The definitive version has been published in Computer Networks, DOI: 10.1016/j.comnet.2015.07.003

# Crowdsourced Network Measurements: Benefits and Best Practices

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

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

**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.

*Keywords:* Crowdsourcing, Network Measurements, Methodology, Best

*Email addresses:* `matthias.hirth@informatik.uni-wuerzburg.de` (Matthias Hirth), `tobias.hossfeld@uni-due.de` (Tobias Hoßfeld), `mellia@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 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, (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 plugins for popular software to conduct measurements, e.g., leveraging a Firefox browser extension [13] or distributing a plugin for BitTorrent clients [14, 15]. Providing measurement devices to end users as done by SamKnows [16] or Ripe Atlas [17] is another means to create large testbeds. Finally, applications such as Skype or the streaming solution of Conviva [18] embed network measurement tools aimed 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 sufficient number of participants in a voluntary measurement context. Some of the projects provide incentives, e.g., access to the observed information, access to other participating measurement probes, or improvement of the participant's network performance [19]. However, these incentives primarily target interested and experienced technical users or other researchers. Incentive mechanisms must be adapted depending on the required target group of participants and the actual desired measurements, causing incentive design to become a difficult challenge [20].

## 3. Crowdsourcing-Based Network Measurements

To complement the existing network measurement techniques, we suggest the usage of paid crowdsourcing as an additional method to acquire results from end users. In this section, we briefly introduce the concept of paid crowdsourcing and define the terminology used in the reminder of this work. Thereafter, we detail the advantages and challenges related to this technique, and 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 the* *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, and administrative overhead is reduced [22]. Crowdsourcing and related topics such as human computation or collective intelligence [23] have drawn much attention in recent years and fostered the development of numerous new services and applications on the Internet such as like Wikis, online labor market places, and reCAPTCHA [24].

For CNM, online labor 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 includes image tagging or simple text transcription; these are called *micro-tasks* and can be accomplished within a few minutes to a few hours. Creative tasks in contrast require certain skills and occasionally specialized tools. Examples are logo design or text production. However, even complex research and development challenges can be solved via crowdsourcing tasks. Different platforms usually implement specific workflows to handle individual task types. Creative tasks offered by design platforms such as 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. These platforms focus on recruiting a very small set of highly specialized workers for a task.

In contrast to the above, CNM tasks belong to the category of micro-tasks. Micro-tasks are simple tasks that are often highly repetitive, e.g., generating consecutive measurement samples. On commercial crowdsourcing platforms focusing on micro-tasks, such as 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 because specialized platforms for mobile crowdsourcing such as Gigwalk [29] already exist.

The process of distributing via a micro-tasking crowdsourcing platform is schematically depicted in Figure 1. The submission of the completed tasks and the aggregation of results [23], including possible quality control mechanisms [1, 30, 31], are omitted in the figure. The main differences between micro-tasking and voluntary participation are the possibility to select dedicated participants and the primarily monetary incentives in paid crowdsourcing. These are also the main reasons for some of the advantages and drawbacks of this technique, which are discussed in the following.

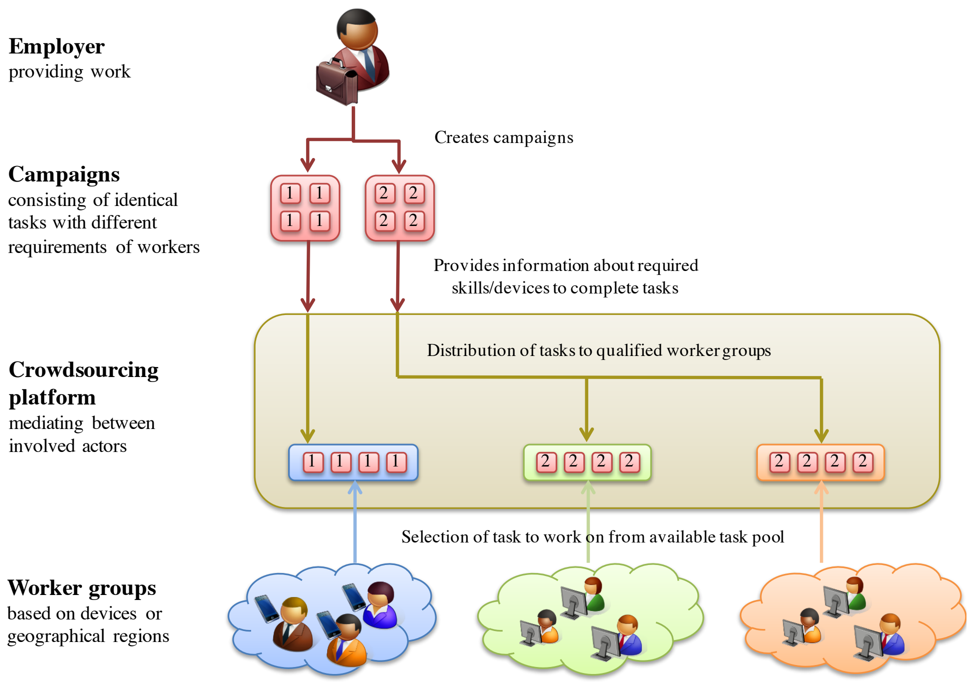


Figure 1: Overview of the work distribution process on a crowdsourcing platform

### *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 detailed comparison to other network measurement techniques is found in Section 4.

CNM usually calls for *low cost 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, fees for the crowdsourcing platform usage and the salary of the workers must be considered.

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 access the Internet using different types of broadband access technologies and a large variety of devices, ranging from desktop PCs to smartphones, enabling a diverse view of the network.

This variety of real end user devices allows 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 easily involve the workers enabling large-scale, realistic *QoE measurements*.

End user measurements are also possible using voluntary participation approaches. However, CNM offers better *controllability of the probes*. The large variety and number of crowdsourcing workers allows researchers to choose only a subset of workers suitable for meeting the specific requirements of the measurement, e.g., in terms of country of origin or hardware and software on the worker's device. Furthermore, measurement tools can be implemented to 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 enables not only large-scale measurement campaigns but also a *rapid generation of measurement results*, with several hundreds of tasks being processed 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 Operating Systems (OSs), software and hardware configurations and their network connection. This must be considered in the design of the measurement. It may 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. Furthermore, the different operating systems and software environments must be considered during the implementation of the measurement software. The latter also must provide means for detecting *untrustworthy workers*. Some crowdsourcing workers try to trick the system, if they anticipate doing so will result in a gain. This could comprise 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 must be addressed in the design is *coordination of the workers*. On crowdsourcing platforms, the workers decide which task to work on. This makes it difficult to schedule measurements at a very specific point in time. If further filtering of the workers is applied, e.g., selecting customers of a given ISP, the group of potential workers shrinks, and fulfilling additional constraints, e.g., a minimum number of simultaneous probes, can become a challenge.

Additionally, recruiting workers from specific commercial platforms can be difficult due to *restrictions and limitations of the crowdsourcing platforms*. Crowdsourcing platforms can roughly be divided into two groups: platforms with specialized use cases and platforms focusing on crowd provisioning. Specialized platforms, such as 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 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. For example, MTurk restricts tasks, such as asking workers to download and install software or to register at other web pages, whereas platforms such as Microworkers do not impose such restrictions. These restrictions must be considered when designing the measurement tools, e.g., by selection of an appropriate platform or by designing a web-based tool.

*Privacy and security constraints* of the worker always must be considered, independent of the regulations of the platform providers. Running a software tool from an unknown employer imposes a certain risk on 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, and Threats) analysis of CNM as 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 CNM. However, 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 must be robust in the face of different hardware and software environments, the technical capabilities of the probes might be limited in some cases, and the experiments are harder to control.

Nevertheless, CNM opens new opportunities for conducting measurements in realistic end user environments, particularly large-scale user studies. Concerning threats, CNM results might be biased by unknown influence factors, e.g., due to limitations of end user devices or malicious workers. Furthermore, the success of CNM 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 provide details on methods to mitigate some of the weaknesses and threats 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 tests, real user feedback also must be included. In the following, we discuss in detail possible parameters for a network measurement setup and to what extent the test requirements could be fulfilled with different network measurement techniques. To illustrate the discussion, we use concrete examples, although the addressed parameters are applicable to a wide range of measurements. Thereafter, we directly compare the different measurement approaches and summarize their capabilities in Table 1.

### *4.1. Parameters of Network Measurements*

The *granularity of network measurement data* is one of the parameters that must be considered when designing a measurement. The granularity of data can vary from packet traces to aggregated flows on a single client,


Figure 2: SWOT analysis of CNM

e.g., to monitor application behavior at the end user device [33], to traffic statistics of a backbone link for dimensioning wide-area networks [34, 35]. Although ISP traces allow the broadest spectrum of measurement granularity, the large amount of data largely requires the analysis of aggregated measures. In distributed testbeds, packet and flow-level data are available, but it is not possible to measure backbone link utilization. The data granularity from voluntary measurements and CNM are limited by the constraints of the end user device because the OS might restrict collecting certain information. Concerning the examples mentioned at the beginning, distributed testbeds and CNM would be suitable for monitoring application behavior on a client device; however, ISP traces would be more appropriate for dimensioning decisions.

In addition to the granularity of the measurement data, experimenters also must adapt the *layers of the network stack* at which the data are 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 analyze 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 OSs with their security and technical constraints. This often limits access to the network stack significantly. Considering the analysis of a cloud office service such as Google Docs, ISP measurements would enable the experimenter to gather information on the lower layers of the network stack. For example, this would allow for deriving traffic patterns for this service. However, cloud services such as Google Docs are usually secured by SSL/TLS connections; therefore, gathering information about the user's interactions with the service's web interface would not be possible. 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 ASs or in densely connected ASs. 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 ASs, but they are largely located in research facilities and are therefore likely to be connected to dedicated broadband access networks. Voluntary participation often accesses a huge number of end user nodes on a global scale, which are located in different ASs. 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 of adjusting the scale of the measurement by hiring a dedicated number of participants from selected geographical locations. An example for a small-scale setup could be a local WiFi installation for interference tests. Here, the probes must be located close 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, worldwide distributed measurement probes from different ASs are required, which can be achieved with voluntary participation, CNM, or a global testbed.

In addition to a sufficiently sized measurement setup, a certain *diversity of the measurement points* is also needed to achieve results representative of 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 type of network access of nodes. CNM and voluntary participation offer a diverse set of hardware devices, such as end user PCs, tablets, or smartphones. 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 patterns of web applications, whereas ISP and testbed-based measurements allow for conducting reference measurements with comparable hardware and software configurations.

All network experiments require *control of the test environment* to a certain extent. This includes 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 occasionally fully controllable environment, in which arbitrary software can be installed and that can be manipulated according to researchers' needs. Voluntary participants or crowdsourcing workers can be asked to install experimental software tools, but remote control of the tools is generally difficult 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, periodic measurements observing changes over time. Single-snapshot measurements are possible using any measurement technique. Long term and repetitive measurements, however, are more difficult to conduct in voluntary participation and CNM because the measurement probes must remain active over a longer period. Repetitive measurements using the same nodes multiple times are also difficult to achieve with voluntary participation and CNM because the availability of the nodes is not guaranteed. In CNM, this issue is diminished because a group of workers can be hired again to redo the test. Using hired workers also helps to enforce time constraints, which are usually more difficult to guarantee in voluntary participation approaches.

In addition to technical parameters, the *human factor* becomes increasingly 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, 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 difficult to trigger large-scale behaviors involving multiple users. CNM also offers access to realistic traffic, and even the triggering of flash crowds is possible. Furthermore, voluntary participation and CNM enable direct collection of actual user feedback.

Finally, the *costs* for conducting an experiment must be considered. The costs for using ISP traces or a testbed vary based on the point of view. Both measurement techniques require significant investments for the hardware and software necessary for the measurement and test infrastructure. However, after this infrastructure is set up, the costs for conducting measurements are relatively small. Voluntary participation and CNM require less initial investment cost because only a reporting system is required to which the measurement results from the probes are sent. 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 examine more closely a direct comparison of the measurement techniques.

Measurements performed by *ISPs* and in *distributed testbeds* are primarily taken using dedicated and specialized hardware. This enables deploying specialized measurement tools that gather information from network layers. However, restrictions are imposed either by test isolation considerations in testbeds or by security constraints in production environments. Direct control over measurement probes enables long term and repetitive measurements. However, the specialized hardware and the dedicated testbeds impose biases on 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 because 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 hardware and software and of end user network connections. The main difference between voluntary participation and CNM is the motivation of the users. Although 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 must consider the incentives but does not necessarily require security features to identify cheaters. Moreover, voluntary tests require a certain amount of public relations management to build up and maintain a user base. CNM, using monetary incentives, can be deployed rather quickly because the required number of participants can be directly recruited. However, the software must implement features to avoid cheating and fraud. CNM can also be used to kick-start voluntary participation approaches by recruiting the initial users.

Table 1 summarizes the different parameters to consider when 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 on the measurement's requirements.

## 5. Use Cases for CNM

Having discussed the general applicability of CNM to 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 largely performed using dedicated testbeds, which allow a biased view because the hardware and the broadband connections are not representative of 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 differenced. In July 2013, we conducted a measurement of the access bandwidth of 500 Microworkers.com users by asking them to 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

| Requirement \ Technique | ISP Measurements | Distributed Testbeds | Voluntary Participation | CNM |
|---|---|---|---|---|
| Granularity of measurement data | Packet level to backbone aggregate | Packet and flow level | Limited by constraints 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 ASs; geographically close; fixed scale | Global scale; multiple ASs; fixed size | Global scale; multiple ASs; unpredictable scale | Global scale; multiple ASs; 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 difficult to achieve | |
| Time scale | Snapshot and repetitive; short to long term | | Snapshot; short term; repetition with the same nodes difficult to achieve | Snapshot; short term; repetition with the same nodes difficult to achieve, but possibility to hire the same people again |
| End user interactions | Realistic interactions recorded in the traces; specific interactions cannot be triggered | Mostly synthetic traffic | Interactions can be measured and to a certain extent can be triggered | Interactions can be measured and triggered |
| Costs | Significant investment costs, thereafter free to use | | Only expenses for reporting infrastructure | Expenses for reporting hardware; worker payments |

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

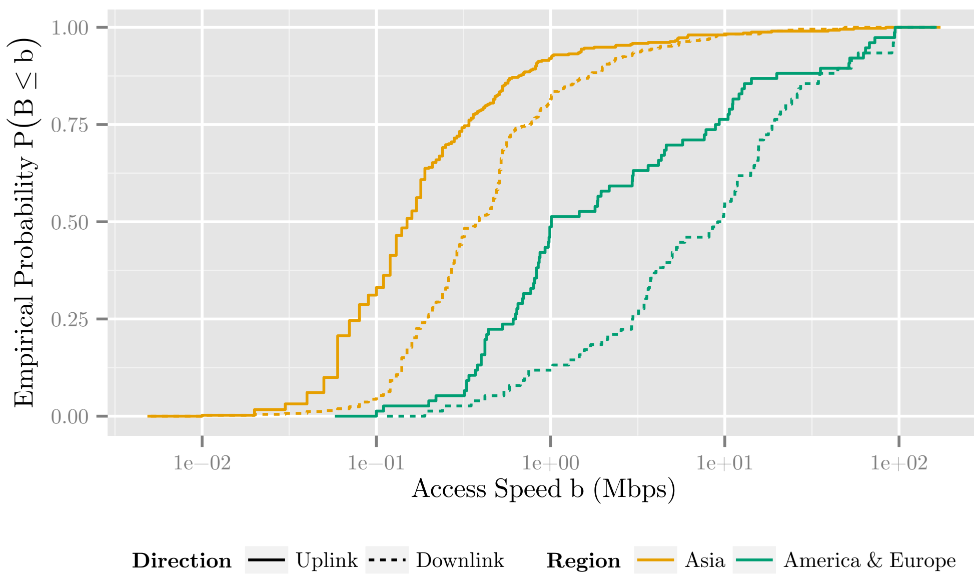


Figure 3: Measurement of access speeds of end users obtained via CNM

measurement provider [39] on 163 PlanetLab nodes, resulting in an average download bandwidth of 174.8 Mbps (Std.: 216.2 Mbps).

Although 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) remains low compared with PlanetLab nodes. The measurements allow us two conclusions. First, network measurements performed on PlanetLab nodes may not be representative of 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 shows that CNM might suffer from biases due to geographical location of the workers. However, this information can be monitored and used during the evaluation phase to normalize the results.

For 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 Western Europe, whereas the crowdsourcing users were largely based in Asia-Pacific and Eastern Europe. This reflects the fact 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

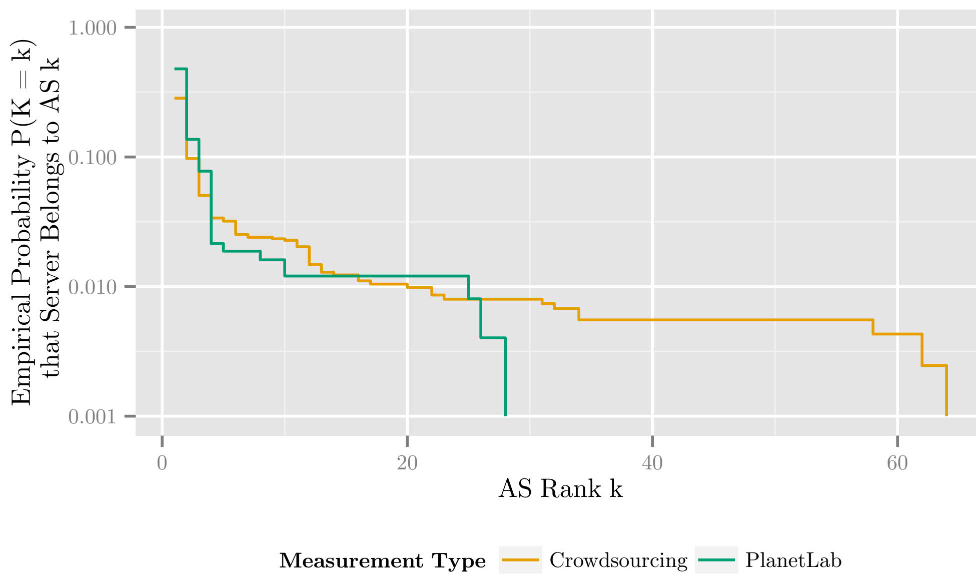


Figure 4: AS Distribution of YouTube servers via crowdsourcing and via PlanetLab

from the measurement show that the capability of PlanetLab to measure a global CDN is rather low because 80% of the requests are directed to US servers. We further analyzed the number of ASs of the YouTube servers as observed by the PlanetLab nodes and the crowdsourcing users. Figure 4 shows the probability 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 observed fewer than 30 ASs, whereas the Crowdsourcing nodes were able to detect more than 60.

### *5.2. QoE and Application-Layer Measurements*

As mentioned above, 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 participant's PC to monitor network parameters and correlate them with application behavior and the QoE. Specialized test applications emulating a given application behavior allow pinpointing the QoE-influencing factors even more easily. In addition to the information retrieved during the measurement, additional details about the workers are commonly available

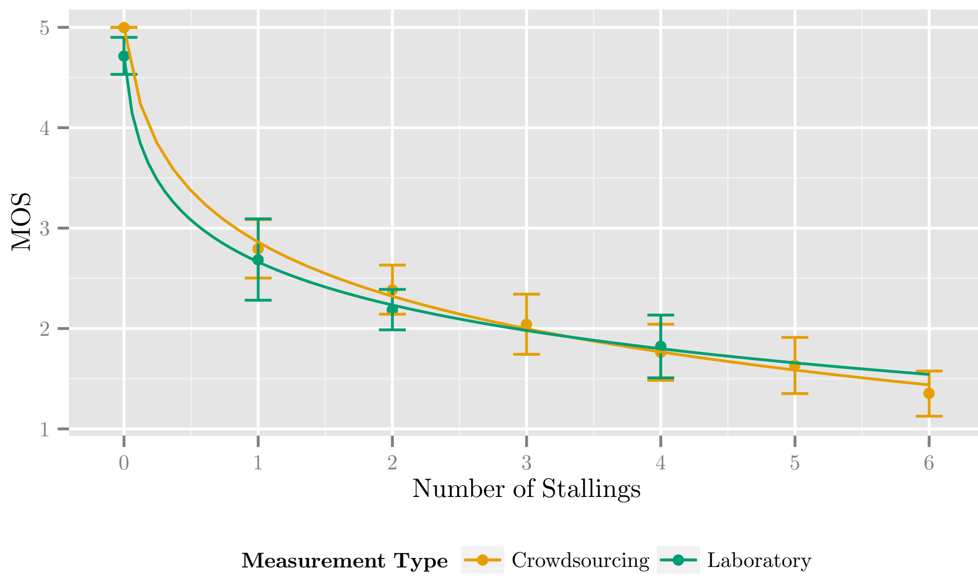


Figure 5: YouTube QoE results from subjective user studies conducted with crowdsourcing and in a laboratory environment

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. For example, we found a major effect of the demographics of the users (to be more precise, their country of origin) on the perception of aesthetic appeal of web sites [41] or images [42]. In addition to cultural differences, the provided incentives and payments in those tests must be considered [20] because they can influence the worker's motivation to work diligently.

Furthermore, CNM can be used to replace cost-intensive laboratory studies for subjective assessments while offering a similar result quality. For example, Figure 5 shows the QoE for YouTube video streaming in the presence of stalling events during the video playback. 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 and by means of crowdsourcing. Both approaches led 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 experiment with both voluntary users from social networks and paid crowdsourcing users. Whereas it took approximately 26 days to acquire approximately 100 voluntary testers, the same task was completed within 36 hours using a commercial crowdsourcing platform at a total cost of $16.

CNM and voluntary participation also offer easy means to gather information on the application layer. Whereas ISP traces only allow indirect information gathering of application information by analyzing packet and flow content, CNM and voluntary participation allow direct access to certain application information directly on the end host. The same is also possible using test beds; however, 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 analyzed the usage and the most important and annoying features of the service Dropbox as perceived by the participants. To gather information about Dropbox usage, we implemented a Dropbox application using the official Dropbox API to interact with the participants' accounts. A survey with questions about the users' experience with Dropbox was then offered. In our case, the Dropbox application automatically gathered all relevant meta information, e.g., the available and used Dropbox space, enabling us to collect objective information without any possible errors introduced, e.g., by erroneously completed questionnaires.

Exemplary results from this study are shown in Figure 6, depicting the CDF of the used Dropbox space of the participants. Other than in Figure 5, in which we have a good accordance of crowdsourcing and laboratory results, we can see the 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 is biased. However, the crowdsourcing results are in good accordance with other measurements [46], which suggests that the crowd-based measurements are more representative than are the results obtained from 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 lack any education in computer science or any experience in network measurements. Therefore, *easy-to-use measurement software*

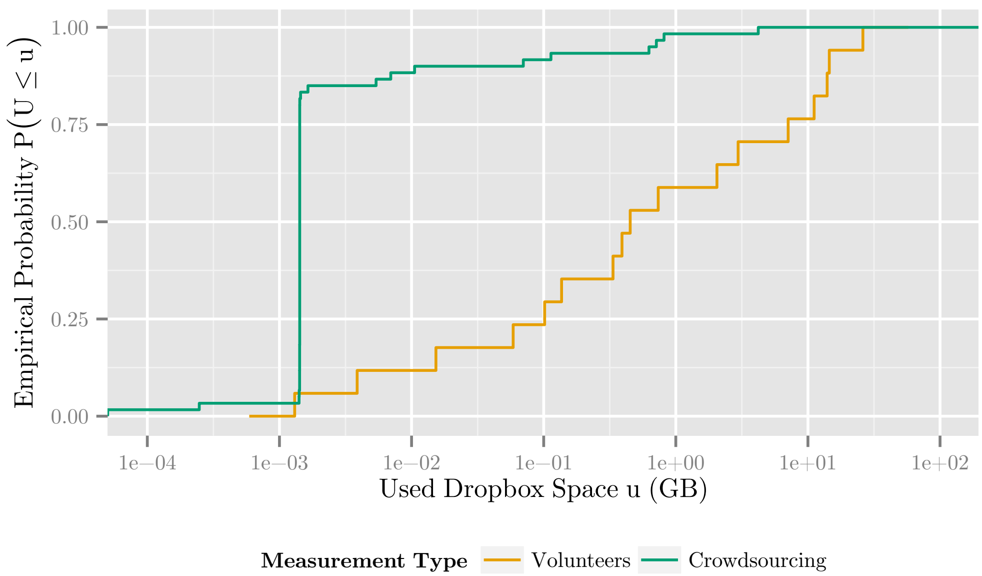


Figure 6: Usage of Dropbox space measured with crowdsourcing and with voluntary participation

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 the user interface of the software being as simple as possible. This implies that the results are automatically collected on a server to avoid any need for workers to address the resulting delivery.

Moreover, the software and the test design must *consider limitations of the target hosts*. Workers perform their tasks on a variety of devices and OSs. They might lack administrative privileges on their computers, for example, in Internet cafés. Under these conditions, installing software and executing certain shell or scripting languages might not be possible. Therefore, we suggest using JavaScript or Java applets because they permit running the measurements directly within the browser of the worker. An analysis of 558 workers from Microworkers showed that 97 % have JavaScript 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 application capability for network measurements; thus, workarounds must be developed. The size of the measurement software should be kept to a minimum because the bandwidth of the end user device is usually limited. Hence, a trade-off exists between complexity and prerequisites on the one hand, and the number of successfully completed tasks on the other hand. This trade-off must be considered carefully during the design of the measurement campaign.

There should be special focus on *choosing the right crowd-provider*. There are many options for recruiting participants including online social networks, online panels and a multitude of paid crowdsourcing platforms. All providers differ — occasionally 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, and the integration of the measurement software into the platform must be considered. A comprehensive overview of all available platforms is not possible due to the large number. We provide a summary of features from two exemplary commercial crowdsourcing platforms and one source of voluntary participants in Table 2. Note that platform implementations and their features typically evolve 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 residents are allowed to create campaigns | Support of international employers | Support of international employers |
| Diversity of participants | Primarily US and Indian workers | International users with a large portion from Asia | Primarily 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 number of testers | None | None | Test must to be designed in a joyful manner to attract participants and to go viral |
| Time to acquire a few hundred of testes | A few hours to a few days | A few hours to a few days | A 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 must be implemented on an own server and embedded in an iFrame | Only plain text descriptions and input are supported, more complex task must be implemented on an own server | Tasks must 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, and the Facebook social network

Depending on the specific requirements of the intended measurement, a careful selection of an appropriate crowd provider is needed. Demographical biases for example, can be avoided by filtering the participating workers or by selecting participants from multiple platforms, e.g., Facebook and a commercial platform, to ensure the required diversity. Limitations in terms of supported tasks are more difficult to address, and different implementations might be required. Although some crowd providers such as 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 requirements on workers' devices or network connections because specialized measurement software must be executed on the devices. Therefore, not all workers are able to complete the task if the device does not fulfill the experiment requirements. *Automated checks of the measurement 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 whether Java is installed at the very beginning of the task. If Java is not available or enabled on a worker's device, detailed information can be provided why the task is not available for the specific worker. Checking the measurement prerequisites automatically also yields insights about possible issues with the task design, e.g., why most workers do not complete the measurement. Furthermore, detailed information about the end user device can also be used to create personalized measurement settings for each worker, e.g., workers with more powerful devices can perform more repetitions than workers can with mobile devices.

In addition to clear communication in case of errors, it is also important to *describe tasks in a clear manner and simple words*. To let a large number of workers complete the task successfully, its description must 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 number of non-native English-speaking workers on international crowdsourcing platforms, a multilingual task description can also increase the completion rate of tasks.

Although 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 because they neither impose any additional effort on the worker nor reveal any additional private information such as email address. 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 that cannot be reproduced. Email communication or forums can help here because they enable more-interactive communication. However, according to our experience, forums should be preferred. In most cases, a majority of the workers face the same issues or have 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. Furthermore, no private information, i.e., the email address, of the worker is revealed. Using 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 to let users who cannot complete the task ask questions. The employer must monitor existing communication channels of the workers, e.g., forums or Facebook pages. During one of our campaigns, a worker stated on Facebook that his virus scanner detected malware in our software. The problem arose because the software tried to access the Internet for the measurements. A short post explaining the measurement details solved the problem and let the other workers continue our task.

Invalid measurement results not only are caused by misunderstandings or errors in the task design but also can be caused by cheating workers who try to receive the payment without performing the tasks properly. Therefore, *cheat detection and avoidance* techniques must be applied. The results from cheating workers can highly affect the results of measurements [48] and impose additional costs [31]. A defensive task design, i.e., making it easier to complete the task in a meaningful manner than to find a means to cheat, can be applied to measurements in which no user interaction is required. If user interaction is required, e.g., the worker must access certain web pages or videos, such interactions can be monitored [52–54] or additional validation questions [48] about the content of the visited pages or videos can be added to verify correct task completion by the worker.

## 7. Future Research Directions

Crowdsourcing remains a very active research area with numerous open research challenges. Here, a distinction must 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, and technical challenges. Use-case-related challenges include, for example, the development of availability measures for ubiquitous crowdsourcing tasks [1]. CNM also imposes new challenges on crowdproviders and crowdsourcing users, which have not yet been addressed or are covered by general crowdsourcing research challenges. In the following, we discuss some future research directions from a crowdsourcing point of view, which could foster the development of CNM.

From the experimenter point of view, it is important 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, no research exists on how to implement more-complex task scheduling or throttling mechanisms other than submitting the tasks individually.

CNM, the measurement software on the user device in particular, also requires new concepts to preserve the *privacy of the crowdsourcing workers*. When conducting measurements, a maximum amount of information about the worker is desired. This includes both technical parameters of the hardware, the geographical location and, occasionally, demographic information. Privacy concerns are considered in recent work in the field of mobile crowdsourcing [61] or 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 significantly more intrusive; thus, new methods must be found to grant detailed information to the experimenter while preserving the user's privacy.

Moreover, CNM creates the need for 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 an image or solving a problem. Although, this also applies for CNM involving explicit user feedback, a great deal of CNM can be conducted simply by running a measurement tool in the background. Here, new pricing schemes are required because the measurements might run for a longer period than regular crowdsourcing tasks, without any user participation. A similar research direction is identified by Vukovic et al. [66] who consider the required resources for completing a crowdsourcing task in the pricing scheme. However, for a CNM, the pricing scheme also must consider the duration of the measurement and should be sufficiently general to consider possible user interactions.

Finally, CNM — like all crowdsourcing approaches — requires additional efforts to transform an existing system into a version that can be used together with crowdsourcing. This raises the question of the *feasibility of CNM*. To use an existing experiment with crowdsourcing workers, architectural and programmatic changes in the measurement software and conceptual challenges such as incentive design and designing cheat-detection and avoidance mechanisms are needed. These adaptations impose personal costs and require additional time until the actual measurement can be conducted. Furthermore, both the costs for the crowdsourcing workers and the administration of the task including support for the workers must be considered. Therefore, comprehensive cost models are desirable to analyze the trade-off between the benefits of crowdsourcing a network measurement and the resulting costs.

## 8. Conclusion

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

**Acknowledgments**

The research leading to these results 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 4770/2-1, TR257/38-1.

[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 (2013).

[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 the 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 the 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 (2011).

[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 the of Internet Measurement Conference, Boston, Massachusetts, USA, 2012.

[6] B. Chun, D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, M. Bowman, Planetlab: An Overlay Testbed for Broad-Coverage Services, Computer Communication Review 33 (2003).

[7] C. Dovrolis, K. Gummadi, A. Kuzmanovic, S. D. Meinrath, Measurement Lab: Overview and an Invitation to the Research Community, Computer Communication Review 40 (2010).

[8] GENI, GENI - Global Environment for Network Innovations, mbox`http://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 (2005).

[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 the 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 the SIGMETRICS Conference, San Diego, California, USA, 2007.

[13] 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.

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

[15] 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 the Symposium on Networked Systems Design and Implementation, Lombard, Illinois, USA, 2013.

[16] SamKnows Ltd, SamKnows -A Free Broadband Masurement 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 (2008).

[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 (2014).

[21] J. Howe, The Rise of Crowdsourcing, Wired Magazine 14 (2006).

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

[23] A. J. Quinn, B. B. Bederson, Human Computation: A Survey and Taxonomy of a Growing Field, in: Proceedings of the Conference on Human Factors in Computing Systems, Vancouver, Canada, 2011.

[24] L. Von Ahn, B. Maurer, C. McMillen, D. Abraham, M. Blum, reCAPTCHA: Human-Based Character Recognition via Web Security Measures, Science 321 (2008).

[25] 99designs Pty Ltd, 99designs, `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 (2013).

[31] M. Hirth, T. Hoßfeld, P. Tran-Gia, Analyzing Costs and Accuracy of Validation Mechanisms for Crowdsourcing Platforms, Mathematical and Computer Modelling 57 (2013).

[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 the 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 Characteristics, IEEE Network, IEEE 11 (1997).

[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 the SIGCOMM Conference, Pisa, Italy, 2006.

[37] S. Banerjee, T. G. Griffin, M. Pias, The Interdomain Connectivity of PlanetLab Nodes, in: Proceedings of the Workshop on Passive and Active Network Measurement, Juan-les-Pins, France, 2004.

[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 the 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 the Workshop on Perceptual Quality of Systems, Vienna, 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 the CrowdMM Workshop, Barcelona, Spain, 2013.

[43] ITU-T, Recommendation P.800.1 – Mean Opinion Score (MOS) Terminology, http://www.itu.int/ (Accessed: February, 2015), 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, volume 77, Springers Computer Communications and Networks Serie, 2013.

[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 the 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 the 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 the 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 the 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 the Conference on Human Factors in Computing Systems, Atlanta, Georgia, USA, 2010.

[50] M. Hirth, T. Hoßfeld, P. Tran-Gia, Anatomy of a Crowdsourcing Platform - Using the Example of Microworkers.com, in: Proceedings of the 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 the 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 the 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 the 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 the 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 the 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 (2014).

[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 the 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 the 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, in: Proceedings of the Collective Intelligence Conference, Cambrige, Massachusetts, USA, 2012.

[61] D. Yang, G. Xue, X. Fang, J. Tang, Crowdsourcing to Smartphones: Incentive Mechanism Design for Mobile Phone Sensing, in: Proceedings of the 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 (2013).

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

[64] Y. Singer, M. Mittal, Pricing Mechanisms for Crowdsourcing Markets, in: Proceedings of the 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 (2013).

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