destination devices) are fixed
* users move based on a random-walk model

In most of the cases, these assumptions are unrealistic and incorrect. For instance, it is known that, in general, the spatial distribution of network nodes moving according to the random waypoint model is non-uniform (e.g., [8]). Moreover, wireless channels can be highly asymmetric and highly time-varying. Unfortunately, there are not many traces of actual data access patterns or realistic models available for wireless users, especially for mobile peer-to-peer settings (e.g., [9]).

**III. Monitoring**

The identification of the appropriate parameters that need be analyzed in order to characterize the specific condition of interest is an important first step in the monitoring process. This will determine at which layers and network points and in which spatio-temporal granularities the monitoring needs to take place. In general, the monitoring techniques can be classified into various categories with respect to the following aspects:

* Use of active vs. passive probing
* User-centric vs. operator-centric vs. administrator-centric perspective
* Placement of monitors
* Spatio-temporal granularity of the collected data
* Data processing and analysis methodology

In user-centric, users may enable their devices to collect measurements and upload these measurements in a central data repository for further analysis. On the other hand, in the operator-centric approaches, the operator initiates the monitoring and data collection. Typically, the number of monitors and their network locations is restricted. The monitoring can take place from the perspective of a user, client device, group of clients (with the same profile), AP, LAN or network domain, entire infrastructure.

Extensive monitoring and collection of data in fine spatio-temporal detail can improve the accuracy of the performance estimates, but also increase the energy consumption and detection delay, as the network interfaces need to monitor the channel over longer time periods and then exchange this information with other devices. Some important aspects that need to be addressed are:

* identification of the dominant parameters through sensitivity analysis studies
* strategic placement of monitors at routers, APs, clients, and other devices
* automation of the monitoring process to reduce human intervention in managing the monitors and collecting data
* aggregation of data collected from distributed monitors to improve the accuracy, while maintaining a low communication and energy overhead
* (cross-)validation study to verify that the collected traces correspond to representative conditions

Often monitoring tools are not without flaws and several issues arise when they are used in parallel for thousands devices of different types and manufacturers. They are limited in their capabilities because they cannot capture all the relevant information due to either hardware limitations, the proprietary nature of hardware and software, or hidden terminals. Furthermore, monitors are subject to issues related to:

* Fine-grain data sampling
* time synchronization
* incomplete information
* data consistency
* vendor-specific information and dependencies, often not publicly available

 The monitoring methodology varies based on how it deals with various inconsistencies across the data collected from different monitors due to synchronization issues (e.g., different clocks), errors, incomplete values, different data due to the different drivers/hardware capabilities (e.g., RSSI values).

Unfortunately, there are not many traces of actual data access patterns or realistic models available for wireless users, especially for mobile peer-to-peer settings (e.g., [9]). Often academics are reluctant to expend the time and energy required to “sanitize” the data sets. Similarly, companies are not eager to disclose information they consider proprietary. The development of realistic, but also general, tractable and elegant models is a non-trivial task.

**IV. Empirical-based measurements, analysis and modeling**

 Often academics are reluctant to expend the time and energy required to “sanitize” the data sets. Similarly, companies are not eager to disclose information they consider proprietary. The development of realistic, but also general, tractable and elegant models based on empirical measurements is a non-trivial task.

It is important to note that the realism of an empirical or synthetic trace depends tightly on the system to be studied. Highlighting the ability of empirically-based models to capture the characteristics of a certain condition (e.g., with respect to workload, topology, mobility) and providing a flexible framework for using them in performance analysis studies is crucial. At the same time, the generation of synthetic traces based on some models to reflect certain network condition, especially in non-simplistic network topologies/architectures can be a particularly hard problem. Furthermore, the scaling properties of synthetic traces and simulators are very important and have not been fully addressed. For example, it is not clear that a simple 20-node simulation can be “stretched” to 10,000 node simulations by a “copy-and-paste” methodology.

Researchers from different disciplines, e.g., statisticians, mathematicians and computer-scientists can have particularly fruitful collaborations in analyzing and modeling measurements. A first step in a statistical analysis of real-life measurements includes the treatment of potentially sparse data/signal as well as missing values and outliers. The approach that will be used depends on the particular objectives of the modeling study.

The maintenance of a database with models and benchmarks for traffic load, mobility patterns, network topologies and representative scenarios can be particularly helpful in cross-validation in performance analysis studies. In such case, for example, a researcher can use benchmarks to test the performance of a proposed mechanism under different traffic load (e.g., heavy vs. normal vs. light or heavy p2p traffic demand vs. web browsing vs. email access patterns) and wireless link (e.g., different packet loss or SNR characteristics of the channel) conditions and compare its performance with other related mechanisms.

**References**

[1] Mark E. Crovella and Azer Bestavros. Self-similarity in world-wide-web traffic: Evidence and possible causes. IEEE/ACM Transactions on Networking, 5(6):835–846, December 1997.

[2] Vern Paxson. Empirically-derived analytic models of wide-area TCP connections. IEEE/ACM Transactions on Networking, 2(4):316–336, August 1994.

[3] Will E. Leland, Murad S. Taqq, Walter Willinger, and Daniel V. Wilson. On the self-similar nature of Ethernet traffic. In Deepinder P. Sidhu, editor, ACM Symposium on Communications Architectures and Protocols (SigComm), pages 183–193, San Francisco, California, September 1993. ACM. also in Computer Communication Review 23 (4), Oct. 1992.

[4] Vern Paxson and Sally Floyd. Wide-area traffic: The failure of Poisson modeling. IEEE/ACM Transactions on Networking, 3(3), June 1995.

[5] W. Willinger, V. Paxson, and M. Taqqu. Self-similarity and heavy tails: Structural modeling of network traffic. In R. Adler, R. Feldman, and M. Taqqu, editors, A Practical Guide to Heavy Tails: Statistical Techniques and Applications, pages 27–53. Birkhauser, 1998.

[6] Mark Crovella and Azer Bestavros. Self-similarity in world-wide-web traffic: Evidence and possible causes. In Proceedings of SIGMETRICS, Philadelphia, PA, May 1996

[7] Michael Mitzenmacher. A brief history of generative models for power law and lognormal distributions. Internet Mathematics, 2003.

[8] Christian Bettstetter, Giovanni Resta, and Paolo Santi. The node distribution of the random waypoint mobility model for wireless ad hoc networks. IEEE Transactions on Mobile Computing, 2(3):257–269, July 2003.

[9] Amit Jardosh, Elizabeth M. BeldingRoyer, Kevin C. Almeroth, and Subhash Suri. Towards realistic mobility models for mobile ad-hoc networks. In ACM International Conference on Mobile Computing and Networking (MobiCom), San Diego, CA, September 2003.

[10] W. Willinger, Taqqu M.S., R. Sherman, and D.V. Wilson. Self-similarity through high-variability: Statistical analysis of ethernet LAN traffic at the source level. ACM Computer Communication Review, 25(4):100–113, October 1995.

[11] Paul Barford and Mark E. Crovella. Generating representative Web workloads for network and server performance evaluation. In ACM Sigmetrics Conference on Measurement and Modeling of Computer Systems, pages 151–160, Madison, Wisconsin, June 1998.

[12] William S. Cleveland, Dong Lin, and Don X. Sun. IP packet generation: statistical models for TCP start times based on connection-rate superposition. In ACM Sigmetrics Conference on Measurement and Modeling of Computer Systems, pages 166–177, Santa Clara, CA, United States, June 2000.

[13] Jin Cao, William S. Cleveland, Dong Lin, and Don X. Sun. On the nonstationarity of Internet traffic. In ACM Sigmetrics Conference on Measurement and Modeling of Computer Systems, pages 102–112, Cambridge, MA, USA, June 2001.

[14] Xiaoqiao Meng, Starsky Wong, Yuan Yuan, and Songwu Lu. Characterizing flows in large wireless data networks. In ACM International Conference on Mobile Computing and Networking (MobiCom), pages 174–186, Philadelphia, PA, September 2004.

[15] Maria Papadopouli, Haipeng Shen, and Manolis 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] Merkouris Karaliopoulos, Maria Papadopouli, Elias Raftopoulos, and Haipeng Shen. On scalable measurement-driven modelling of traffic demand in large WLANs. In IEEE Workshop on Local and Metropolitan Area Networks, Princeton NJ, USA, June 2007.

Note: This draft is heavily based on the monograph on Maria Papadopouliand Henning Schulzrinne. Peer-to-peer Computing for Mobile Networks: Information Discovery and Dissemination. Springer Eds., ISBN: 978-0-387-24427-3, 2009.