Standardizing Large-Scale Measurement Platforms

Marcelo Bagnulo
University
Carlos III of Madrid
marcelo@it.uc3m.es

Philip Eardley
Trevor Burbridge
BT
philip.eardley@bt.com
trevor.burbridge@bt.com

Brian Trammell
ETH Zurich
trammell@
tik.ee.ethz.ch

Rolf Winter
University of
Applied Sciences
Augsburg
rolf.winter@
hs-augsburg.de

This article is an editorial note submitted to CCR. It has NOT been peer reviewed. The authors take full responsibility for this article's technical content. Comments can be posted through CCR Online.

ABSTRACT

Over the last few years, we have witnessed the deployment of large measurement platforms that enable measurements from many vantage points. Examples of these platforms include SamKnows and RIPE ATLAS. All told, there are tens of thousands of measurement agents. Most of these measurement agents are located in the end-user premises; these can run measurements against other user agents located in strategic locations, according to the measurements to be performed. Thanks to the large number of measurement agents, these platforms can provide data about key network performance indicators from the end-user perspective. This data is useful to network operators to improve their operations, as well to regulators and to end users themselves. Currently deployed platforms use proprietary protocols to exchange information between the different parts. As these platforms grow to become an important tool to understand network performance, it is important to standardize the protocols between the different elements of the platform. In this paper, we present ongoing standardization efforts in this area as well as the main challenges that these efforts are facing.

Categories and Subject Descriptors

C.2.3 [Computer-Communication Networks]: Network Operations – *network monitoring.*

Keywords

Measurement platforms, design, standardization, IETF.

1. INTRODUCTION

The set of things that are known about the Internet is small compared to those things that are unknown. This seems strange, since the Internet is an engineered, distributed system, designed, planned, constructed and continuously extended and optimized. As such, the engineers behind the Internet and its operation should really know everything about it. However, the Internet's distributed nature, its heterogeneity, information hiding practices in network management, and the sheer scale of the network makes it a daunting task at best to assess the state of the network.

"Measuring" the Internet would be beneficial for all involved stakeholders ranging from Internet Service Providers (ISPs) to consumers. For example, consumers today don't really know whether the data rate they in their contract with their ISP matches the actual rate they typically get, or whether the rate varies during the day. Network operators could use a better view of the network beyond their own network boundary for improved traffic engineering or for network diagnostics. Regulators could use better data to protect the interest of customers, and to guide better public policy decisions. Network equipment manufacturers could build better equipment based on data that shows how the Internet is evolving in various segments of the network and under various conditions.

To capture the state of the network and to probe it continuously, a large and diverse set of vantage points is necessary, i.e. a large number of measurement agents that are deployed in various locations in the network. The larger the number of measurement agents, the more complete is the picture these measurements paint. Therefore, for many measurement tasks, a large-scale measurement platform comprising tens of thousands of individual measurement agents is needed.

To fulfill the pressing need for data, several efforts to build large measurement platforms have started over the last few years. However, deploying such platforms is a challenging task because of the number of measurement agents involved. The result is a considerable number of measurement platforms with a diverse degree of success. These platforms typically cannot interact with each other and their results are difficult if not impossible to compare even though they are aiming to measure the same thing in many cases.

There are many potential places where standards for such large-scale measurement platforms could be useful. On the most extreme end, one could think of protocols that would allow arbitrary agents to interact with each other to perform measurements with or for them. This clearly has far reaching implications when it comes to security, privacy and protection of the intentions of the party that deployed the measurement agents. On the other end of the spectrum, standards could merely describe data formats and protocols to exchange measurement data between platforms. Many platforms deployed have very specific goals and collect only data related to these goals. It would be beneficial if this data could be shared so that other platforms could use it instead of reproducing the same data. The metrics underlying the exchanged data need to be well-specified in order that they can be interpreted correctly. Anything in-between those two extremes is certainly possible.

This paper briefly describes some of the existing measurement platforms, after which different areas where standardization would be beneficial are identified. Also, an analysis of the potential difficulties in the standardization work is presented as well as the benefits that it would bring to the different stakeholders.

2. LARGE-SCALE MEASUREMENT PLATFORMS

There are a number of existing platforms, with a widely different number of vantage points that collect sometimes overlapping data and that hardly interact with each other. Moreover, even when they are aiming to measure the same characteristic of the Internet they potentially use different underlying metrics or measurement methodologies. Therefore, the results they produce are difficult to compare or merge into a combined data set. In this section, we provide a very short survey of some of the more relevant large-scale measurement platforms. Due to lack of space and the large number of platforms, the survey is not exhaustive but merely illustrative of the existing diversity. These infrastructures come in a variety of forms: they may be software-based or hardware-based, they might use active or passive measurements, and be commercial or academic.

SamKnows [1] launched their fixed-line broadband performance measurement project in 2008 in conjunction with Ofcom, the UK telecommunications regulator. They have since then deployed over 20,000 hardware probes in participants' homes around the world, on behalf of regulators and ISPs. The probes run active tests throughout the day when the users are not using their broadband connection. The use of hardware probes provides high confidence in the accuracy of the results, as consistent hardware and software in the probes means that results are unaffected by varying client PC specifications. The use of hardware probes also means that they can be positioned to be able to detect user traffic and suppress testing in order to avoid corrupting the test results and impacting on user experience.

The platform supports a wide range of tests, both at the network and the application layer. New measurements can be remotely deployed on the probes. A whitepaper detailing the tests covered and the methodology employed is [2].

Bismark [3] is a hardware-based measurement platform. The effort is being led by the Georgia Institute of Technology and aims to provide a platform for researchers to study specific network phenomena. One such example is 'buffer bloat' – the detrimental effect that excessively large buffers can have on network performance. While its deployment model uses hardware-based probes similar to SamKnows, the goal of Bismark is not to benchmark broadband services against one another, but rather to study very specific network issues in a high level of detail. The project was launched in 2011 and at present has about 300 deployed probes, predominantly in North America.

RIPE Atlas [4] again uses hardware probes but unlike SamKnows and Bismark, these are not necessarily targeted to be installed behind an end user's broadband connection – they may be installed anywhere in an ISP's network. The project is operated by RIPE and funded through grants and the ISPs' purchase of probes. At the time of writing there were approximately 2000 probes deployed worldwide with the majority in Europe.

Ookla [5] operates the website speedtest.net, a popular web-based speed test that allows consumers to run ad-hoc checks of how their broadband connection is performing. The test results are also collected on the server side, and this data is made available to researchers for further study. Their usage of a purely software-based approach means that results may be distorted by client PC misconfiguration, poor internal home network performance or cross-traffic. The large sample size they collect reduces the impact of such effects to some extent, although sample self-selection and repeated tests from the same premise can also bias results. Ookla publishes some data via its website for use by researchers, and sells other data to ISPs. Independent studies have also used speedtest.net [6].

Netalyzr [7] comes from the University of Berkeley and focuses on diagnosing network capabilities from a web browser, such as detecting service blocking and NATs, rather than performance or continual reliability testing.

3. STANDARDIZATION COMPONENTS

This section identifies different elements of large-scale measurement platforms that could be standardized. These are divided into the following broad areas, which are depicted in Figure 1:

- Architecture
- Tests
- Control and Report Protocols
- Preprocessing of data

We expand on each of these areas next.

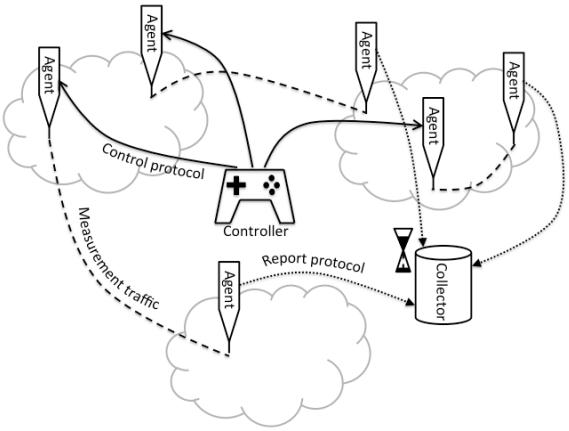

Figure 1. High-level interaction of components

3.1 Architecture

Even though the different available measurement platforms differ greatly, it would be useful to define a reference architecture for a measurement platform that comprises the main components that most if not all the platforms encompass. At the very least this should provide a common framework and terminology to discuss about measurement infrastructures. We can identify the following basic elements of a measurement platform:

- **Measurement agents** perform network measurements. They are pieces of code that can be executed in specialized hardware (hardware probe) or on a general purpose device (like a PC or mobile phone). In some cases they can be classified by their location. For instance in those platforms that measure the performance of the access network, there are a large number of measurement agents located in end user premises and a much smaller number located in the ISP's network, so that the former agents perform tests against the latter ones. Measurements may be active (the agents generate test traffic), passive (agents observe user traffic), or some hybrid form of the two.

- **Controller**(s) manage measurement agents by informing them which tests they should perform and when. This is a fundamental component since it is in charge of scheduling measurement activities performed by the agents. A particular measurement agent is controlled by one Controller, which avoids the possibility of the measurement agent receiving conflicting schedules or overloading the device or network. However there may be several Controllers, each controlling a subset of the measurement agents.
- **Collector**(s) accept measurement results from the agents, once their tests are complete. Collectors generally perform some analysis of their own (e.g. correlation and aggregation) and store or provide the results for reporting purposes. A given measurement agent may report to multiple collectors for data replication/partitioning or for redundancy.

While these are the main components of a measurement platform, it is possible to envision more sophisticated platforms that encompass additional elements, such as an aggregator of results, a hierarchy of controllers, and external data sources providing additional input to tests or analysis, such as 'looking glasses' for current topology information. However, these additional components need not be enumerated by a reference architecture. We believe a simple reference architecture containing the essential elements will prove to be a powerful tool.

3.2 Tests

Tests – the measurements performed by the agents – are one key area for standardization. There are different aspects of these tests that are appropriate for standardization. The most fundamental one is the metric i.e. a detailed specification of how to measure specific phenomena in the network. These specifications must be sufficiently detailed so that two independent implementations of the test produce coherent and comparable results.

Other aspects of tests appropriate for standardization include measurement agent placement. For example, if the goal is to define "Internet access speed", it is important to define the location of the measurement agents in order to guarantee that there is a coherent definition of what "Internet access" means. Similarly, as many of the tests will not be singleton tests but a series of tests, it is also important to standardize the scheduling strategy for test execution (i.e. periodic tests, Poisson distributions, etc). Finally, it may also be relevant to define standard environmental conditions for test execution. A notable example of such condition is the absence of cross traffic while performing a test that could potentially distort the test results. As absence of traffic can be defined in many different ways, a standard definition of the no cross traffic condition would help to produce coherent tests.

3.3 Control and Report Protocols

Another obvious area for standardization is related to control and report protocols.

The control protocol is used between the controller and the measurement agents. This protocol allows the controller to inform the measurement agents about which tests they should perform, the related parameters for the test (e.g. destination address, ports and other test specific parameters), and the scheduling for the test. Additionally, they could convey other mandatory or desirable conditions (e.g. perform this test when there is no cross traffic). The control protocol will also specify when and how to report results.

The report protocol is used between the measurement agents and the collector, so that the measurement agents can convey the results of the tests they have performed to the collector for processing and analysis. It is entirely possible that existing standards can be leveraged for the data transport part, however data export format definitions might be needed to allow interoperable data exchange.

3.4 Preprocessing

Finally, another area for standardization is related to preprocessing of results. This includes data preparation for analysis (e.g. how to eliminate outliers) as this may severely impact the obtained aggregated results. Other areas related to preprocessing include curation (i.e. how to consistently store data so it can be safely used in the future by other parties) and how to handle privacy issues (i.e. standard ways to anonymize data)

4. USE CASES FOR STANDARDIZATION

Different use cases have somewhat different motivations for deploying large measurement platforms and they results in different standardization requirements.

Regulators and consumer advocacy groups want large scale measurements in order to compare the performance of ISPs. By providing better information to customers they create more effective competition and drive up the performance (and/or drive down the price) of broadband access.

They require measurements so that different broadband providers can be accurately and fairly compared. They also need metrics that reflect the end users' 'quality of experience' (QoE) (and not some network metric of little real impact). Similarly governments would like to understand "how does the country's broadband compare with the rest of the world?"

The implications for standards are mainly the following:

- **Tests** need standardization, especially the metric definitions. "Speed" should be measured the same way for different ISPs, and "reliability" should be the same compound metric. Comparability is not enough: these definitions should also match how customers perceive "speed" and "reliability". Scheduling is less important to standardize since the results are averaged over many measurements although there will be requirements for the number of conducted tests in order to avoid disrupting the user traffic.
- **Preprocessing** is also important, since the removal of outliers and how results are aggregated affects the final result. However this cannot be fully standardized, due to differences in conditions among different networks, and the presence of anomalies in the raw measurements. Negotiation between the regulator and operator may be necessary to agree whether it is fair (or unfair) to include these anomalous measurements in the average.
- **Control and report protocols** are not particularly important to standardize. The regulator generally obtains measurement results from a single entity that runs the measurement agents, controller(s) and collector(s), and their contract defines what tests are done and how the results are reported. Control and reporting standardization may become important in a future time, to allow multiple such operators ("measurement service providers") to collaborate, but such a market is at present far from developing.

End users want large-scale measurements in order to check that their broadband service is performing as expected and to diagnose faults or impairments.

They require measurements that: provide metrics that reflect their individual QoE; accurately assess whether a service level agreement has been met (particularly for corporate customers); and isolate an issue to the home (or enterprise) network, the ISP's network or the service provider, so they know who to complain to.

The implications for standards are:

- **Tests** should be standardized They want to understand whether the QoE is as good as it should be. They want scheduled tests (to regularly check the SLA) and one-off ad-hoc tests. Again, tests should be comparable and should reflect perceived service quality, but there is more room for uncertainty in values than in the regulatory use case.
- In this case, there is no need to standardize the **control and report protocols,** or **preprocessing.**

Operators want large-scale measurements for several reasons. Similarly to the regulator use case, they want to understand how their service compares with competitors. They also want to understand the perspective of their end users better. In these cases, the implications for standardization are similar to the use cases above.

Large-scale measurements can also support network management and planning, service monitoring, and troubleshooting.

For network planning purposes, measurement can identify bottlenecks and observe the impact of emerging applications on traffic patterns. In this case, tests are run proactively on a sample of the measurement agents at the edge (i.e., operator-owned CPE or enterprise edge routers).

In the case of troubleshooting, the goal is to quickly identify and fix problems reported by customers, or to identify and fix these even before they result in customer support calls. Troubleshooting is inherently iterative; i.e., more detailed tests may be triggered by initial results..

Fault identification requires more reactive tests. For example, when a customer phones up to report a fault, the operator could immediately run a series of tests on that line to try and isolate where the fault is (e.g. in the operator's network or at home), what is causing it, (e.g. component failure or network overload) and how to fix it (e.g. reroute around a lossy link, or reboot the CPE).

These use cases require measurements that can locate an issue within a network, and measure network parameters along multiple paths and at multiple layers.

The operator may also want to test the impact of a new capability, such as IPv6, before it is offered to customers.

The implications for standards are:

- **Control and Report protocols**. The multiple observation points required by this use case can be provided by edge devices as well as end systems. Embedding measurement agents at these points requires standard control and report protocols, so that devices from multiple vendors can communicate with the same Controller and Collector.
- **Tests** benefit from standardization, as well, to ensure comparability of results produced by end systems and edge device agents from different vendors. Test standardization also allows tests to inter-operate between multiple vendors. Troubleshooting use cases, especially, may require new tests to be standardized.
- A standard for **preprocessing** is, as in end user cases, unimportant.

In theory, it is possible for one common measurement platform to serve all the above use cases. Different parties can operate their own measurement agents (and controller and collector) or have agreements to be able to test against other parties' agents (e.g. in-network agents). Multiple parties can also agree on a single test schedule for a measurement agent (which is not unreasonable as the measurement agent is owned by a single final authority). This also allows testing and assurance that the device on which the measurement agent is implemented, along with the network it is testing, is capable of operating the specified tests (e.g. devices have CPU/memory/storage limits).

While results could be shared directly with multiple parties' collectors, in reality it appears sufficient to report to a single party who can manage security and data protection issues. This also reduces the reporting load on the measurement agent and network. Data sharing can then take place at off-line rather than via the report protocol.

5. STANDARDIZATION OPPORTUNITIES

When it comes to the standardization opportunities described previously, the obvious first question to ask is which standards development organization (SDO) should specify which of the standardization components outlined in section 3. It should be noted that some SDOs have already ongoing work that covers certain aspects that are useful for large measurement infrastructures. Other, more recent, efforts have been triggered by people that have developed and deployed these platforms.

The Internet Engineering Task Force (IETF) IPPM Working Group has been working since 1997 on metrics, methodologies and protocols between test equipment, focusing on active measurement. The set of metrics defined by IPPM is not exhaustive, and there are certainly opportunities for extension of the IPPM metrics to be more useful for large-scale measurement. As an example, there could be definitions of metrics that take into account user experience, or more complicated preconditions for testing, as well as the definition of passive or hybrid metrics to be compatible with existing active metrics. However, given its long history and current activity, IPPM seems an ideal place to standardize new metrics. Among the standardization opportunities identified within IPPM, we can find metrics for critical measurements such as bulk transfer and buffer bloat. The establishment of a registry for metrics that would allow the control and report protocol to refer succinctly to tests will be useful.

Initiated by participants in the USA Federal Communications Commission's Measuring Broadband America (MBA) initiative, the IETF has recently started to explore whether additional standardization could be done in the context of broadband performance measurement. A mailing list (named LMAP) has been created to discuss this effort, and there is an initial document [8] that describes both the MBA architecture and where protocol support is needed. However at the time of this writing, there is no official working group within the IETF. The protocols under discussion in LMAP would allow the components described in section 3 to interact in a standardized manner.

The Broadband Forum (BBF) develops specifications for broadband wire-line solutions. Its project "Broadband Access Service Attributes and Performance Metrics" (WT-304) [9] started in autumn 2012. WT-304 builds on the earlier work of TR-143 [10] which only defined throughput and response time tests, by adding more performance tests such as loss and jitter and tests with emulated streaming, browsing and so on. WT-304 also aims for a more flexible capability that can for example measure particular segments of the network, measure across multiple networks, schedule continuous tests and allow on-demand triggering of tests.

The IEEE project P802.16.3 "Mobile Broadband Network Performance Measurements" also started in autumn 2012 [11]. It addresses end-to-end measurements to characterize the performance of mobile broadband networks from a user perspective. It is not limited to any particular air interface. One interesting topic is the need to collect metadata associated with a measurement, such as the device's location, the cell ID and maybe radio resource control parameters – and perhaps even a measurement could be triggered based on the value of some metadata. The remaining battery power and network cost/allowance are examples of environmental conditions for test execution.

Both the BBF and IEEE heavily reference IPPM metrics.

6. STANDARDIZATION CHALLENGES

In the previous sections we concluded that there is considerable standardization potential in the area of large-scale measurement platforms and that there are many benefits that could result from realizing such potential. Nevertheless, we should acknowledge that there are several challenges that standardization efforts will have to surmount. In this section we identify some of them that are apparent in this early stage of the process.

Scope

As we have observed in section 2, a significant number of (more or less) large measurement platforms exist today. Different platforms have been designed and deployed with different use cases in mind, as we have identified in section 4. Because of that, the design of the different platforms correspond to different requirements and result in different designs. The standardization process would need to specify only one solution, which would need to encompass as many of the requirements as possible at the same time. Given the different nature of the intended use cases for the measurement platforms it is far from obvious if it would be possible or even desirable to fulfill a large and possibly divergent set of requirements. As part of the standardization process, the first stage will involve scoping the outcome, and deciding which use cases will be covered and which ones will be discarded. This is likely to be a painful process and may jeopardize the entire standardization effort. Moreover, once the scope has been defined, it is likely that part of the community involved in the process will lose interest as their use case was defined to be out of scope. This may result in a significant decrease in the energy behind the standardization process, which again may jeopardize the whole process.

Protocol wars

The control and report protocols described in section 3 would need to be fairly generic protocols that can be operated by multiple device platforms. For the control protocol, any protocol that allows the (secure) transport of tests plans with corresponding parameters from the controller to the measurement agents should be sufficient. Similarly, for the report protocol, any protocol that allows the (secure) transport of the results of the tests should suffice. Note that different tests will require different input parameters and will produce different results, so transport protocols should be able to carry arbitrary information. There are a number of options for such protocols, ranging from defining new protocol(s) or reusing existing protocol(s) such as a RESTful API leveraging HTTP, NETCONF or IPFIX. Similarly, there are several ways to represent test control information and test results, including JSON, XML or new TLV formats.

Each such choice implies complex tradeoffs, dependencies, and implementation constraints; for example, as IPFIX provides transport and framing as well as an extensible, network-centric information model, it addresses more of the problem at hand in reporting but is less flexible to support interaction models other than that for which it was designed, while choosing HTTP for reporting would also require the definition of representation and semantics, but would be more flexible in deployment.

The issue is not merely technical. Since many people have their favorite protocol in mind, the possibility of a protocol war over which protocol(s) will be used is a real risk. Moreover, different protocols are likely to meet some requirements of some use cases more easily than others, resulting in an additional risk of the protocol war contaminating the already complicated requirement/use case discussion.

Definition of metrics

While there is a significant amount of work on the definition of metrics produced by the IPPM working group at the IETF, the actual tests performed by existing platforms are loosely related to existing standards, as the existing standards are simply too generically defined to be useful in an operational platform. The metrics as currently defined by IPPM, for example, leave many open parameters, such as the so-called "P-Type" (packet type), which allows any of the metrics to be defined with any type of packet. A consequence of this was the failure of an initial attempt to define an IPPM metric registry [12] which was subsequently deprecated [13] due to its lack of usability: the combinatorial nature of the open parameters in the metric rendered the registry useless. In order to produce usable metrics that produce comparable results it is necessary to define the metrics more narrowly. This not only implies a reduction in the number of open input parameters, but also the necessity of a more precise definition of the conditions in which the tests are executed. For example, it is important to define if 'cross traffic' is present or not during the execution of the measurement, since it may affect the results. The challenge here is to limit the number of possible tests to the ones that are useful in the real world and to define them precisely enough so that two implementations performing the same tests produce comparable results. The IPPM working group has made an initial attempt in that direction [14].

Bulk Transfer Metrics

Many if not all of the existing platforms perform tests on TCP throughput. This is only natural, as TCP performance is one of the main benchmarks that users, ISPs and regulators care about since it is the foundation for most services and applications. However, the definition of a proper metric for TCP throughput has been the holy grail of the measurement community for the last 20 years. The problem is discussed in [15]. The fundamental problem is that TCP specifications leave many aspects open to implementers, and thus different TCP implementations perform differently even if they all comply with the TCP specification. An additional

problem is that different operating systems implement different flavors of TCP, which again affects the performance. The proper definition of a TCP throughput metric that reflects the actual user experience remains elusive.

Regulatory implications

Another challenge that a standardization effort in this area is likely to encounter is related to the involvement of the regulator and the implications that such involvement may have for the ISPs. As we mentioned already, several of these platforms have been used by regulators to benchmark Internet access products in different countries. Producing standards in this area would potentially allow the regulators to mandate that ISPs support these standards. The ISPs may be concerned about the potential additional costs.

SDO competition

Last but not least, there is the challenge of which Standard Development Organization should cover which aspects of this work. As of today, there are at least three SDOs working on this, namely the IETF, the BBF and the IEEE. While there is some coordination between them through liaisons, it remains to be seen whether they will coordinate or they end up working in overlapping parts of the problem space, resulting in partially competing standards.

7. CONCLUSIONS

There is a lot of interest from different parties, ranging from researchers and operators to regulators to assess the state of the Internet, where "state" refers to a wide range of various metrics. This interest has led to a number of measurement platforms of considerable size that are deployed on the Internet today.

These platforms all differ, but on a higher layer of abstraction one can make out components that most platforms implement in one way or the other. This paper attempts to describe which aspects of these components would benefit from standardization along with the benefits and challenges of doing so.

It appears that while the most urgent place for standardization is in the general area of metrics and measurement methodologies, there are other components that would benefit from standardization. In some use cases targeted by existing measurement platforms, standardization on data export and control protocols would come with significant benefits for parties that are interested in the deployment and control of these infrastructures. This is particularly so when such infrastructure involves measurement agents from different vendors or operated by different parties.

The first steps towards creating standards in this area have been taken and at least three SDOs have initiated some form of activity. Since the involved parties are, however, interested in very different and potentially conflicting goals regarding these activities, the road to standardization will be a difficult one and there are many potential problems the standardization process will face.

8. ACKNOWLEDGMENTS

The work of Marcelo Bagnulo, Trevor Burbridge and Philip Eardley has been partially funded by the EU-FP7 Leone project.

The work of Brian Trammell and Rolf Winter has been partially funded by the EU-FP7 mPlane project.

9. REFERENCES

[1] http://www.samknows.com.

[2] SamKnows, July 2011. Test Methodology White Paper.

[3] N. Feamster, May 2011. TheBISMark Project: Broadband Measurements from the Network Gateway.

[4] R. Kisteleki, May 2011. RIPE Atlas, First results.

[5] http://www.ookla.com/

[6] I. Canadi, P. Barford, J. Sommers, 2012. Revisiting broadband performance. In *Proceedings of the 2012 ACM conference on Internet measurement conference*, DOI=http://doi.acm.org/10.1145/2398776.2398805

[7] C. Kreibich, N. Weaver, B. Nechaev, V. Paxson, 2010. Netalyzr: illuminating the edge network. *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. DOI=http://doi.acm.org/10.1145/1879141.1879173

[8] H. Schulzrinne, W. Johnston, J. Miller, September 2012. Large-Scale Measurement of Broadband Performance: Use Cases, Architecture and Protocol Requirements. Internet-Draft, draft-schulzrinne-lmap-requirements-00

[9] WT-304. Broadband Access Service Attributes and Performance Metrics. http://www.broadband-forum.org/technical/technicalwip.php

[10] TR-143, May 2008. Enabling Network Throughput Performance Tests and Statistical Monitoring. http://www.broadband-forum.org/technical/download/TR-143.pdf

[11] P802.16.3 Project: Mobile Broadband Network Performance Measurements. http://www.ieee802.org/16/mbnpm/index.html

[12] E. Stephan, August 2005. IP Performance Metrics (IPPM) Metrics Registry. BCP 108, RFC 4148.

[13] A.Morton, April 2011. RFC 4148 and the IP Performance Metrics (IPPM) Registry of Metrics Are Obsolete. RFC 6248.

[14] M. Bagnulo, T. Burbridge, S. Crawford, P. Eardley, A. Morton, January 15, 2013. A registry for commonly used metrics. Independent registries. Internet-Draft, draft-bagnulo-ippm-new-registry-independent

[15] M Mathis, M Allman, July 2001. A Framework for Defining Empirical Bulk Transfer Capacity Metrics. RFC 3148.