Computer Networks 000 (2015) 1–14

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Crowdsourced network measurements: Benefits and best practices

Matthias Hirth[a,*], Tobias Hoßfeld[b], Marco Mellia[c], Christian Schwartz[a], Frank Lehrieder[a]

[a] University of Würzburg, Chair of Communication Networks, Germany
[b] University of Duisburg-Essen, Chair of Modeling of Adaptive Systems, Germany
[c] Politecnico di Torino, Dipartimento di Elettronica e Telecomunicazioni, Italy

ARTICLE INFO

Article history:
Received 31 October 2014
Revised 24 June 2015
Accepted 3 July 2015
Available online xxx

Keywords:
Crowdsourcing
Network measurements
Methodology
Best practices

ABSTRACT

Network measurements are of high importance both for the operation of networks and for the design and evaluation of new management mechanisms. Therefore, several approaches exist for running network measurements, ranging from analyzing live traffic traces from campus or Internet Service Provider (ISP) networks to performing active measurements on distributed testbeds, e.g., PlanetLab, or involving volunteers. However, each method falls short, offering only a partial view of the network. For instance, the scope of passive traffic traces is limited to an ISP's network and customers' habits, whereas active measurements might be biased by the population or node location involved. To complement these techniques, we propose to use (commercial) crowdsourcing platforms for network measurements. They permit a controllable, diverse and realistic view of the Internet and provide better control than do measurements with voluntary participants. In this study, we compare crowdsourcing with traditional measurement techniques, describe possible pitfalls and limitations, and present best practices to overcome these issues. The contribution of this paper is a guideline for researchers to understand when and how to exploit crowdsourcing for network measurements.

1. Introduction

The Internet has become an integral part of everyday life. Since its original design, it has experienced dramatic changes in terms of the number and types of nodes and applications such that the complexity of the system severely poses limitations on understanding its behavior. In this context, network measurements are crucial to shedding light on eventual issues, supporting the understanding of arising problems, and improving system design with the ultimate goal of enhancing the end user's Quality of Experience (QoE). Measurements must cover several technical aspects, for example, signal strength and radio coverage on the *link layer* and topology, routing and dynamic traffic changes on the *network layer*. For optimizing the QoE as perceived by users, however, *application layer measurements* on the end user device and subjective studies at the *user level* are gaining increasing importance to identify current and future network challenges and their effects on end users. To this end, measurement probes are required both within the network to measure technical parameters, and on the edge of the network to measure the QoE of individual users for specific applications.

Currently, coarse measurements are conducted that consider (a) passive observations of traffic in cooperation with Internet Service Providers (ISPs) and network operators,

* Corresponding author. Tel.: +49 931 3186954.
E-mail addresses: matthias.hirth@informatik.uni-wuerzburg.de (M. Hirth), tobias.hossfeld@uni-due.de (T. Hoßfeld), mellia@polito.it (M. Mellia), christian.schwartz@informatik.uni-wuerzburg.de (C. Schwartz), lehrieder@informatik.uni-wuerzburg.de (F. Lehrieder).

http://dx.doi.org/10.1016/j.comnet.2015.07.003
1389-1286/

(b) actively running experiments in testbeds, either in isolation or connected to the public Internet, or (c) asking voluntary participants to run a measurement tool. We propose Crowdsourced Network Measurements (CNM) as additional means for researchers to complement the view and broaden the scope of previous techniques.

Passive measurement studies performed in ISP or campus networks offer a very detailed and possibly complete view, but only on a limited portion of the Internet. Thus, it is difficult to generalize results. In addition, it is well known how difficult it is to access real traffic traces, and the raw measurement data are rarely published due to privacy and business issues. Therefore, academic testbeds, e.g., PlanetLab, are available to the research community. Their worldwide sites offer a detailed but sparse and biased view of the current Internet. Testbed nodes are typically located in academic institutions or research facilities with high-speed Internet access and are shared among several experiments that could interfere with one another. Hence, those measurements are not useful for quality estimation on application and user levels. Furthermore, diversity in terms of devices and Internet access is much more limited than the device diversity typical Internet users have access to today. Instrumenting measurements on real devices of volunteers can overcome these limitations and provide a realistic and diverse view—depending on the number of involved end users.

With voluntary participants, the initial phase of user acquisition can be challenging, and maintaining a constant user pool to perform repetitive experiments is often difficult. Moreover, it is usually difficult to run dedicated experiments exclusively for users in a given geographical location, or with specific devices, if the number of volunteers in the project is not high. To overcome the issues of the aforementioned measurement techniques, we propose crowdsourced network measurements, i.e., recruiting users of paid crowdsourcing platforms to run the measurement software on their own devices. CNM can be viewed as a special use case of crowdsensing, in which user devices act as environmental sensors [1]. However, in the proposed approach, we limit our interest to monitoring technical network conditions. Similar difficulties to those experienced with crowdsensing arise here, e.g., the appropriate incentive design or the validation of the observations. In contrast to existing work, we do not focus on the realization of a special measurement case; instead, we focus on the general benefits and limitations of CNM. Furthermore, we note the challenges to realizing network measurements with crowdsourcing users and show possible solutions and best practices. The results can be used for evaluating the suitability of CNM for specific problems, and the best practices may help to avoid common pitfalls.

The reminder of this work is structured as follows. Section 2 reviews general network measurement techniques. The concept of CNM is introduced in Section 3, and its advantages and challenges are discussed. Section 4 discusses different parameters considered when designing network measurements and to what extent they can be realized with the different techniques. Furthermore, a comparison of CNM and general measurement techniques is given here. Section 5 illustrates the advantages of CNM using some exemplary use cases. Practical guidelines for conducting CNM and avoiding common pitfalls are given in Section 6. Section 7 summarizes use-case-driven research challenges to be addressed by crowdsourcing providers and researchers to improve further the applicability of CNM. Section 8 concludes the paper.

2. General network measurement techniques

Network measurements are primarily conducted using existing infrastructure at an ISP, in testbeds or with the help of voluntary participants. In the following, we further detail the basic principles of these approaches.

2.1. Network measurements by ISPs

ISPs have direct access to their network components, e.g., routers and Points-of-Presence, and thus they are able to gain detailed knowledge about their network. This includes complete information about the structure of the network and the traffic within the network. Measurements of application behavior are possible to a certain extent by using advanced tools that extract information from packet traces, e.g., using deep-packet inspection methodologies [2–4]. The amount of data that must be processed causes new challenges, but sampling strategies and today's processing power allow easy scaling to several Gb/s [5]. This type of measurement allows the drawing of a very accurate picture of a specific part of the Internet. The ability to perform passive analysis using off-the-shelf hardware has made such measurements quite popular among the research community, in which novel methodologies are being devised to extract increasing amounts of valuable information from passive traces.

2.2. Distributed testbeds

Testbeds, such as PlanetLab [6], M-LAB [7], GENI [8], and GLab [9], consist of hundreds of nodes located inside a country or worldwide. These testbeds allow us to run distributed experiments in a well-specified environment that supports even complex measurement setups. In contrast to ISP measurements, testbeds offer the possibility of a broader view of the Internet, due to the distributed geographical locations and the different Internet connections of the nodes. Testing novel applications on PlanetLab has become the de facto standard in the research community. Similarly, PlanetLab is popular for running active measurements to gather information about the status of the Internet. However, the limited and often special position of testbed nodes decreases the generality of results.

2.3. Voluntary participation of internet users

Another possible means of performing network measurements relies on voluntary participants. DIMES [10], iPlane [11], or DipZoom [12] are among the first attempts in this direction. Measurement tools have been made available to the community and volunteers asked to participate in these experiments. Unfortunately, the majority of the participating hosts are PlanetLab nodes, with some nodes from academia and a few handfuls of residential hosts.

To access a broader range of end user devices, projects attempt to ease the installation of software. One technique employed is distributing plug-ins for popular software to

Please cite this article as: M. Hirth et al., Crowdsourced network measurements: Benefits and best practices, Computer Networks (2015), http://dx.doi.org/10.1016/j.comnet.2015.07.003