

Measuring networks for enhancing user experience (MUSE)

Renata Teixeira
Inria Paris-Rocquencourt

1 Introduction

The last decade has seen a drastic popularization of access to Internet applications and services. People rely on the Internet for everyday tasks ranging from business to entertainment. In many households, the Internet has completely replaced traditional means to deliver phone, TV, or video-on-demand. As our lives become more dependent on connectivity, it is easy to understand people’s frustration when their Internet experience is not up to their expectation — for example, when their download is too slow, when their voice-over-IP call is choppy, or when the movie they have paid for keeps stalling. Most importantly, most Internet users are not tech-savvy and hence cannot fix performance problems and anomalous network behavior by themselves [3]. The complexity of most Internet applications makes it hard even for networking experts to fully diagnose and fix problems. This illustrates a bigger problem: the increasing gap between the complexity of our networked systems and everyday users’ abilities to operate these systems when poor network performance or anomalies occur.

The goal of the MUSE¹ research group at Inria is to improve user online experience by building personalized networking technology that guides network performance and diagnosis based on user quality of experience. A key challenge of this research is to infer the user quality of experience based solely on the observation of systems/network metrics. The vast majority of network performance studies have focused on well-established network performance metrics (round-trip time, loss rate, and throughput) rather than how these metrics relate to user perception. Understanding the more subjective quality of experience has been limited to few applications, such as VoIP [5], online-gaming [1], and video playback [10]. These studies are often based on the Mean Opinion Score (MOS). Building application-specific models is useful in improving the performance of individual applications, but we cannot easily generalize these models to capture a particular user’s online experience, which involves multiple applications and contexts over time.

Understanding how users react to different levels of network performance is challenging. User perception is subjective and contextual making it hard to build analytical models. Different users may have different tolerance levels to network performance and the same user may have different expectations under different circumstances. As a result, we take an empirical approach to measure network performance annotated with user perception and then use the resulting datasets to build empirical models. Unfortunately, it is challenging to “measure” users. We must develop methods to measure the user perception of network performance as users perform their routine online tasks. It is hence important that these methods are not too intrusive. Otherwise, users are unlikely to participate in the experiment. In addition, we must capture perception of a large number of users at different levels of performance and in a variety of scenarios.

This white paper briefly describes our experiences building HostView to measure quality of experience and on using the HostView dataset to build predictors of user dissatisfaction with network performance. We then describe

¹MUSE stands for measuring networks for enhancing user experience.

2 HostView

We have designed HostView [8, 9], an end-host tool that collects network performance data annotated with user feedback on the quality of experience with online applications. HostView collects network packet traces and application process executable names, which we can use to infer network performance per application, as well as a number of end-host performance metrics such as WiFi signal strength/noise and machine CPU load and memory utilization. HostView also records user feedback on application performance with two mechanisms: a system-triggered questionnaire based upon the Experience Sampling Methodology (ESM) [2], and a user-triggered mechanism called an “I’m annoyed!” button. The ESM mechanism prompts the user no more than three times per day with a questionnaire about their experience with online applications. The “I’m annoyed!” button is always displayed at the corner of the screen and we ask users to click on it when they are dissatisfied with their application performance. The initial version of HostView had approximately 100 downloads and we obtained data for more than two weeks with user feedback for 19 users.

We have used the HostView dataset to develop a method to predict user dissatisfaction with network performance (i.e., given only network performance data as input, can we predict whether the user is dissatisfied or not?) [6]. The main challenges of modeling user dissatisfaction with network performance comes from the scarcity of user feedback and the fact that poor performance episodes are rare. We developed a methodology to build training sets in face of these challenges. Our analysis of HostView data showed that non-linear support vector machines achieve high accuracy to solve the classification problem. In particular, these results showed that a predictor learned for one application does not apply to another. Hence, in order to accurately predict user dissatisfaction with one application we need labeled samples of this particular application to train predictors. In HostView data, we only have enough labeled samples if we combine data from multiple users and still we can only build models for few applications.

3 Future direction

We are currently building a new version of HostView and collaborating with sociologists (within the “User-Centric Networking”, UCN, project) to gain deeper insights into users online activities and their perception of network performance. In our on-going work we are improving a few aspects of HostView: the user experience sampling method, the performance and usability of the tool, and extending the data collection to other platforms including mobile devices and the browser. We plan to release the new version of HostView as part of the ethnographic study we are conducting within the UCN project and then release HostView for a larger number of users.

After we obtain richer datasets with samples of network performance labeled with user perception, then we plan to improve the predictors of user perception of network performance. Our future research will explore a number of open questions that must be addressed before we can incorporate these models into a system that will run online on users’ devices to automatically detect instances when quality of experience degrades. Does a model learned from one user apply to another? Do models vary per user context? How often do we need to re-train models? If models are specific for a given user, then how can we personalize models per user without requiring users to give explicit feedback (which may prevent wider adoption)? Another set of challenges relate to whether we can design lightweight predictors to run online without overloading the user’s device.

Acknowledgements

The original HostView work and the initial predictor of user dissatisfaction was the thesis of Diana Joumblatt supervised in collaboration with Nina Taft and Jaideep Chandrashekar. George Rosca is developing the new version of HostView co-supervised by Anna-Kaisa Pietilainen. The new user study we are conducting with the UCN project is in collaboration with Peter Tolmie, Tom Rodden, Tom Lodge, Richard Mortier, and Murray Goulden.

This project is partially supported by the European Community's Seventh Framework Programme (FP7/2007-2013) no. 611001 (User-Centric Networking).

References

- [1] Kuan-Ta Chen, Polly Huang, and Chin-Laung Lei. How sensitive are online gamers to network quality? *Commun. ACM*, 49(11):34–38, 2006.
- [2] S. Consolvo and M. Walker. Using the Experience Sampling Method to Evaluate Ubicomp Applications. *IEEE Pervasive Computing Magazine*, 2(2), 2003.
- [3] R. E. Grinter, W. K. Edwards, M. Chetty, E. Poole, J.-Y. Sung, J. Yang, et al. The ins and outs of home networking: The case for useful and usable domestic networking. *ACM Transaction on Computer-Human Interaction*, 16(2), 2009.
- [4] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2009.
- [5] W. Jiang and H. Schulzrinne. Modeling of packet loss and delay and their effect on real-time multimedia service quality. In *Proc. NOSSDAV*, 2000.
- [6] D. Joumlatt, J. Chandrashekar, B. Kevton, N. Taft, and R. Teixeira. Predicting user dissatisfaction with internet application performance at end-hosts. In *Proc. of IEEE INFOCOM (mini-conference)*, 2013.
- [7] D. Joumlatt, R. Teixeira, J. Chandrashekar, and N. Taft. Perspectives on Tracing End-Hosts: A Survey Summary. In *SIGCOMM CCR*, April 2009.
- [8] D. Joumlatt, R. Teixeira, J. Chandrashekar, and N. Taft. HostView: annotating end-host performance measurements with user feedback. *ACM SIGMETRICS Performance Evaluation Review*, 38(3), 2010.
- [9] D. Joumlatt, R. Teixeira, J. Chandrashekar, and N. Taft. Performance of networked applications: The challenges in capturing the user's perception. In *ACM SIGCOMM Workshop on Measurements Up the Stack*, 2011.
- [10] S. Tao, J. Apostolopoulos, and R. Guérin. Real-time monitoring of video quality in IP networks. *IEEE/ACM Trans. Netw.*, 16(5):1052–1065, October 2008.