Crowdsourced network measurements: Benefits and best practices

(Article begins on next page)

Crowdsourced Network Measurements: Benefits and Best Practices

Matthias Hirth[1], Tobias Hoßfeld[a,b], Marco Mellia[c], Christian Schwartz[a], Frank Lehrieder[a]

[a] *University of Würzburg, Chair of Communication Networks*
[b] *Now at: University of Duisburg-Essen, Chair of Modeling of Adaptive Systems*
[c] *Politecnico di Torino, Dipartimento di Elettronica*

Abstract

Network measurements are of high importance both for the operation of networks and for the design and evaluation of new mechanisms. Therefore, several approaches exist for making network measurements, ranging from analyzing live traffic traces from campus or Internet Service Providers (ISPs) networks, to performing active measurements on distributed testbeds, e.g, PlanetLab, or involving volunteers. Each method falls short in offering only a partial view of the network. For instance, the scope of traffic traces is limited to the ISP's network and customers' habits, while 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 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 comprises a guideline for researchers when and how to exploit crowdsourcing for network measurements.

Keywords: Crowdsourcing, Network Measurements, Methodology, Best Practices

Email addresses: `matthias.hirth@informatik.uni-wuerzburg.de` (Matthias Hirth), `tobias.hossfeld@uni-due.de` (Tobias Hoßfeld), `mellia@tlc.polito.it` (Marco Mellia), `christian.schwartz@informatik.uni-wuerzburg.de` (Christian Schwartz), `lehrieder@informatik.uni-wuerzburg.de` (Frank Lehrieder)

1. Introduction

The Internet has become an integral part of everyday life. Since its original design, it has experienced dramatic changes in terms of number and types of its nodes and the running applications, so that the complexity of the system severely poses limitation to understand its behavior. In this context, network measurements are crucial to shed 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 have to cover several technical aspects: signal strength and radio coverage on the *link layer*; topology, routing and dynamic traffic changes on the *network layer*, etc. For optimizing the QoE as perceived by the users, however, *application layer measurements* on the end-user device and subjective studies at the *user level* gain more and more importance to identify current and future network challenges and their impact on the end-users. To this end different 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, measurements are conducted coarsely considering (a) passive observations of traffic in cooperation with Internet Service Providers (ISPs) and network operators, (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 an additional methodology 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 world wide located sites offer a broad but very limited view of the current Internet. Most 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 each other. Hence, those measurements are not useful for quality estimation on the application and user levels. Furthermore, diversity in terms of devices and Internet access is much more limited than the one typical Internet users have access to today. Instrumenting measurements on real devices of volunteers can overcome those 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 especially challenging and maintaining a constant user pool to perform repetitive experiments is often hard. Moreover, it is usually difficult to run dedicated experiments exclusively for users from 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 seen as a special use case of crowdsensing, where user devices act as environmental sensors [1]. However, in the proposed approach we limit on monitoring to technical network conditions, but still similar difficulties to crowdsensing arise here, e.g., an 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 in this work on the general benefits and limitations of CNM. Further we point out the challenges in realizing network measurements with crowdsourcing users and show possible solutions, as well as best practices. The discussion results can be used for evaluating the suitability of this measurement approach for specific problem and the best practices help avoiding 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 as well as challenges are discussed. Section 4 discusses different parameters considered while designing network measurements and to which extend they can be realized with the different network measurement techniques. Further a comparison of CNM and general network measurement techniques is given here. Section 5 illustrates the advantages of CNM using some exemplary use-cases. Practical guidelines for conducting CNM and to avoid common pitfalls are given in Section 6. Section 7 summarizes use-case driven research challenges to be addressed by crowdsourcing providers and researcher to further improve the applicability of CNM. Section 8 concludes the paper.

2. General Network Measurement Techniques

Network measurements are mainly 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 to and thus are able to gain detailed knowledge about their own 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 has to be processed causes new challenges, but sampling strategies and today's processing power allow to scale to several Gb/s easily [5]. This type of measurement allows drawing a very accurate picture of a specific part of the Internet. The ability to perform passive analysis using off-the-shelf hardware has made them quite popular among the research community, where novel methodologies are being devised to extract more and more valuable information from passive traces.

2.2. Distributed Testbeds

Testbeds, like PlanetLab [6], M-LAB [7], GENI [8], and GLab [9], consist of hundreds of nodes on a country or worldwide scale. These testbeds allow running distributed experiments in a well-specified environment that supports even complex measurement setups. In contrast to ISP measurements, testbeds offer possibly a broader view of the Internet, due to the sparse geographical location and the different Internet connections of the nodes. Testing novel applications on PlanetLab has become the de-facto standard in the research community. Similarly, it is popular to use PlanetLab to run 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 possibility to perform network measurements relies on voluntary participants. DIMES [10], iPlane [11], DipZoom [12], or DASU [13] 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 turns out to be PlanetLab nodes, with some nodes from academia and a few handful residential hosts.

To access a broader range of end-user devices, projects attempted to ease the installation of software, e.g., distributing plug-ins for widely used software to conduct measurements, e.g., leveraging Firefox browser extension [14], or distributing a plugin for BitTorrent clients [13, 15]. Providing measurements devices to end-users as done by SamKnows [16] or Ripe Atlas [17] is another means to access large testbeds. And finally, applications like Skype or the streaming solution of Conviva [18] embed network measurement tools aiming at specific service monitoring.

In contrast to paid crowdsourcing, no monetary incentives are involved here. However, a thoughtful incentive design including, e.g., the type of incentive and when to grant it, is crucial for motivating a large enough number of participants in a voluntary participation measurement. Some of the projects provide incentives, e.g., access to the observed information, access to other participating measurement probes, or improvement of the participants network performance [19]. However, these incentives mainly target technical interested and experienced users, or other researchers. Incentive mechanisms have to be adapted depending on the required target group of participants and the actual measurement, causing incentive design to become a hard challenge [20].

3. Crowdsourcing-Based Network Measurements

To complement the existing network measurement techniques, we suggest the usage of paid crowdsourcing as an additional way to acquire results from end-users. In this section we briefly give an introduction in the concept of paid crowdsourcing and define our terminology used in the reminder of this work. Afterwards, we detail on the advantages and challenges related to this measurement technique, as well as the resulting strengths, weaknesses, opportunities and threads.

3.1. General Overview of Paid Crowdsourcing

The term *Crowdsourcing* was initially defined by Jeff Howe in 2006 as *"...the act of taking a job traditionally performed by a designated agent and outsourcing it to an undefined, generally large group of people in form* *of an open call"* [21]. In contrast to outsourcing, the granularity of work in terms of the size of the tasks is usually small in crowdsourcing, as well as the administrative overhead is reduced [22]. Crowdsourcing and related topics like human computation or collective intelligence [23] have drawn a lot of attention in recent years and fostered the development of numerous new services and application in the Internet, like Wikis, online labor market places, or reCAPTCHA [24].

For the CNM online labour markets, or so-called crowdsourcing platforms, are of interest, because they offer easy and fast access to a large number of *workers*. These platforms act as a mediator between the *employers* who submit work and the human *workers* completing the tasks. Crowdsourcing platforms support a wide range of *tasks*. The simplest type of tasks including image tagging or simple text transcription, are so called *micro-tasks* that can be accomplished within a few minutes to a few hours. Creative tasks in contrast require a certain skill set and sometimes specialized tools. Examples are logo design or text production. However, even complex research and development challenges can be solved via crowdsourcing tasks. Different platform provides usually implement specific workflows to handle the individual task types. Creative tasks offered by design platforms like 99designs [25] or research and development challenges on InnoCentive [26] are often challenge based, i.e. only the best or most suitable solution submitted by the participants is paid. Therefore, these platforms focus on recruiting a very small set of highly specialized workers for a task.

In contrast to this CNM tasks belong to the category of micro-tasks. Micro-tasks simple tasks with are often highly repetitive, e.g., generating consecutive measurement samples. On commercial crowdsourcing platforms focusing on micro-task, like Amazon Mechanical Turk (MTurk) [27] or Microworkers.com [28], the tasks are usually grouped in larger units, which we refer to as *campaigns*. The workers are usually completely abstracted from the employer via the platform, which handles the distribution of the tasks automatically. However, some platforms allow directly selecting a subset of the anonymous crowd based on certain criteria, e.g., their current country or demographic properties. Further filtering criteria, e.g., the worker's device, might also be available in the future, as already specialized platforms for mobile crowdsourcing like Gigwalk [29] exist. The process of distributing via a micro-tasking crowdsourcing platform is schematically depicted in Figure 1. The submission of the completed tasks, as well as the aggregation of results [23] including possible quality control mechanisms [1, 30, 31] are omitted in the figure.

The possibility to select dedicated participants and the mainly monetary incentives in paid crowdsourcing are the main differences to voluntary participation. These are also the main reasons for some of the advantages and drawbacks of this measurement technique, which are addressed in the following.

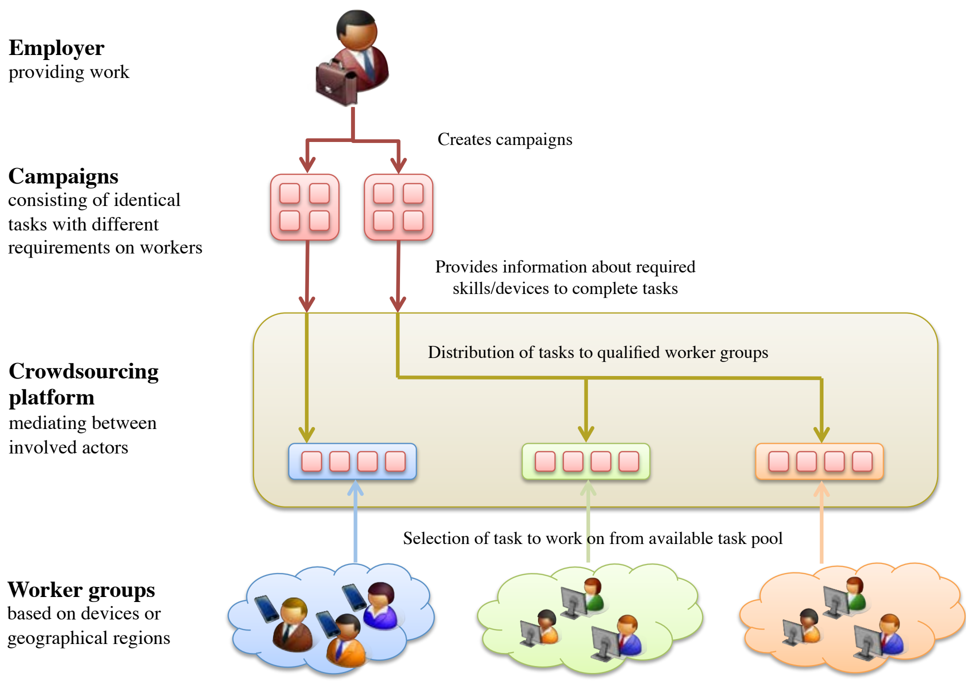

Figure 1: Overview of the work distribution process in a crowdsourcing platform.

3.2. Promising Advantages

In the following we discuss some promising advantages of CNM, which are later illustrated by selected use cases in Section 5. A detail comparison to other network measurement techniques can be found in Section 4.

CNM usually only causes *low costs for measurements*, even for very large-scale experiments. Almost no infrastructure is required to conduct a measurement, except for handling the reporting of the measurement probes. However, the fees for the crowdsourcing platform usage and the salary of the workers have to be taken into account.

The workers, and consequently the CNM probes, exhibit a very high *diversity*. Crowdsourcing workers are usually distributed all over the world, allowing researchers to conduct measurements from various geographical locations and multiple ISP networks. The workers also access the Internet using different types of broadband access technologies and a large variety of devices, ranging from desktop PC devices to smart phones, enabling a diverse view on the network.

This variety of real end-user devices in daily use allow measurements in *realistic scenarios*. Consequently, the results are not biased by special equipment or by special high-speed Internet connections often used in research facilities. Crowdsourcing-based probes can also be instrumented to collect information about commonly used software on end-user devices. Moreover, measurements on the end-user device can also easily involve the workers enabling large-scale and realistic *QoE measurements*.

End-user measurements are also possible using voluntary participation approaches, however, CNM offers a better *controllability of the probes*. The large variety and number of crowdsourcing workers allows researchers to chose only a subset of workers, which is suitable for meeting the requirements of the measurement, e.g., in terms of country of origin or hard- and software on the worker's device. Further, measurement tools can be implemented in such a way that they gather exactly the required level of detail of the measurement data without any additional censoring in a post-processing step.

Finally, the large number of workers on commercial crowdsourcing platforms offers a 24/7 workforce with thousands of workers being online at the same time. This does not only enable large-scale measurement campaigns, but also a *rapid generation of measurement results* with several hundreds of tasks being processed within in a few hours or even minutes.

3.3. Emerging Challenges

CNM enables several new possibilities for network measurements. However, it can be difficult to adapt current measurement approaches to incorporate crowdsourcing workers, because of several emerging challenges.

The *diversity of end-user devices*, one of the major advantages of this approach, can cause significant issues during the test setup. CNM probes might differ in their operation systems (OSes), software and hardware configurations, and their network connection. This has to be considered in the design of the measurement. It can be necessary to adapt, e.g., the duration of the measurement or the amount of transferred data based on the available bandwidth of the measurement probes. Further, the different operation systems and software environments have to be taken into account during the implementation of the measurement software.

The measurement software also has to provide means for detecting *untrustworthy workers*. Some crowdsourcing workers try to trick the measurement software, if they expect a gain from their action. This could be a faster completion of the task or multiple payments for the same task. Therefore, additional effort is required to add security checks to the software.

Another challenge that needs to be address in the test design is the *coordination of the workers*. On crowdsourcing platforms, the workers can decide which task to work on. This makes it difficult to schedule measurements at a very specific point of time. If further filtering of the workers is applied, e.g., for customers of a given ISP, the group of potential workers shrinks and it can become challenging to fulfill additional measurement constraints, e.g., a certain number of simultaneous probes.

Also recruiting workers from specific commercial platforms can be difficult due to *restrictions and limitations of the crowdsourcing platforms*. Crowdsourcing platforms can roughly be grouped in platforms with specialized use cases and platforms focusing on crowd provisioning. Specialized platforms, like CrowdFlower [32] offer sophisticated frameworks including quality control and crowd management for a given use case, e.g., content annotation. These platforms cannot be used for recruiting participants for network measurements at all, because they avoid a direct interaction between employers and workers. Crowd providers, e.g., MTurk or Microworkers, provide access to the registered workers and are suitable for a vast number of crowdsourcing tasks, but offer less quality assurance support. However, these platforms differ in their terms of use. MTurk for example restricts tasks, like asking workers to download and install software, or to register at other web pages, while platforms like Microworkers do not impose such restrictions. These restrictions have to be considered while designing the measurement tools, e.g., by selection of an appropriate platform or designing a web based tool.

Privacy and security constrains of the worker always have to be considered, independent of the regulations of the platform providers. Running a software tool from a unknown employer imposes a certain risk to a worker. Therefore, users may try to use sandbox environments to run the software or use fake identities to participate in tests that require registration. This, in turn, can result in biased measurement results.

3.4. SWOT Analysis of CNM

After discussing the advantages and challenges we continue with a SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis of CNM shown in Figure 2. The large number of measurement points and the relatively low cost for conducting the measurement are some of the strengths of the CNM. Still, the most important benefit of CNM is the direct access to end-user devices. This, however, is also the main reason for the weaknesses of the approach. When using end-user devices as measurement probes, the measurement software has to robust in the face of different hard- and software environments, the technical capabilities of the probes might be limited in some cases, and the experiments are harder to control.

Nevertheless, CNM open new opportunities for conducing measurement in realistic end-user environments, especially large-scale user studies. Concerning threads, CNM results might be biased by unknown influence factors, e.g., due to limitations of the end-user device or malicious workers. Further, the success of CNM measurements is difficult to predict, because a successful experiment depends no longer only on technical factors, but also on the willingness of the workers to participate.

In the next section we present a more detailed comparison of CNM and other measurement approaches. Later, in Section 6, we detail on methods to mitigate some of the weaknesses and threads of CNM.

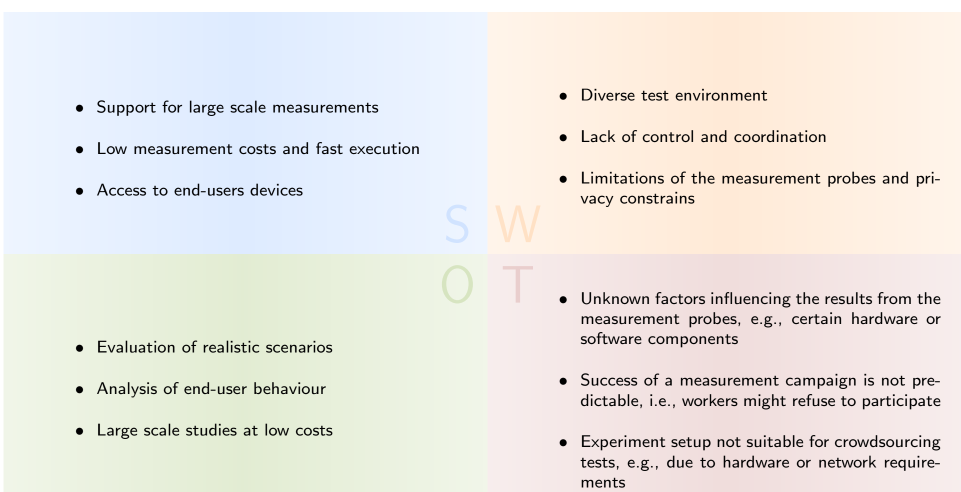

Figure 2: SWOT analysis of CNM

4. CNM vs. General Network Measurement Techniques

Depending on the addressed research question, different numbers, locations, and technical equipment for the measurement points are required. For some evaluations real user feedback has to be included as well. In the following we discuss in detail possible parameters for a network measurement setup and to which extent they could be fulfilled with different network measurement techniques. To illustrate the discussion, we use concrete examples, however, the addressed parameters are applicable to a wide range of measurements. Afterwards we directly compare the different measurement approaches and summarize their capabilities in Table 1.

4.1. Parameters of Network Measurements

The *granularity of network measurements data* is one of the parameters that needs to be considered while designing a measurement. The granularity of the data from network measurements can vary from packet traces to aggregated flows on a single client, e.g. to monitor application behaviour at the end user device [33], to cumulative traffic statistics of a backbone link for dimensioning wide area networks [34, 35]. While ISP traces allow a broadest spectrum of measurement granularity, the large amount of monitored data mainly requires the analysis of aggregated measures. In distributed testbeds, packet and flow level data is available, while it is not possible to measure backbone link utilization. The data granularity from voluntary measurements and CNM are limited by the constrains of the end-user device, because the OS might restrict capturing certain information. With regard to the examples mentioned at the beginning, distributed testbeds and CNM would be suitable for monitoring application behaviour at a client device, however ISP traces would be more appropriate for dimensioning decisions.

Besides the granularity of the measurement data, experimenters also need to adapt the *layers of the network stack* at which the data is collected according to the experiment. With custom ISP measurement solutions, it is possible to measure at any layer of the network stack, because the hardware can be fully controlled. However it is not possible to analyse application layer data encrypted at the end-hosts. In a testbed, some network layers might not be accessible due to restrictions of shared testbeds. In voluntary participation and CNM, the measurement tools run on standard OSes with their common security and technical constrains. This often limits the access to the network stack significantly. Considering the analysis of a cloud office service like Google Docs, ISP measurements would enable the experimenter to gather information on the lower layers of the ISO/OSI Model. This would, e.g., allow for deriving traffic patterns for this service. However, cloud services like Google Docs are usually secured by via ssl connections, therefore, it would not be possible to gather information about the user's interactions with the service's web interface. These interactions can only be captured directly on the end-device, e.g., via browser plugins. Consequently, voluntary participation or CNM could be used here.

Another important measurement parameter is the *scale of a measurement*. The scale can be defined, e.g., by the number of measurement points, their geographical distance, or their distance in terms of inter-AS hops, i.e. number of hops within an autonomous system (AS). ISP based measurements are limited by the number of nodes available to the ISP and the operations area of the ISP. Moreover, ISP nodes are naturally located in the same ASes or in ASes densely connected. Testbeds can scale from a few nodes to several hundreds, which are located at a single local site or are globally distributed. Global testbeds generally include nodes from multiple ASes, but testbed nodes are mainly located in research facilities and therefore likely to be connected to dedicated broadband access networks. Measurements based on voluntary participation can access a huge number of end-user nodes on a global scale, which are located in different ASes. However, the scale of the measurement, in terms of participating nodes, geographical distances and inter-AS distances is not controllable. CNM is comparable to voluntary participation, but provides means in adjusting the scale of the measurement by hiring a dedicated number of participants from selected geographical locations. An example for a small-scale measurement setup could be a local Wifi installations for interference tests. Here, the probes have to be located closely together, which usually requires a dedicated testbed. A possible application for a large scale measurement setup is, e.g., the analysis of content distribution networks [36]. In this case, world wide distributed measurement probes from different ASes are required, which can be achieve with voluntary participation, CNM, or a global testbed.

Besides a sufficient size of measurement setup also a certain *diversity of the measurement points* is needed to achieve results representative for a larger number of real network users. ISP and testbed based measurements are conducted using servers or dedicated measurement hardware, which are not common end-user devices. The same applies to the network access of these nodes. CNM and voluntary participation offer a diverse set of hardware devices, like end-user PCs, tablets, or smart-phones. For both approaches, the type of participating measurement nodes can be influenced by providing specialized measurement software, e.g., only for Linux or iOS. In the future crowdsourcing platforms might additionally allow hiring only users with given device specifications. Therefore, CNM and voluntary participation enable the evaluation of device specific influence factors, e.g., on the traffic patters of web applications, whereas ISP and testbed based measurements allow for conducting reference measurements with comparable hard- and software configurations.

All network experiments require *control of the test environment* to a certain extend. This includes the installed software tools, scheduling of measurements, and adaption of experiment parameters. Professional monitoring solutions are available in production environments of ISPs, but it is difficult to install experimental software tools or to influence the network significantly for test purposes. Testbeds in contrast offer a highly configurable and sometimes fully controllable environment, where arbitrary software can be installed and which can be manipulated according to the researchers needs. Voluntary participants or crowdsourcing workers can be asked to install experimental software tools, but a remote control of the tools is generally hard to achieve. In both cases, the network parameters can hardly be influenced.

The *time scale* of a measurement is another parameter in the design of network measurement. It can vary from a single snapshot to a long-term measurement, or with periodic measurements observing the change over time. Single-snapshot measurements are possible using any measurement technique. Long-term and repetitive measurements, however, are harder to conduct in voluntary participation and CNM, because the measurement probes need to remain active over a longer period of time. Repetitive measurement using the same measurement nodes multiple times are also difficult to achieve with voluntary participation and CNM, because the availability of the nodes it not guaranteed. In CNM this issue is diminished, as a group workers can be hired again to redo the measurement. Using hired workers also helps to enforce time constrains which are usually even more difficult to guarantee in voluntary participation approaches.

Besides technical parameters, the *human factor* becomes more and more relevant in network research. On the one hand, end-users generate traffic patterns through their interactions, which can affect the infrastructure to a high degree. On the other hand side the QoE becomes an important factor in measuring the satisfaction of customers. ISP traces already include a realistic traffic pattern from end-users, but it is not possible to trigger specific user interactions, e.g., flash-crowds. Testbeds usually do not produce real end-user traffic, but when using synthetic generators, predefined traffic patterns can be emulated. Using voluntary participation can help to collect real end-user traffic, but it is hard to trigger large scale behaviours involving multiple users. CNM also offers access to realistic traffic, however, even the triggering of flash-crowds is possible. Furthermore, voluntary participation and CNM additionally enable direct collection of actual user feedback.

Finally, the *costs* for conducting an experiment have to be considered. The costs for using ISP traces or a testbed vary based on the point of view. Both measurement techniques require significant investment costs for the hardware and software necessary for the measurement and test infrastructure. However, after this infrastructure is set up, the costs for conducting measurements is relatively small. Voluntary participation and CNM require less initial investment costs, because only a reporting system is required where the measurement results from the probes are send to. However, in CNM every measurement introduces additional costs for the workers' salary and the commission for the crowdsourcing platform.

4.2. Comparison of Network Measurement Techniques

After discussing several design parameters of network measurements individually, we now have a closer look at a direct comparison of the measurement techniques.

Measurements performed by *ISPs* and in *distributed testbeds* are mainly using dedicated and specialized hardware. This enables deploying specialized measurement tools, which gather information from the network layers. Still, restrictions are imposed either by test isolation considerations in testbeds or by security constrains in production environments. The direct control over the measurement probes enables long term and repetitive measurements. However, the specialized hardware and the dedicated testbeds impose biases in the measurements, which do not reflect end-user conditions. Both measurement techniques also fall short in providing direct end-user feedback or information about realistic end-user devices.

Voluntary participation and *CNM* measurement probes are intended to be real end-user devices. Consequently, the availability of individual measurement nodes varies significantly, as users may go offline or only participate in a single test. Moreover, the duration the participants contribute to a measurement cannot be predicted. However, both measurement techniques offer a rather realistic view of currently used end-user hard- and software, as well as end-user network connections. The main difference between voluntary participation and CNM is the motivation of the users. While voluntary participation is based on altruism or non-monetary incentives, crowdsourcing workers are profit oriented. This results in different challenges when designing measurement tools. Software for voluntary tests needs to consider the incentives, but not necessary requires security features to identify cheaters. Moreover, voluntary tests need public relations management to a certain extent in order to build up and maintain a user base. CNM using monetary incentives can be deployed rather quickly, as the required number of participants can be directly recruited. However, the software needs to implement features to avoid cheating and frauds. CNM can also be used to kickstart voluntary participation approaches, by recruiting the initial users.

Table 1 summarized the different parameters to consider while setting up network measurements and which realizations of those parameters are possible with the presented network measurement techniques. This overview can be used to select an appropriate measurement technique based no the measurement's requirements.

5. Use Cases for CNM

After discussing the general applicability of CNM for certain aspects of network measurements, we now illustrate CNM using examples of a few selected use cases.

5.1. Realistic End-user Probes

Network measurements are mainly performed using dedicated testbeds, which allow a biased view, because the hardware and the broadband connections are not representative for real end-user systems. For instance, a significant share of PlanetLab nodes is located within a "Global Research and Education Network" [37], and the available bandwidth of the nodes and real end-users show large differences. In July 2013, we conducted a measurement of the access bandwidth of 500 Microworkers.com users by asking them perform a commercial speed test [38] and hand in the link to the evaluation page. The results are depicted in Figure 3. In April 2014, we conducted a similar measurement using the command line tool of the same measurement provider [39] on 163 PlanetLab nodes, resulting in an average download bandwidth of 174.8 Mbps (Std.: 216.2 Mbps).

Requirement \ Technique	ISP Measurements	Distributed Testbeds	Voluntary Participation	CNM
Granularity of measurement data	Packet level to backbone aggregate	Packet and flow level	Limited by constrains of end-user device	
Measurement layer	Any except application layer	Partly network layer, no application layer	Application layer, additional information about real end user devices	
Measurement scale	Limited by owned nodes; limited ASes; geographically close; fixed scale	Global scale; multiple ASes; fixed size	Global scale; multiple ASes; unpredictable scale	Global scale; multiple ASes; scale and location can be controlled by hiring participants
Diversity of measurement points	Dedicated measurement/server hardware		Realistic end-user devices; often devices in research networks	Realistic end-user devices
Controllability of the measurement	Professional monitoring tools available; experimental software cannot be deployed in production environment	Highly configurable; experimental software can be deployed	Experimental software can be deployed; remote control of software hard to achieve	
Time scale	Snapshot and repetitive; short to long-term		Snapshot; short-term; repetitive with the same nodes hard to achieve	Snapshot; short-term; repetitive with the same nodes hard to achieve, but possibility to hire the same people again
End-user interactions	Realistic interactions recoded in the traces; specific interactions cannot be triggered	Mostly synthetic traffic	Interactions can be measured and to a certain extend be triggered	Interactions can be measured and triggered
Costs	Significant investment costs, afterwards free to use		Only expenses for reporting infrastructure	Expenses for reporting hardware; worker payments

Table 1: Parameters of network measurements and their feasibility in different network measurement techniques.

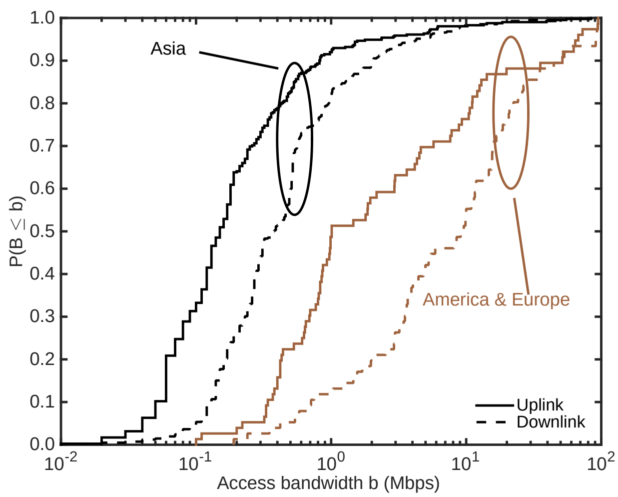

Figure 3: Measurement of access speeds of end users obtained via CNM.

Even if we only consider users from America and Europe where the access speed is significantly higher than in Asia, the average download bandwidth of 17.21 Mbps (Std.: 24.09 Mbps) is still low compared to PlanetLab nodes. The measurements allow two conclusions. First, network measurements performed on PlanetLab nodes might not be representative for real end-user devices. Therefore, additional reference measurements using at least a few end-users probes might be advisable for future measurement studies. Second, the measurement show that CNM might suffer from biases due to geographical location of the workers or their hard- and software configuration. However, this information can be monitored and used during the evaluation phase to normalize the results.

As a second example, we use crowdsourcing users recruited from Microworkers.com and PlanetLab nodes to measure the YouTube CDN in [40]. Most of the available PlanetLab nodes were located in the US and West-Europe, the crowdsourcing users were mainly based in Asia-Pacific and East-Europe. This reflects that the PlanetLab vantage points are overrepresented in areas with high education density, whereas the chosen crowdsourcing platform is very popular in developing countries. The results from the measurement show that the capability of PlanetLab to measure a global CDN is rather low, since 80% of the requests are directed to US servers. We further analyzed the number of ASes of the YouTube servers as observed by the Planetlab Nodes and the crowdsourcing users. Figure 4 shows the probabil-

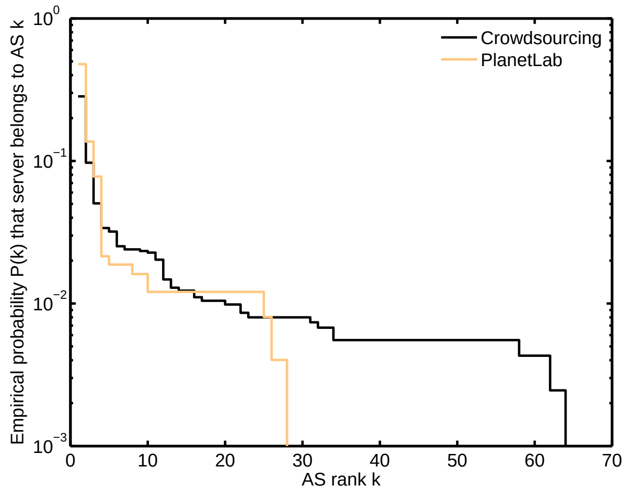

Figure 4: AS Distribution of YouTube servers via crowdsourcing and via PlanetLab.

ity that a server belongs to AS with rank k, where rank k is based on the number of YouTube servers within the AS. Here, the Planetlab nodes only observed less then 30 ASes, whereas the Crowdsourcing nodes were able to detect more than 60.

5.2. QoE and Application Layer Measurements

As mentioned before, one of the major drawbacks of testbed and ISP based network measurement techniques is the lack of user feedback. In contrast, CNM and voluntary approaches can be used to conduct large scale QoE measurements of real applications. Additional measurement tools can be deployed on the participants PC to monitor the network parameters and correlate them with the application behaviour and the QoE. Specialized test apps emulating a given application behaviour allow pinpointing the QoE-influencing factors even more easily. Besides the information retrieved during the measurement, additional details about the workers are commonly available via the Crowdsourcing platform, e.g., the worker's country of origin. This allows identifying additional influence factors or reducing the number of questions to the users for collecting relevant data. We found for example a major impact of the demographics of the users (to be more precise his country of origin) on the perception of aesthetic appeal of web sites [41] or images [42]. Beside cultural differences, the provided incentives and payments in those tests have to be considered [20], because they can influence

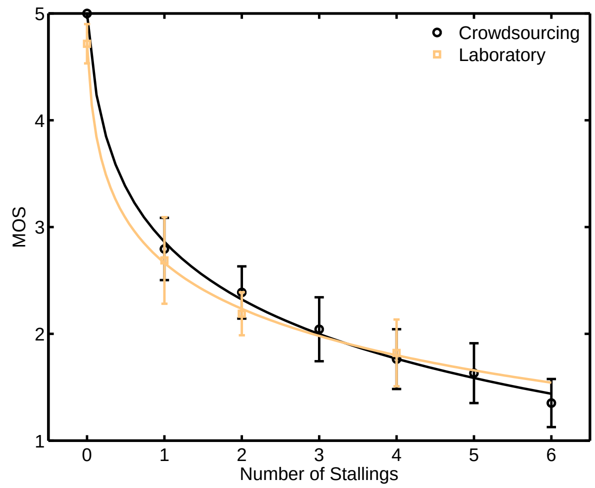

Figure 5: YouTube QoE results from subjective user studies conducted with crowdsourcing and in a laboratory environment.

the workers motivation to work diligently.

Furthermore, CNM can be used to replace cost intensive laboratory studies for subjective assessments, while offering a similar result quality. Figure 5 shows for example the QoE for YouTube video streaming in the presence of stalling events during the video playout. In this case we measured the QoE using the mean opinion score (MOS) [43], which can be determined from subjective ratings of test participants by averaging over multiple repetitions using the same stimulus. The subjective studies were conducted in a laboratory environment as well as by means of crowdsourcing. Both approaches lead to the same QoE results [44].

In contrast to voluntary participation, the costs for CNM are higher, but CNM enables a faster completion of the test. In [40], we describe a QoE experiments with both voluntary users from social networks and paid crowdsourcing users. While it took about 26 days to acquire about 100 voluntary testers, the same task was completed within 36 hours using a commercial crowdsourcing platform at total costs of $ 16.

CNM and voluntary participation also offer easy means to gather information on the application layer. While ISP traces only allow indirect information gathering of application information by analysing packet and flow content, CNM and voluntary participation allow direct access to certain application information direly on the end host. The same is also possible

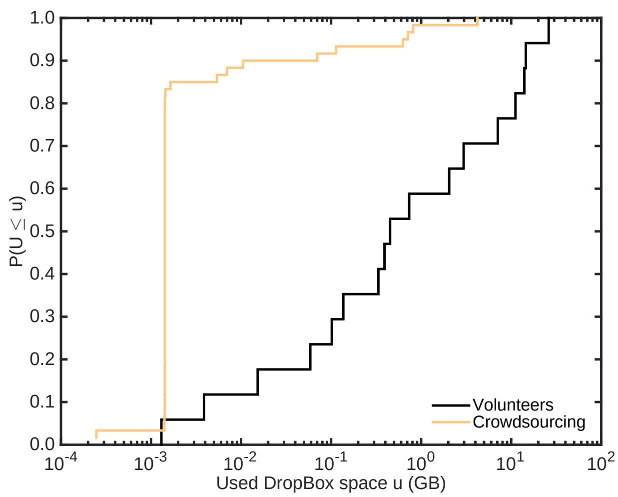

Figure 6: Usage of Dropbox space measured with crowdsourcing and with voluntary participation.

using test beds, however, here no real end-user interactions are available. On example for gathering application information using CNM and voluntary participation is given in [45]. In this study we analysed the usage as well as the most important and annoying features of the service DropBox as perceived by the participants. To gather information about the DropBox usage, we implemented a DropBox application using the official DropBox API to interact with the participants' account and a survey with questions about the users' experience with DropBox. In our case the DropBox application automatically gathered all relevant meta information, e.g., the available and used DropBox size, enabling us to collect objective information without any possible errors introduced, e.g., by erroneous filled questioners

Exemplary results from this study are shown in Figure 6, depicting the CDF of the used DropBox space of the participants. Other then in Figure 5, where we have a good accordance of crowdsourcing and laboratory results, we can see that both curves differ significantly in this experiment. This indicates that one or both of our test groups show biases, but from the measurement itself it is not clear which of them. However, the crowdsourcing results are in good accordance with other measurements [46], which leads to the assumption that the crowd-base measurement are more representative than the results obtained from the volunteers.

6. Best Practices

Previous crowdsourcing studies [20, 40–42, 45, 47, 48] experienced several pitfalls and practical problems. In the following we present a set of guidelines, which we developed to mitigate or avoid those problems.

Most of the workers neither have an education in computer science, nor are they experienced in network measurements. Therefore an *easy-to-use measurement software* is needed, because executing complicated shell commands is too complex and error-prone. In contrast, the required interaction between the worker and the software should be kept at a minimum with an user interface of the software being as simple as possible. This also means that the results are automatically collected on a server to avoid that workers having to deal with result delivery.

Moreover, the software as well as the test design has to *consider limitations of the target hosts*. Workers perform their tasks on a variety of devices and OSes. They might lack administrative privileges on their computers, for example, in Internet cafés. Under these conditions, installing software, as well as the execution of certain shell or scripting languages might not be possible. Therefore, we suggest using JavaScript or Java applets, as they permit running the measurements directly within the browser of the worker. An analysis of 558 workers from Microworkers showed that 97 % have JavaScipt enabled and 53 % have a working Java Runtime Environment. However, the security mechanisms of the browsers prevent the execution of shell commands. Good reasons exist for this behavior, but it limits the applicability to network measurements and workarounds have to be developed. The size of the measurement software should be kept to a minimum, as the available bandwidth of the end-user device is usually limited. Hence, a trade-off exists between the complexity and the prerequisites of the task on the one hand, and the number of successfully completed tasks on the other hand. This trade-off has to be considered carefully during the design of the measurement campaign.

Special attention should also be paid on *choosing the right crowd-provider*. There are lots of options for recuiting paricipants crowd-based measurements, ranging from online social network and online panels to a multitude of paid crowdsourcing platforms. All of providers differ — sometimes significantly — in the supported types of tasks, demographics of their users [49, 50], and their features for employers and workers [51]. In particular, the platform access, the diversity of participants, the costs per task and for qualification tests, payment features, the performance to acquire testers, as well as the integration of the measurement software into the platform have to be considered. A comprehensive overview of all available platforms is not possible due to the large number, therefore we provide a summary of features from two exemplary commercial crowdsouring platforms and one source of voluntary participants in Table 2. It has to be noted that the platform implementations, as well as the features are typically adapted over time. The information provided in the table reflects the status at the end of October, 2014.

Feature \ Platform	MTurk	Microworkers	Facebook
Platform access	Only US residence is allowed to create campaigns	Support of international employers	Support of international employers
Diversity of participants	Mainly US and Indian workers	International users with a large portion from Asia	Mainly friends or acquaintances
Costs per task	One cent to a few dollars (depending of the task length)	Ten cents to a few dollars (depending on the task length)	free
Variable payment features	Bank transfer, Amazon.com gift cards	Micropayment services, wire card, credit card	Not applicable
Costs for qualification tests	free	Ten cents to a few dollars (depending on the task length)	Free
Effort to acquire a large amount of testers	None	None	Test has to be designed in a joyful manner to attract participants and to go viral
Time to acquire a few hundred of testes	Few hours to a few days	Few hours to a few days	Few days to a few weeks
Support of specialized participant groups	Worker groups can be selected by qualifications e.g. obtained by qualification test or overall performance, or by given attributes e.g. country	Worker groups can be selected by overall performance, special attributes e.g. country, and deliberately formed by selecting individual workers	No direct support of grouping participants
Integration of measurement software into the platform	Forms are directly supported, more complex tasks have to be implemented on an own server and embedded in an iFrame	Only plain text descriptions and input is supported, more complex task have to be implemented on an own server	Tasks have to be implemented on an own server and can be embedded using an iFrame

Table 2: Comparison of two commercial crowdsourcing platforms Amazon Mechanical Turk and Microworkers, as well as the Facebook social network.

Depending on the specific requirements of the intended measurement, a careful selection of an appropriate crowd provider is required to overcome platform specific limitations. Demographical biases for example, can be overcome by filtering the participating workers or by selecting participants from multiple platforms, e.g. Facebook and a commercial platform, to assure the required diversity. Limitations in terms of supported tasks are harder to solve and different implementations might be required. While some crowd providers like Microworkers.com allow employers to pay for downloading and installing software, this is not possible on MTurk. Browser based solutions using JavaScript or Java applets can still be deployed here.

Independent of the crowdsourcing provider, most crowdsourcing tasks are performed via a web interface, e.g., showing images and providing an input field where tags can be added. However, CNM often impose more requirements on the workers' device or network connection, because specialized measurement software has to be executed on the device. Therefore, not all workers are able to complete the task, if the device does not fulfill the experiment's requirements. *Automated checks of the measurements prerequisites* at the beginning of the task can help here to minimize the time a worker spends on a task he cannot complete. For example, consider a measurement setup containing a Java applet. The experimenter should automatically check if Java is installed right at the beginning of the task. In case Java is not available or enabled at the workers' device, detailed information can be provided why the task is not available for the specific worker. Checking the measurement prerequisites in an automatic way gives also insights about possible issues of the task design, e.g., why most of the workers' do not complete the measurement. Further, detailed information about the end-user device can also be used create personalized measurement settings for each worker, e.g., workers with more powerful devices can perform more test repetitions then workers on mobile devices.

Besides a clear communication in case of errors, it is also important to *describe tasks in clear manner and simple words.* To achieve that a large number of workers can complete the task successfully, its description has to be easy to understand. Step-by-step instructions and screen-shots help workers to complete the task in a short amount of time. Technical and scientific terms should be avoided. Considering the large amount of non-native English speaking workers on international crowdsourcing platforms, multilingual task description can also increase the completion rate of tasks.

Even if the task description is detailed and well structured, some workers might face problems with the given task. Thus, it is necessary to *provide support for worker feedback and questions.* Feedback forms, forums, or email communication can be used for this purpose. Simple forms are recommended for optional feedback on the task, as they do not impose any additional effort to the worker, nor do they reveal any additional private information like email addresses. However, feedback forms only provide a one-way communication channel from the worker to the employer. This can be a significant disadvantage, e.g., if the workers faced issues during the task execution, which cannot be reproduced. Email communication or forums can help here, because they enable a more interactive communication. However, according to our experience, forums should be preferred. In most cases the majority of the workers faces the same issues or has the same questions. Therefore, a forum thread can help to answer multiple questions at once or provide possible solutions to a large number of people with a single post. Further, no private information, i.e., the email address, of the worker nor the employer is revealed. Using the worker feedback, the employer can support the workers, improve the task description, or modify the task design if required. In multi-step tasks, feedback should be possible at every stage not only at the end, since users that cannot complete the task cannot ask questions otherwise. The employer should also monitor existing communication channels of the workers, e.g., forums or Facebook pages. During one of our campaigns, a worker stated in Facebook that his virus scanner detected malware in our software. The problem arouses because the software tried to access the Internet for the measurements. A short post explaining the measurement details solved the problem and the other workers continued our task.

However, invalid measurement results are not only caused by misunderstandings or errors in the task design, but can also be caused by cheating workers, who try to receive the payment without performing the tasks properly. Therefore, *cheat detection and avoidance* techniques need to be applied. Results from cheating workers can highly affect the results of measurements [48] and impose additional costs [31]. A defensive-task design, i.e., it is easier to complete the task in a meaningful manner than to find a way to cheat, can be applied to measurements where no user interaction is required. If user interaction is required, e.g., the worker has to access certain web pages or videos, these interactions can be monitored [52–54] or additional validation questions [48] about contents of the visited page or video can be added to verify that the worker completed the task correctly.

7. Future Research Directions

Crowdsourcing is still a very active research area with numerous open research challenges. Here a distinction has to be made between challenges arising from specific use-cases and challenges related to crowdsourcing in general. General challenges as identified by, e.g., [55–57], include mechanisms for quality control, workflow design, motivation schemes for workers, legal and ethical issues, repeatability of experiments, interconnection of crowdsourcing and machine clouds, as well as technical challenges. Use-case related challenges include for example, the development of availability measures for ubiquitous crowdsourcing tasks [1]. CNM also imposes new challenges to crowd-providers and crowdsourcing users, which have not been addressed yet or are covered by general crowdsourcing research challenges. In the following we discuss some of the future research directions from a crowdsourcing point of view, which could foster the development of CNM.

From the experimenter point of view it is imported to *control the execution time of the task* to influence duration of the measurement. There is already some work [58–60] on real-time crowdsourcing, i.e., minimizing the time between the submission of a task and its completion. However, to the best of our knowledge, there does not exist any research on the how to implement more complex task scheduling or throttling mechanisms other then submitting the tasks individually.

CNM, especially the measurement software on the user-device, also requires new concept to preserve the *privacy of the crowdsourcing workers*. While conducing measurements, a maximum amount of information about the worker is desired. This includes both technical parameters of the hardware, as well as the geographical location and sometimes demographic information. Privacy concerns are considered in recent work in the field of mobile crowdsourcing [61] or in on commercial crowdsourcing platforms [62]. However, performing dedicated measurements on the application layer of the end-user device, e.g., using browser plugins, is potentially way more intrusive and thus new ways have to be found to grant detailed information to the experimenter while preserving the users privacy.

Moreover, CNM create the need for a further development of *pricing and incentive schemes*. Current pricing schemes, e.g., [63–65] assume that a task involves active participation of the worker, e.g., tagging and image or solving a problem. While this also applies for CNM involving explicit user feedback, a lot of CNM can be conducted by simply running a measurement tool in the background. Here, new pricing schemes are requited, because the measurements might run for a longer period of time then regular crowdsourcing tasks, but do not require any user participation. A similar research direction was identified in [66], considering the required resources for completing a crowdsourcing task in the pricing scheme. However, for a CNM the pricing scheme also has to consider the duration of the measurement and should be general enough to also consider possible user interactions.

Finally, CNM — like all crowdsourcing approaches — requires additional efforts to transform an existing system into a crowdsource-able version and this raises the question about the *feasibility of CNM*. To use an existing experiment with crowdsourcing workers, architectural and programmatic changes in the measurement software, as well as conceptual challenges like incentive design and designing cheat-detection and avoidance mechanisms are needed. These adaption impose personal costs and require additional time until the actual measurement can be conducted. Further, the costs for the crowdsourcing workers, as well as the administration of the task including the support for the worker have to be considered. Therefore, comprehensive cost models are desirable to analyse the trade-off between the benefits of crowdsourcing a network measurement and the resulting costs.

8. Conclusion

Even if CNM imposes several new challenges to experimenters, it can be a valuable tool to achieve a realistic view of the network from an end-user perspective. In comparison to data provided by ISPs, CNM data is not as detailed but uses less biased vantage points, and thus offers a broader view on the Internet. The crowdsourcing measurement nodes are more diverse in terms of available software and hardware, however less reliable than a dedicated testbed. This requires more effort during the development and deployment of a measurement tool, however the measurement nodes represent a more realistic environment. Furthermore, crowdsourcing-based measurement tools enable new possibilities to monitor user behavior, gather user feedback and conduct user level measurements like in video QoE. In contrast to approaches using voluntary participation, the measurements can be accomplished faster and with more control of the participating measurement points. CNM should not be seen as a replacement for common measurement techniques, but more as an additional measurement possibility for specific use cases. Especially the combination with existing measurement techniques seems promising.

Acknowledgements

The research leading to these results has received funding from the European Union under the FP7 Grant Agreement n. 318627 (Integrated Project "mPlane") and from the Deutsche Forschungsgemeinschaft (DFG) under Grants HO 4770/1-1, TR257/31-1.

[1] M. Vukovic, R. Das, S. Kumara, From Sensing to Controlling: The State of the Art in Ubiquitous Crowdsourcing, International Journal of Communication Networks and Distributed Systems 11 (1).

[2] V. K. Adhikari, S. Jain, Z.-L. Zhang, YouTube Traffic Dynamics and its Interplay with a Tier-1 ISP: An ISP Perspective, in: Proceedings of SIGCOMM Conference, New Delhi, India, 2010.

[3] B. Ager, N. Chatzis, A. Feldmann, N. Sarrar, S. Uhlig, W. Willinger, Anatomy of a Large European IXP, in: Proceedings of SIGCOMM Conference, Helsinki, Finland, 2012.

[4] A. Finamore, M. Mellia, M. Meo, M. Munafo, D. Rossi, Experiences of Internet Traffic Monitoring with tstat, IEEE Network 25 (3).

[5] P. M. Santiago del Rio, D. Rossi, F. Gringoli, L. Nava, L. Salgarelli, J. Aracil, Wire-speed Statistical Classification of Network Traffic on Commodity Hardware, in: Proceedings of Internet Measurement Conference.

[6] B. Chun, D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, M. Bowman, Planetlab: An Overlay Testbed for Broad-Coverage Cervices, Computer Communication Review 33 (3).

[7] C. Dovrolis, K. Gummadi, A. Kuzmanovic, S. D. Meinrath, Measurement Lab: Overview and an Invitation to the Research Community, Computer Communication Review 40 (3).

[8] GENI, GENI - Global Environment for Network Innovations, mboxhttp://www.geni.net/ (Accessed: February, 2015).

[9] GLab, GLab - German Lab, http://www.german-lab.de (Accessed: February, 2015).

[10] Y. Shavitt, E. Shir, DIMES: Let the Internet Measure Itself, Computer Communications Review 35 (5).

[11] H. V. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, A. Venkataramani, iPlane: An Information Plane for Distributed Services, in: Proceedings of Symposium on Operating Systems Design and Implementation, Seattle, Washington, USA, 2006.

[12] Z. Wen, S. Triukose, M. Rabinovich, Facilitating Focused Internet Measurements, in: Proceedings of SIGMETRICS Conference, San Diego, California, USA, 2007.

[13] M. A. Sánchez, J. S. Otto, Z. S. Bischof, D. R. Choffnes, F. E. Bustamante, B. Krishnamurthy, W. Willinger, DASU: Pushing Experiments to the Internet's Edge, in: Proceedings of Symposium on Networked Systems Design and Implementation, Lombard, Illinois, USA, 2013.

[14] M. Dhawan, J. Samuel, R. Teixeira, C. Kreibich, M. Aollman, N. Weaver, V. Paxson, Fathom: A Browser-based Network Measurement Platform, in: Proceedings of Internet Measurement Conference, Boston, Massachusettes, USA, 2012.

[15] D. R. Choffnes, F. E. Bustamante, Z. Ge, Crowdsourcing Service-Level Network Event Monitoring, in: Proceedings of SIGCOMM Conference, New Delhi, India, 2010.

[16] SamKnows Ltd, SamKnows -A free broadband measurement service for consumers, `http://www.samknows.com/` (Accessed: February, 2015).

[17] Ripe NCC, Ripe Atlas, `http://atlas.ripe.net/` (Accessed: February, 2015).

[18] Conviva Inc., Conviva - Intelligent Control Platform, `http://www.conviva.com` (Accessed: February, 2015).

[19] D. R. Choffnes, F. E. Bustamante, Taming the Torrent: A Practical Approach to Reducing Cross-isp Traffic in Peer-to-peer Systems, Computer Communications Review 38 (4).

[20] T. Hoßfeld, C. Keimel, M. Hirth, B. Gardlo, J. Habigt, K. Diepold, P. Tran-Gia, Best Practices for QoE Crowdtesting: QoE Assessment with Crowdsourcing, Transactions on Multimedia 16 (2).

[21] J. Howe, The rise of crowdsourcing, Wired magazine 14 (6).

[22] T. Hoßfeld, M. Hirth, P. Tran-Gia, Modeling of Crowdsourcing Platforms and Granularity of Work Organization in Future Internet, in: Proceedings of International Teletraffic Congress, San Francisco, Californica, USA, 2011.

[23] A. J. Quinn, B. B. Bederson, Human Computation: A Survey and Taxonomy of a Growing Field, in: Proceedings of Conference on Human Factors in Computing Systems, Vancouver, Canada.

[24] L. Von Ahn, B. Maurer, C. McMillen, D. Abraham, M. Blum, reCAPTCHA: Human-Based Character Recognition via Web Security Measures, Science 321 (5895).

[25] 99 designs Pty Ltd, 99 designs, `http://www.99designs.com` (Accessed: February, 2015).

[26] InnoCentive Inc., InnoCentive Challenges Programms, `http://www.innocentive.com` (Accessed: February, 2015).

[27] Amazon Inc., Amazon Mechanical Turk, `http://www.mturk.com` (Accessed: February, 2015).

[28] Weblabcenter Inc., Microworkers, `http://www.microworkers.com` (Accessed: February, 2015).

[29] Gigwalk Inc., Gigwalk, `http://www.gigwalk.com` (Accessed: February, 2015).

[30] M. Allahbakhsh, B. Benatallah, A. Ignjatovic, H. R. Motahari-Nezhad, E. Bertino, S. Dustdar, Quality Control in Crowdsourcing Systems: Issues and Directions, IEEE Internet Computing 17 (2).

[31] M. Hirth, T. Hoßfeld, P. Tran-Gia, Analyzing Costs and Accuracy of Validation Mechanisms for Crowdsourcing Platforms, Mathematical and Computer Modelling 57 (11).

[32] CrowdFlower Inc., CrowdFlower, `http://www.crowdflower.com` (Accessed: February, 2015).

[33] B. Staehle, M. Hirth, R. Pries, F. Wamser, D. Staehle, YoMo: A YouTube Application Comfort Monitoring Tool, in: Proceedings of Workshop on New Dimensions in the Assessment and Support of Quality of Experience for Multimedia Applications, Tampere, Finland, 2010.

[34] K. Thompson, G. J. Miller, R. Wilder, Wide-Area Internet Traffic Patterns and Aharacteristics, Network, IEEE 11 (6).

[35] R. Van De Meent, Network Link Dimensioning: A Measurement & Modeling Based Approach, Ph.D. thesis (2006).

[36] A.-J. Su, D. R. Choffnes, A. Kuzmanovic, F. E. Bustamante, Drafting Behind Akamai (Travelocity-Based Detouring), in: Proceedings of SIGCOMM Conference, Pisa, Italy, 2006.

[37] S. Banerjee, T. G. Griffin, M. Pias, The Interdomain Connectivity of PlanetLab Nodes, in: Procedings of Workshop on Passive and Active Network Measurement, Juan-les-Pins, France.

[38] Ookla, Speedtest.net, `http://www.speedtest.net` (Accessed: February, 2015).

[39] Matt Martz, Speedtest.net commandline interface, `http://github.com/sivel/speedtest-cli` (Accessed: February, 2015).

[40] V. Burger, M. Hirth, C. Schwartz, T. Hoßfeld, Increasing the Coverage of Vantage Points in Distributed Active Network Measurements by Crowdsourcing, in: Proceedings of Conference on Measurement, Modelling and Evaluation of Computing Systems, Bamberg, Germany, 2014.

[41] M. Varela, T. Mäki, L. Skorin-Kapov, T. Hoßfeld, Increasing Payments in Crowdsourcing: Don't look a Gift Horse in the Mouth, in: Proceedings of Workshop on Perceptual Quality of Systems, Vieanna, Austria, 2013.

[42] J. Redi, T. Hoßfeld, P. Korshunov, F. Mazza, I. Povoa, C. Keimel, Crowdsourcing-based Multimedia Subjective Evaluations: A Case Study on Image Recognizability and Aesthetic Appeal, in: Proceedings of CrowdMM Workshop, Barcelona, Spain, 2013.

[43] ITU-T, Recommendation P.800.1 – Mean Opinion Score (MOS) terminology, http://www.itu.int/ (accessed: February, 2015) (Jul. 2006).

[44] T. Hoßfels, R. Schatz, E. Biersack, L. Plissonneau, Internet Video Delivery in YouTube: From Traffic Measurements to Quality of Experience, in: E. Biersack, C. Callegari, M. Matijasevic (Eds.), Data Traffic Monitoring and Analysis: From Measurement, Classification and Anomaly Detection to Quality of Experience, Springers Computer Communications and Networks series, Volume 7754, 2013, pp. 264–301.

[45] P. Amrehn, K. Vandenbroucke, T. Hoßfeld, K. de Moor, M. Hirth, R. Schatz, P. Casas, Need for Speed? On Quality of Experience for File Storage Services, in: Proceedings of Workshop on Perceptual Quality of Systems, Vienna, Austria, 2013.

[46] I. Drago, M. Mellia, M. M. Munafo, A. Sperotto, R. Sadre, A. Pras, Inside Dropbox: Understanding Personal Cloud Storage Services, in: Proceedings of Internet Measurement Conference, Boston, Massachusetts, USA, 2012.

[47] B. Gardlo, M. Ries, T. Hoßfeld, Impact of Screening Technique on Crowdsourcing QoE Assessments, in: Proceedings of Radioelektronika Confernce, Brno, Czech Republic, 2012.

[48] T. Hoßfeld, R. Schatz, M. Seufert, M. Hirth, T. Zinner, P. Tran-Gia, Quantification of YouTube QoE via Crowdsourcing, in: Proceedings of Workshop on Multimedia Quality of Experience - Modeling, Evaluation, and Directions, Dana Point, California, USA, 2011.

[49] J. Ross, L. Irani, M. Silberman, A. Zaldivar, B. Tomlinson, Who are the Crowdworkers?: Shifting Demographics in Mechanical Turk, in: Proceedings of Conference on Human Factors in Computing Systems, Atlanta, Gorgia, USA, 2010.

[50] M. Hirth, T. Hoßfeld, P. Tran-Gia, Anatomy of a Crowdsourcing Platform - Using the Example of Microworkers.com, in: Proceedings of Workshop on Future Internet and Next Generation Network, Seoul, Korea, 2011.

[51] D. Vakharia, M. Lease, Beyond Mechanical Turk: An Analysis of Paid Crowd Work Platforms, in: Proceedings of the iConference, Newport Beach, California, USA, 2015.

[52] J. M. Rzeszotarski, A. Kittur, Instrumenting the Crowd: Using Implicit Behavioral Measures to Predict Task Performance, in: Proceedings of Symposium on User Interface Software and Technology, Santa Barbara, California, USA, 2011.

[53] R. K. Mok, W. Li, R. K. Chang, A User Behavior Based Cheat Detection Mechanism for Crowdtesting, in: Proceedings of SIGCOMM Conference, Chicago, Illinois, USA, 2014.

[54] M. Hirth, S. Scheuring, T. Hoßfeld, C. Schwartz, P. Tran-Gia, Predicting Result Quality in Crowdsourcing Using Application Layer Monitoring, in: Proceedings of International Conference on Communications and Electronics, Da Nang, Vietnam, 2014.

[55] M. Sabou, K. Bontcheva, A. Scharl, Crowdsourcing Research Opportunities: Lessons from Natural Language Processing, in: Proceedings of Conference on Knowledge Management and Knowledge Technologies, Graz, Austria, 2012.

[56] A. Kittur, J. V. Nickerson, M. Bernstein, E. Gerber, A. Shaw, J. Zimmerman, M. Lease, J. Horton, The Future of Crowd Work, in: Proceedings of Conference on Computer Supported Cooperative Work, San Antonio, Texas, USA, 2013.

[57] Y. Zhao, Q. Zhu, Evaluation on Crowdsourcing Research: Current Status and Future Direction, Information Systems Frontiers 16 (3).

[58] J. P. Bigham, C. Jayant, H. Ji, G. Little, A. Miller, R. C. Miller, R. Miller, A. Tatarowicz, B. White, S. White, et al., VizWiz: Nearly Real-Time Answers to Visual Questions, in: Proceedings of Symposium on User Interface Software and Technology, New York, New York, USA, 2010.

[59] M. S. Bernstein, J. Brandt, R. C. Miller, D. R. Karger, Crowds in Two Seconds: Enabling Realtime Crowd-powered Interfaces, in: Proceedings of Symposium on User Interface Software and Technology, Santa Barbara, California, USA, 2011.

[60] M. S. Bernstein, D. R. Karger, R. C. Miller, J. Brandt, Analytic Methods for Optimizing Realtime Crowdsourcing, arXiv preprint arXiv:1204.2995.

[61] D. Yang, G. Xue, X. Fang, J. Tang, Crowdsourcing to Smartphones: Incentive Mechanism Design for Mobile Phone Sensing, in: Proceedings of Conference on Mobile Computing and Networking, Istanbul, Turkey, 2012.

[62] M. Lease, J. Hullman, J. P. Bigham, M. S. Bernstein, J. Kim, W. Lasecki, S. Bakhshi, T. Mitra, R. C. Miller, Mechanical turk is not

anonymous, Available at SSRN: http://ssrn.com/abstract=2228728 or http://dx.doi.org/10.2139/ssrn.2228728.

[63] S. Faradani, B. Hartmann, P. G. Ipeirotis, What's the Right Price? Pricing Tasks for Finishing on Time, in: Proceedings of Workshop on Human Computation, San Francisco, California, USA, 2011.

[64] Y. Singer, M. Mittal, Pricing Mechanisms for Crowdsourcing Markets, in: Proceedings Conference on World Wide Web, Rio de Janeiro, Brazil, 2013.

[65] J. Wang, P. G. Ipeirotis, F. Provost, Quality-based pricing for crowdsourced workers, NYU Working Paper.

[66] M. Vukovic, Crowdsourcing for Enterprises, in: Proceedings on Conference on Services-I, Los Angeles, California, USA, 2009.