

Bauer, Steven; Lehr, William

Conference Paper

Measuring Mobile Broadband Performance

29th European Regional Conference of the International Telecommunications Society (ITS):
"Towards a Digital Future: Turning Technology into Markets?", Trento, Italy, 1st - 4th August,
2018

Provided in Cooperation with:

International Telecommunications Society (ITS)

Suggested Citation: Bauer, Steven; Lehr, William (2018) : Measuring Mobile Broadband Performance,
29th European Regional Conference of the International Telecommunications Society (ITS):
"Towards a Digital Future: Turning Technology into Markets?", Trento, Italy, 1st - 4th August, 2018,
International Telecommunications Society (ITS), Calgary

This Version is available at:

<https://hdl.handle.net/10419/184930>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Measuring Mobile Broadband Performance

Steven Bauer and William Lehr¹
MIT

Abstract

The FCC has been running the Measuring Broadband America program since 2011. They have released six well-received yearly reports so far, analyzing the data collected from wired broadband providers. In 2012, the FCC announced an analogous Measuring Mobile Broadband Performance effort and they have been collecting data since 2014. However, the FCC has not released any reports analyzing the mobile data. In this paper we investigate the technical and policy challenges confronting this and other efforts to collect and report on mobile broadband performance.

At the core of the debate over mobile performance measurement are questions such as: what constitutes scientifically valid performance measurement for mobile broadband networks? How does one draw valid inferences about user experience, correctly attributing impairments when they exist to different actors or components in the system? Or, how can mobile performance measurement reporting protect privacy while still remaining informative? Answering these questions is more difficult for mobile than fixed broadband for multiple reasons that we document.

We explore the current status of mobile broadband performance measurement, building on earlier work that focused on measuring the performance of fixed broadband services.² Traditionally, drive and walk testing of mobile networks (conducted by companies such as RootMetrics and P3 Communications) have competed with crowd-sourced measurements (like Ookla's Speedtest app and the FCC's app). For each approach, we examine the measurement methodology, analysis procedures, and data availability. We find that while information describing the general structure of the measurement and reporting methodologies is typically available, important details regarding the measurement methodology and how the raw data is modified to produce the final reports is often missing.

We dispute published claims³ that driving testing is "scientific," whereas other methodologies are suspect. We conclude that there is no single best approach and that the availability of a

¹ This work was supported in part by the U.S. National Science Foundation under Grants 1345256, 1413973, and 1413905. The views expressed here are the authors' own. The authors may be contacted at Stephen Bauer (corresponding author) bauer@mit.edu; and William Lehr, wlehr@mit.edu.

² See, for example, Bauer, Lehr and Mou (2016), Bauer, Lehr and Hung (2015), Bauer, Clark & Lehr (2010), and Bauer, Clark & Lehr (2009).

³ See Recon Analytics (2016), "Scientific Testing: the most reliable way to test a network," White Paper prepared by Recon Analytics LLC, August 2016, available at <http://reconanalytics.com/wp-content/uploads/2016/08/Scientific-Testing-Report-2016-08-09.pdf>.

multiplicity of measurement platforms and approaches offers significant benefits. Although there is no single best mobile measurement approach for all questions, we explain why it is important to carefully match the test methodology and available data to the questions being asked.

Finally, our analysis explores the implications for broadband policy of this evolving ecosystem for mobile broadband measurement.

Table of Contents

1. Introduction.....	2
2. Motivations for Mobile Broadband Measurements	3
3. Challenges in Measuring Mobile Broadband	5
4. Comparison of Mobile Measurement Methods	8
4.1. Drive Testing	11
4.2. Walk Testing.....	15
4.3. App Based Panels for Mobile Testing	16
4.4. Crowd sourced mobile measurements	18
5. Summing Up and Further Thoughts	19
6. References.....	20

1. Introduction

The FCC has been running the Measuring Broadband America program since 2011. They have released six well-received yearly reports so far, analyzing the data collected from wired broadband providers. In 2012, the FCC announced an analogous Measuring Mobile Broadband Performance effort and they have been collecting data since 2014. However, the FCC has not released any reports analyzing the mobile data. In this paper we investigate the technical and policy challenges confronting this and other efforts to collect and report on mobile broadband performance.

At the core of the debate over mobile performance measurement are questions such as: what constitutes scientifically valid performance measurement for mobile broadband networks? How does one draw valid inferences about user experience, correctly attributing impairments when they exist to different actors or components in the system? Or, how can mobile performance measurement reporting protect privacy while still remaining informative? Answering these questions is more difficult for mobile than fixed broadband for multiple reasons that we document.

We explore the current status of mobile broadband performance measurement, building on earlier work that focused on measuring the performance of fixed broadband services.⁴ Traditionally, drive and walk testing of mobile networks (conducted by companies such as RootMetrics and P3 Communications) have competed with crowd-sourced measurements (like the Ookla's Speedtest app and the FCC's mobile app). For each approach, we examine the

⁴ See, for example, Bauer, Lehr and Mou (2016), Bauer, Lehr and Hung (2015), Bauer, Clark & Lehr (2010), and Bauer, Clark & Lehr (2009).

measurement methodology, analysis procedures, and data availability. We find that while information describing the general structure of the measurement and reporting methodologies is typically available, important details regarding the measurement methodology and how the raw data is modified to produce the final reports is often missing.

We dispute published claims⁵ that drive testing is "scientific," whereas other methodologies are suspect. We conclude that there is no single best approach and that the availability of a multiplicity of measurement platforms and approaches offers significant benefits. Although there is no single best mobile measurement approach for all questions, we explain why it is important to carefully match the test methodology and available data to the questions being asked.

Finally, our analysis explores the implications for broadband policy of this evolving ecosystem for mobile broadband measurement.

2. Motivations for Mobile Broadband Measurements

Measurement of mobile broadband performance in the United States as well as globally supports multiple public policy objectives:

- 1) Provides end users with information on mobile network performance and coverage, helping them to make better choices when selecting providers and plans, and managing their usage;⁶
- 2) Helps policymakers better target universal service funds to address issues in underserved areas;⁷

⁵ See Recon Analytics (2016), "Scientific Testing: the most reliable way to test a network," White Paper prepared by Recon Analytics LLC, August 2016, available at <http://reconanalytics.com/wp-content/uploads/2016/08/Scientific-Testing-Report-2016-08-09.pdf>.

⁶ Market efficiency depends on the ability of buyers and sellers to make informed choices. Consumers need to understand the different capabilities of their broadband options, including their wireless choices (both from cellular operators and options, including self-provisioning) to plan their investments in devices, applications, content services, as well as their subscription choices. Consumers with access to independent (and presumptively, trustworthy) sources of transparent performance information with which to evaluate providers is important in making initial subscription decisions, which is a major source of competitive discipline for providers. However, performance information, including real-time access, can also assist users in moderating their behavior (e.g., choosing where or when to access rich media and on which devices). This also provides important market signals that can contribute to making markets more competitive and efficient.

⁷ Policymakers are interested in making sure that everyone has access to broadband, and ideally, to both mobile and fixed broadband services since they are imperfect substitutes for each other, and offer complementary capabilities (mobility vs. higher data rate/lower data cost, typically). For some subscribers, mobile broadband may be their only available or effective option for obtaining broadband access (e.g., for the homeless trying to find a job), while for others, it may offer valuable complementary benefits or a preferred choice relative to fixed (e.g., cord-cutters), and for still others, living in communities with both fixed and mobile broadband available may still be under-served with respect to the range of competitive choices available. For further discussion, see Lehr, W (2009) "Mobile Broadband and Implications for Broadband Competition and Adoption," a white paper prepared on behalf

- 3) Supports evaluation of mobile network operators (MNOs) to ensure they are meeting the coverage and quality obligations associated with their license obligations and other regulatory commitments;
- 4) Creates incentives and strengthens competition between operators; and,
- 5) Provides empirical basis for resolving disputes associated with customer performance.⁸

Analogous to our earlier work documenting and critiquing speed measurements of wired networks, we recognize that the debate over mobile broadband performance measurement engages the interests of all market participants – policymakers, users, and ISPs alike, and consequently will need to reconcile strategic interests that may be expected to conflict. While we do not believe there is a single method for measuring and reporting mobile broadband performance that is appropriate for all contexts, we recognize the analytic and cost benefits from narrowing the range of methods and from moving toward collective agreement on best practices. Although we believe multiple measurement methods are desirable both because no single method is appropriate for all decision-making contexts and because there is virtue in having multiple sources of measurements (to encourage trust if users can select a measurement source and method they may be more comfortable with and to support cross-validation), we also believe that it is possible to identify inappropriate or misapplied methods. We believe the measurement research community has a role to play in identifying and calling out measurement practices and data that do not reflect best practices, or worse, are simply in error. Having a few generally accepted methods for measuring mobile broadband and having clear documentation to support published measurements will aid data aggregation, analysis, and interpretation.

In this paper, we explain why the mobile broadband measurements at times vary significantly. Differences in measurement objectives and methodologies account for most of the discrepancies. The proper interpretation is not to try and select a best measurement approach, but to encourage the emergence of a healthy measurement ecosystem. As we explain, having multiple, diverse sources and approaches for broadband performance measurements, and especially for mobile broadband performance will enhance the benefits of measurement overall and will enable better insights into mobile broadband performance and the user experience.

of Broadband for America, November 2009, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2446011.

⁸ Potential disputes may arise between any of the potential stakeholders, including customers, MNOs, edge-providers of applications and content, equipment providers, or stakeholders. Having empirical data available is often an important basis of evidence that can aid in the resolution of such disputes. Moreover, the anticipation that evidence will be available can have a salubrious impact on the incentives of stakeholders to honor their commitments. However, if the potential sources of data and how it may be used are not trustworthy then the prospect of its availability can contribute to making matters worse. Having multiple sources of measurement data can contribute to having a more trust worthy data ecosystem. For further discussion of some of these issues, see Lehr, W., E. Kenneally, and S. Bauer (2015), "The Road to an Open Internet is Paved with Pragmatic Disclosure & Transparency Policies," The 43rd Research Conference on Communication, Information and Internet Policy (TPRC.ORG), available at SSRN: <https://ssrn.com/abstract=2587718> or <http://dx.doi.org/10.2139/ssrn.2587718>.

3. Challenges in Measuring Mobile Broadband

Understanding mobile broadband performance is more difficult than fixed broadband for many reasons. Fundamentally new questions are raised that have to be answered for mobile broadband measurement. For instance, no one would think about measuring fixed broadband except at locations where fixed broadband already exists, but in measuring mobile broadband the potential exists to ask about broadband performance in places where actual users have not yet even tried to use a mobile device. In mobile there are an infinite number of geographic locations, orientations, client devices, and surrounding context that could be tested or measured. Infinite is an appropriate characterization here because very small changes along the continuous dimensions of the location, orientation, and surrounding environment of a mobile device sometimes have a significant impact on measured performance.⁹

Mobile itself is an important qualifier on broadband to distinguish this class of measurement from measurements of fixed wireless networks. Fixed wireless networks are, for the most part, naturally tested using methodologies similar to fixed wired broadband.¹⁰ Indeed, the FCC's Measuring Broadband America has included satellite broadband providers offering residential broadband services for several years.

So what types of mobility are actually being tested?¹¹ Testing of mobile networks might be conducted by physically moving a test device or sets of test devices about a geographic area.

⁹ The capacity of radio frequency (RF) spectrum to support simultaneous communications is interference limited and since interference takes place at a receiver when it cannot disentangle the intended signal from other RF signals, the capacity of spectrum is technology-limited. With advances in radio technology (smart antennas, new modulation schemes, MIMO, cognitive radios) and new wireless architectures (5G, smaller cells), the capacity of the RF to support multiple simultaneous users in the same spectrum has greatly increased, helping to facilitate the growth of wireless services of all types. Small variations in antenna placement and orientation, modulation of the power, changes in the choice of frequency used (which impacts the ability of signals to penetrate objects), and a host of other factors can significantly impact the quality of the RF connections ability to support data transport. These problems arise with wired connections but to a much lesser degree. For further discussion, see Matheson, R., & Morris, A. C. (2012). The technical basis for spectrum rights: Policies to enhance market efficiency. *Telecommunications Policy*, 36(9), 783-792, available at https://www.brookings.edu/wp-content/uploads/2016/06/0303_spectrum_rights_matheson_morris.pdf; and Lehr, W. and J. Chapin (2010), "On the convergence of wired and wireless access network architectures," *Internet Economics and Policy*, 22 (2010) 33-41.

¹⁰ The challenges of measuring performance associated with RF do arise with fixed wireless broadband (e.g., the PHY-layer variability is greater factor with wireless than with wired solutions, see Footnote 9), but arise more prominently when trying to diagnose the source of problems and understanding the performance issues associated with satellite-based services (e.g., greater latency than for terrestrial services, more susceptible to climate-related interference issues like clouds, etc.). Such complexities do arise when comparing fixed wireless (terrestrial and satellite based) and fixed wired broadband options, but these are secondary.

¹¹ Within the Future Internet Architecture (FIA) research community, richer notions of mobility are under consideration (see <http://mobilityfirst.winlab.rutgers.edu/>). Mobility can occur in any dimension – time, space *or* context. The sort of mobility we are concerned with here relates to mobility of the end-user's location in time and space, however, the larger challenge is to enable mobility of all network resources.

Tests might be conducted either while the device is being moved (at various speeds) or while the device is temporarily stationary at fixed locations (say standing in a subway station). The precise model of geographic mobility being tested could vary considerably. One objective might simply be to gather a measurement sample in as many geographic locations as possible being somewhat indifferent to the physical speed of the device while the measurement is being taken. The mobility of such measurements might be naturally dictated by existing vehicular or pedestrian traffic in an area. The selection of which roads or paths through an area then would define the mobility model being tested – is a test taken on the highway while driving at 65 miles per hour or at the gas station adjacent to the highway.¹² Both would result in a measurement sample in the same geographic area. Differing models of mobility therefore imply testing different capabilities of a mobile broadband service (seamless handoffs between cell sites in a fast driving car, switching models between LTE and WiFi when users transition between different contexts such as car to home, etc).

Moreover, since the performance of broadband (fixed or mobile) varies over time,¹³ the mobile measurements that are challenged with having to take measurements at many different locations typically, are less likely to have as many measurements over time that are comparable, unless many more mobile measurements are collected at the multiple locations.

How performance is characterized as devices move about the physical space and are handed off between cell sites, roaming partners and wireless technologies at various speeds matters greatly for how performance statistics are reported and aggregated. For example, cellular roaming might be attributed to either the home cellular provider's network or the visited network. Failure to roam efficiently (interpreted here to mean failure to select the fastest available wireless connectivity solution) may account for poor measurement performance e.g., Google's Fi telephone service prefers WiFi whereas other cellular providers may prefer cellular networks and how those choices are made may be controlled by the cellular provider, the end-user, the device, or the application – thereby altering the interpretation of results.

Entire networks (groups of nodes) may move at once. For example, during an emergency, public safety officials may switch an entire group of first-responders from one set of network authorities to another, requiring the migration of data, connections, and potentially even spectrum resources. A user or application may move its network connection (e.g., moving from cellular to WiFi, or switching from wireless to fixed broadband), or the context of the connection may change (e.g., the user switches to viewing health records subject to stricter data protection rules, the user switches between social and business communications, etc.).

¹² Of course, data analysis techniques may be used to extend the analytic results, significantly improving the quality of the inferences that can be derived. With enough data, measurement approaches that offer non-overlapping inference potentials at small sample sizes may become closer substitutes as the size of the samples becomes large enough. The data analysis (inference approaches) that may be applied can vary significantly in sophistication, which adds another layer of complexity in terms of comparing different performance reports, even when those are based on the same data. However, a truism of data analysis that still pertains is that the raw input data (i.e., the measurements) matter a lot, or to put it more succinctly: "garbage in, garbage out".

¹³ Broadband performance varies over all time-scales. Time-of-day (coinciding with peak usage behavior) variations are typical as are longer term variations (e.g., reflecting changes in network infrastructure as MNOs upgrade capacity and make other changes impacting performance in phases).

What applications to test in different mobile contexts is another dimension that further complicates mobile measurement. Users increasingly are accustomed to nearly ubiquitous coverage for some applications under the most stringent of mobility models (i.e., voice/text message everywhere, support for mapping application, etc. at most vehicular speeds), but users have also been acclimated to playing an active role in helping to improve service quality (e.g., moving to where they know they get the best signal when using their smartphones or in selecting which applications to utilize where). Furthermore, expectations of performance for high-resolution video on the other hand may assume scenarios with limited, nomadic-style mobility.¹⁴ However, perhaps we are heading toward a world where even those applications are expected to perform well on mobile networks as passengers stream video content while travelling at highway speeds. On the other hand, tests of bulk upload and download speeds that attempt to understand applications performance scenarios like uploading vacation photos or downloading a movie before a long trip might be testing under mobility scenarios where a user is not moving in geographic space as much.

A different mobile measurement challenge arises when the objective of mobile measurement is to deliberately try to seek out and find the locations and conditions under which mobile service expectations are not being met.¹⁵ If the goal is to identify coverage (availability) gaps where no mobile broadband coverage exists, the objective is not to characterize geographic areas where users already are using mobile broadband, instead it is to identify dead spots within otherwise covered areas or finding the boundaries of existing coverage zones.¹⁶ These types of measurements are used to target and justify funding for universal service and to evaluate whether license obligations are not being met (with respect to coverage and build-out requirements).

¹⁴ Most mobile video consumption is quasi-stationary. Users migrate to locations where they can have an acceptable viewing experience, which includes having acceptable wireless connectivity. Although passengers may want to watch videos at highway speeds, it is not something we want the drivers doing and pre-downloading videos for high-speed transport access (in cars and planes) is readily available and often offers a better option than mobile streaming, or reduced resolution (acceptable on small-screen devices such as smartphones) may change the application performance requirements.

¹⁵ Identifying where fixed broadband services are available is challenging, but is significantly easier than for mobile broadband services. When the availability of fixed broadband services was undertaken in 2010, the data was collected from providers on a Census Block (CB) level. CBs vary in size but were designed to capture fixed broadband availability mapping data collected that information on a Census Block Level (CBs) level. CBGs vary in geographic size by location (for example a single block in a city but potentially much larger in rural areas), but are designed to capture an area with up to a few hundred inhabitants. Mobile providers may claim service is available in their licensed coverage areas even when availability is limited to only a portion of the coverage area, and license coverage areas are typically larger than CBs. Moreover, given the vagaries of RF propagation, there are typically many anomalies, irregularities, and coverage gaps in actual coverage when compared to the coverage predicted by standard propagation models. The availability of coverage and capacity for wireless networks depends on how far users are from the base station and what else is going on in the wireless environment. While similar problems arise with wired networks, the variability with distance and environmental factors is much greater with wireless.

¹⁶ As users of WiFi and cellular services readily experience, dead spots can occur over fine geographic distances such inside one's home.

While fixed broadband services are associated with a specific physical location and are typically shared among multiple users in that residence, mobile broadband tends to be associated with a single user. This is beginning to change as mobile broadband serves collections of users and devices in vehicles or contexts where mobile is a substitute for wired connectivity.¹⁷ The prevalence of single users, however, presents privacy challenges for measurement methodologies that rely upon testing from actual user devices. These privacy challenges impose significant constraints on data collection. While access to fixed broadband metadata (ignoring content, but tracking timing and destination of traffic for broadband user) poses a threat to individual privacy, the privacy concerns surrounding mobile broadband data are significantly greater (the capture of location data in particular).¹⁸

4. Comparison of Mobile Measurement Methods

In this section we examine multiple methods for measuring mobile broadband performance. For each approach, we attempt to examine the measurement methodology, analysis procedures, and data availability. We find that while information describing the general structure of the measurement and reporting methodologies is typically available, important details regarding the measurement methodology and how the raw data is modified to produce the final reports is often missing. As we explain, this lack of methodological transparency makes it difficult for the research community and others to discriminate between good and bad implementations of measurement methodologies, and without this capability, it will be more difficult to build trust in the evolving measurement ecosystem.

In general, each methodology offers different benefits and drawbacks and having more than one approach enables opportunities for cross-validating results.¹⁹ Recognizing this, many national regulators as well as private firms increasingly employ multiple approaches seeing the differing perspectives provided as complementary. A diagram from a management consultancy that provides engineering support services illustrates how they combine measurements from multiple sources into a (marketing term) "360-degree" perspective on mobile involving driving testing, constructed panels, as well as crowded-sourced public data.

¹⁷ Although mobile hotspots and device-applications may enable sharing of mobile services among multiple users and family subscription plans enable sharing of data quotas, the usage of mobile devices is more typically associated with a single user and is more likely to be personalized.

¹⁸ Observing the behavior of fixed broadband traffic from a home raises significant privacy challenges, which are duplicated and worsened (because of the increased personalization of the service and the greater resolution that affords in deriving inferences about the traffic that that enables) with a mobile user in the user's home, however, the privacy risks are significantly amplified with mobile broadband because observations of the mobile user's traffic enables tracking the user's behavior through time and space, not just when the user is in a particular location.

¹⁹ Indeed, pooling information from several measurement sets may provide more information than the aggregate of the parts; however, due care must be taken when trying to pool data.

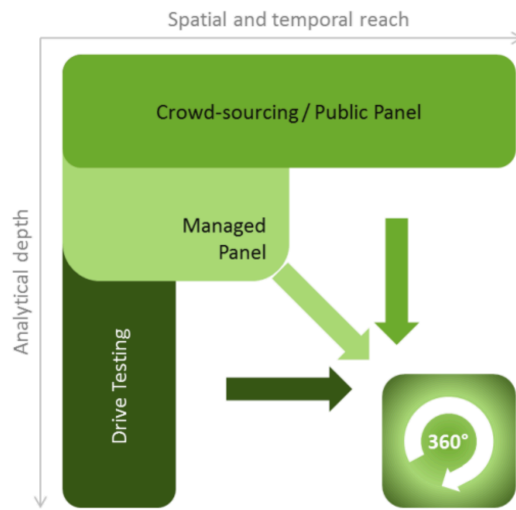


Figure 1: P3 Communications²⁰ integration of data from multiple testing methodologies to provide combined perspective on mobile performance.

This aligns with our own view that different measurement methodologies often offer unique and valuable insights. We note that this is in contrast to statements from some stakeholders that one method is scientifically superior to others.²¹ We recognize such claims as marketing hype and find it unfortunate and misleading in light of the real and important scientific issues that do arise in trying to assess the merits of alternative mobile measurement strategies. We explain further below why we think rank-ordering methodologies a fool's errand since different contexts call for different measurement strategies, but do believe that thoughtful reviews and comparisons of alternative measurements is important to identify bad measurements (of potentially good methodologies, or worse, of bad methodologies) and to differentiate among good measurements that may be more or less appropriate in different decision contexts.

In a move we applaud, the FCC in a public notice explaining mobile broadband providers transparency rule requirements allowed mobile provide providers a great deal of latitude in how the provider elected to measure actual end user performance. Indeed, all the methodologies we discuss in this paper were permitted.

“Mobile BIAS providers may, for example, achieve a representative sampling of end users by running measurement clients on end-user devices (e.g., using consumer speed test data), by placing measurement clients in locations near a representative set of mobile broadband consumers (e.g., by combining drive-test data with an estimate of the reduction in speed from drive-test locations to user locations), or by measuring performance in the network and estimating the relationship between this measured performance and that experienced by end users.”

²⁰ See https://www.btg.org/wp-content/uploads/2016/09/20160922_P3_Amsterdam_SmallCell_Event_Huawei_PUBLIC.pdf.

²¹ See Footnote 5 *supra*.

Similarly, a draft report released by BEREC²² of a summary of responses from EU member countries (see Table 1) highlights the diverse methodologies being employed within each country to understand mobile coverage. The most common methodology utilized by the National Regulatory Authority (NRA) or component authority in each country was drive testing which was utilized in some manner by twenty-seven of the thirty-four respondents. Theoretical models of mobile coverage were used next most often in eleven of the countries. Eight of the countries employed some type of app-based testing, either panel based (one country) or crowd sourced (eight with two more having plans to do so in the future). Furthermore, within these countries, there may be more than one entity conducting measurements and they may be employing additional diverse measurement strategies. Illustrating this, we present panel and crowd sourced app testing of mobile broadband in the UK, even though the UK BEREC respondents said that the UK relied on drive and walk testing alone.

This attests to the fact that the evolving mobile broadband measurement ecosystem is diverse and rapidly changing, and not all parts of it know what others are doing. The difficulty of tracking developments in this rapidly changing environment has posed a real challenge for the authors of this paper and demonstrates the utility of releasing a draft as a status check in July 2018.²³

²² See Appendix A in BEREC's "Draft Common Position on monitoring mobile Coverage" June, 2018. Note that is explicitly targeting assessments of coverage. (downloaded on July 1, 2018 from the https://berec.europa.eu/eng/document_register/subject_matter/berec/public_consultations/8165-draft-common-position-on-monitoring-mobile-coverage).

²³ When we were writing Bauer, Clark, and Lehr (2010), we were challenged by the measurement platforms making changes in real-time, occasionally in response to issues we had brought to their attention (see Bauer, S., D. Clark, and W. Lehr (2010), "Understanding broadband speed measurements," 38th Research Conference on Communication, Information and Internet Policy (TPRC), George Mason University School of Law, Arlington, VA, October 1-3, 2010).

Country Code and NRA or Competent Authority	Theoretical calculation	Drive Testing	Walk Testing	App-based (Panel)	App-based (Crowdsourcing)
AT Austria (RTR)	No	No	No	Yes	Yes
BE Belgium (BIPT)	No	Yes	No	Not yet	Not yet
BG Bulgaria (CRC)	Yes	Yes	No	No	No
CH Switzerland (BAKOM)	No	No	No	No	No
CY Cyprus (OECPR)	No	No	No	No	No
CZ Czech Republic (CTU)	Yes	Yes	No	No	Yes
DE Germany (BNetzA)	No	Yes	No	No	No
DK Denmark (DBA)	No	No	No	No	No
EE Estonia (ETRA)	No	Yes	No	No	No
FI Finland (FICORA)	No	Yes	No	No	No
FR FRANCE (Arcep)	No	Yes	No	No	No
FYROM Macedonia (AEC)	No	Yes	No	No	No
GR Greece (EETT)	No	Yes	No	No	No
HR Croatia (HAKOM)	No more	Yes	No	No	Yes
HU Hungary (NMHH)	No	Yes	No	No	No
IE Ireland (ComReg)	Yes	Yes	Yes	No	Eventually
IS Iceland (PFS)	Yes	Yes	No	No	No
IT Italy (AGCOM)	Yes	Yes	No	No	Yes
LT Lithuania (RRT)	Yes	Yes	No	No	No
LV Latvia (SPRK)	No	No	No	No	No
ME Montenegro (EKIP)	Yes	Yes	No	No	No
MT Malta (MCA)	No	Yes	No	No	No
NL Netherlands (ACM)	No	Yes	No	No	No
NO Norway (NKOM)	Yes	Yes	No	No	Yes
PL Poland (UKE)	No	Yes	No	No	No
PT Portugal (ANACOM)	Yes	Yes	No	No	Yes
RO Romania (ANCOM)	Yes	Yes	No	No	No
SE Sweden (PTS)	No	No	Yes	No	No
RS Serbia (RATEL)	No	Yes	Yes	No	Yes
SK Slovakia (RU)	Yes	Yes	No	No	No
SI Slovenia (AKOS)	Yes	Yes	Yes	No	Yes
TR Turkey (ICTA)	No	Yes	No	No	No
UK United Kingdom (OFCOM)	No	Yes	Yes	No	No

Table 1: Survey of National Regulator Authorities from BERIC's current draft "Common Position on monitoring mobile coverage" June, 2018.

4.1. Drive Testing

Drive testing is a class of measurement methodologies that evaluate the coverage and quality of service provided by mobile networks using common measurement equipment mounted in cars that are used to systematically collect measurements. The most important distinguishing feature of drive testing is the consistent collection of measurements -- minimizing the variation in as many factors that might influence measurement results as possible. The ability to systematically control significant variables that are known to impact performance (e.g., collecting a sample that reflects consistent speed, time-of-day, application use, or device configuration) enhances the ability of analysts to draw inferences and control for measurement noise. This style of mobile testing is employed to provide a common picture of the quality of service of mobile over a predetermined, surveyed area.

While any single drive test organization will have a well-defined methodology, there is no standard drive test methodology across organizations. So, for instance, the drive test methodologies employed by Root Metrics differs from the drive test methodologies employed by California in its CalSPEED project.

There are generally two main types of drive testing. Testing that utilizes typical consumer grade phones (such as shown in Figure 2 below) as well as driving testing that uses specialized receivers.²⁴



Figure 2: Measuring cases with cell phones (left) are mounted in windows of drive test car (right).

Typical metrics collected from a phone-based drive test configuration include those listed in Table 2 below. These metrics range from lower level radio statistics such as received signal strengths to higher-level metrics such a download throughputs or ping times.

Measurement	Description	Use
Protocol Log	RRC and NAS protocol log	Capturing cell broadcast and system information Performing calculation on timing of events Troubleshooting failures
Cell-ID / E-ARFCN	Physical cell identity/E-ARFCN	Confirmation of the carrier in use; identifying the cell sector and base station in use
RSRP	Reference signal received power	Measure cell site received signal coverage in dBm (3GPP)
RSRQ	Reference signal received quality	Measure of cell site received signal quality in dB (3GPP)
SINR/CINR/SNR	Channel quality measure	Supplemental to RSRQ; mobile-manufacturer-specific implementations; not 3GPP
DL-SCH/UL-SCH Throughput	DL/UL shared channel throughput	Measure of throughput the mobile unit is achieving on shared UL or DL channels
DL-SCH/UL-SCH BLER	DL/UL shared channel BLER	Measure of error rate mobile unit is experiencing on the UL or DL shared channels
#DL RB UL RB	DL/UL resource blocks allocated	Measure of how much of the shared channels are allocated to the mobile by the scheduler.
#DL MCS/UL MCS	DL/UL modulation and coding	Indication of the efficiency of the transport channel under current network conditions; range is BPSK/QPSK/16QAM/64QAM
Wideband/Narrow Band CQI	WB / NB channel quality information	Measures the channel quality across the allocated frequencies; used for link adaption
Rank	Number of TX/RX paths in use.	Determines where MIMO is active
UETX Power	Mobile transmit power	Monitors how much power the mobile needs to use to reach the eNodeB; used for link adaption.
Application Throughput	Performance of the end-user application	Measure of the end-user experience; making both UDP and TCP measures
Ping	Length of time to reach a specific IP address	Measures network latency

Table 2: Phone based metrics collected during a drive test (Table copied from [1])

Other drive testers employ specialized receivers that are able to obtain a raw view of the RF spectrum availability, unconstrained by the limitations of conventional phone hardware or network provider device settings. Employing such specialized receivers can allow the testers to identify sources of RF interference, enabling better insight into the quality of the physical layer connectivity that the mobile broadband service is relying on. Figure 3 below shows how

²⁴ See <https://www.viavisolutions.com/en-us/literature/drive-testing-lte-white-paper-en.pdf>

interfering signals that exceeded a configured threshold may be identified along a drive test path. Also, the receiver hardware can be tuned to monitor both a network of interest as well as competitors in a given area. This can be used to support the analysis of how coverage compares between competing providers with time/location matched samples (something that is otherwise often difficult to accomplish).

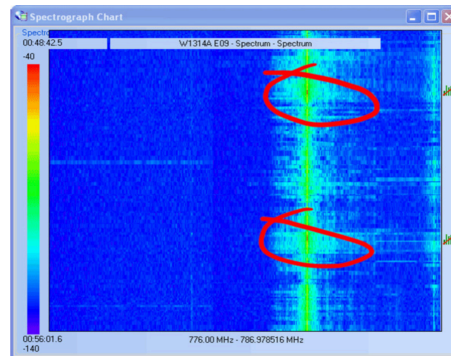


Figure 3: Spectrograph example demonstrating the identification of interfering signals that exceeded a configured threshold along a drive test path. From <https://www.viavisolutions.com/en-us/literature/drive-testing-lte-white-paper-en.pdf>

Drive test methodologies can also be constructed to test typical user applications and activities including web browsing tests and video streaming tests such as tests of YouTube. A high-level description of the drive test methodology employed by a P3 connect mobile benchmark test conducted in Sweden in 2017 is described below.²⁵

The P3 connect Mobile Benchmark in Sweden took place from August 19th to September 9th, 2017. All samples were collected between 8am and 10pm. The network tests covered 21 larger cities, of which 7 count more than 100 000 inhabitants. Additionally, our test routes led through 33 smaller towns as well as the connecting roads. This combination of test areas had been carefully selected to provide a significant series of test results covering the Swedish population. The areas chosen for the 2017 test account for more than 4 million people, or 42 per cent of the total Swedish population.

P3 conducted the tests with two drive-test cars, equipped with arrays of Samsung Galaxy S7 Cat 9 smartphones (Voice) as well as a mixed allocation of Samsung Galaxy S7 and Sony Xperia XZ Cat 9 smartphones (Data) for the simultaneous measurement of voice and data services. Data performance was measured using four smartphones in each car – one per operator. One car was equipped with four Samsung Galaxy S7 while the other car was carrying four Sony Xperia XZ in order to respect the variable data performance of different smart phones in different networks. In order to further reflect the customer experience, the radio access technology was set to LTE preferred mode. The web tests accessed web pages according to the widely recognized Alexa ranking. In addition, the static “Kepler” test web page as specified by ETSI (European Telecommunications Standards Institute) was used.

²⁵ See Sweden 2017 Methodology, <https://www.connect-testlab.com/sweden-2017-methodology-1> (accessed on June 20, 2017).

In order to test the data service performance, files of 3MB and 1MB for download and upload respectively were transferred from or to a test server located on the Internet. In addition, the peak data performance was tested in uplink and downlink directions by assessing the amount of data that was transferred within a 7 seconds time period. Another discipline was the playback of Youtube videos. It took into account that Youtube dynamically adapts the video resolution to the available bandwidth. So, in addition to success ratios, start times and playouts without interruptions, Youtube measurements also determined the average video resolution.

All the tests were conducted with the best-performing mobile plan available from each operator.

The above description provides quite a bit of relevant information that is important to evaluate any inferences that might be based on the data, but also highlights questions of interest. For example, focusing on the best-performing mobile plans from each operator does not tell one about how those operators other plans may perform and it is unclear from the text whether the "best-performing" is also the most-commonly used plan, although perhaps that might be inferred. Furthermore, P3, which is an engineering and management consultancy²⁶ relies on proprietary measurement techniques. While this is understandable since conducting drive testing is expensive and P3's expertise is a source of competitive advantage, it does make it hard to evaluate their results. There are so many details that are left unspecified.

This differs from the CalSPEED drive test methodology. CalSPEED employs the open source popular Iperf testing tool²⁷ to conduct throughput, latency and loss tests. We consider CalSPEED here not because they are widely used but rather because they (commendably) document most of the details of their test methodology, employ open source code that we could inspect, and make available their raw drive test data. In their latest test in 2017 they employed ten drivers to conduct approximately 16,000 tests across the state of California between October 5th and November 10th. Looking at the raw data for test locations, it appears that CalSPEED drivers frequently stop at designated test locations to conduct each set of measurement tests. In Figure 4, we depict all test locations (blue dots) for a single CalSPEED driver over a single day of drive testing. The scarcity of measurements conducted on this day by one driver surprised us. This differs from most other drive testing methodologies that conduct many more measurements as vehicles are in motion over the course of a testing day.

What one should take from this discussion is not that fully-open testing is necessarily better than testing approaches that rely on proprietary methods and data, but rather recognize that each strategy offers costs and benefits. Providing full-disclosure of all aspects of the data and methods makes it harder to recover the costs of creating those measurements in the first-place and adds to the measurement costs (e.g., because of documentation and data access requirements), but it does make it easier for third-parties to validate the data and potentially implement alternative scenarios. Full-disclosure also may complicate and increase the costs of protecting confidential strategic information, and thereby may make it easier to gain cooperation from network operators in identifying information that will contribute to the informativeness of the measurements.

²⁶ See <https://www.p3-group.com/en/>.

²⁷ See <https://iperf.fr/>.

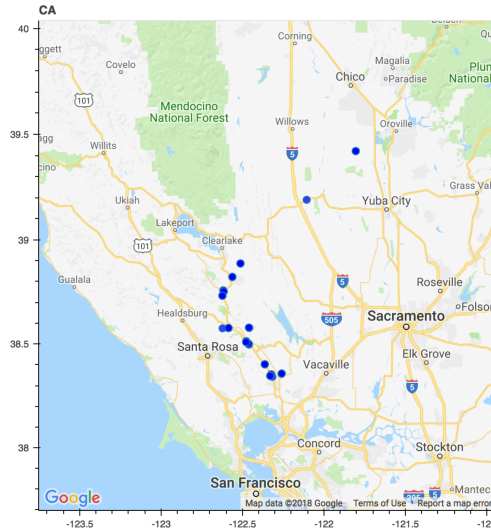


Figure 4: All test locations (blue dots) for a CalSPEED driver over a single day of drive testing. The scarcity of measurements in this drive test methodology is striking compared with other drive tests methodologies. The main takeaway point from this is that how, when and where tests are conducted in a drive test are not comparable.

To further highlight the complexity of evaluating alternative drive test approaches, consider that in the United States, the FCC guidance to mobile providers looking to fulfill their open internet transparency rule requirements required mobile providers to model or estimate the reduction in speed between outside drive-test locations and typical user locations.²⁸ We did not find any information about the actual models that providers may have used to respond to this requirement and have not been able to assess whether this was even addressed. However, we recognize that the requirement was motivated by a significant potential source for measurement discrepancies, as noted earlier. First, mobile performance is worse the further one is from the base station, but also when one is inside rather than outside. Most mobile use actually takes place indoors (and in many cases, mobile traffic is off-loaded to WiFi-connected fixed broadband services). Second, as already noted, end-users are accustomed to the way wireless service performance can be significantly impacted by even small variations in phone placement. Adjusting for differences indoor versus outdoor use depends on the nature of building materials (brick vs. wood) and different assumptions by different operators for different parts of their coverage areas could result in very significant variances in adjusted-performance.

4.2. Walk Testing

Walk testing is substantially similar to drive testing but employs network testers that walk along typical pedestrian paths or locations like subways, sports stadiums, airports and malls. In most cases walk testing is employed in areas otherwise inaccessible to car-based tests. In areas that have pedestrian access and vehicular access, sidewalks and adjacent roads in NYC perhaps, the

²⁸ The FCC Guidance on Open Internet Transparency Rule Requirements, May 19, 2016 states, “Mobile BIAS providers may, for example, achieve a representative sampling of end users by running measurement clients on end-user devices by [...] placing measurement clients in locations near a representative set of mobile broadband consumers (e.g., by combining drive-test data with an estimate of the reduction in speed from drive-test locations to user locations)”

area is most often tested with drive tests, not walk tests. Primarily this is because the per-measurement and geographic coverage costs of drive testing are lower than walk testing.

The walk test measurements are conducted with test equipment affixed to the test-taker's back or carrying harness. Figure 5 depicts such a pack with mounts for six cell phones. While the slower speed of movement of devices in walk tests place less stringent conditions on the network service than drive testing, the complexity of accounting for diverse physical environments that are typically walk tested more than make up for the difference. The test areas are typically chosen to coincide with locations where users spend more time and where access to mobile broadband is identified as being especially important (e.g., offices, homes, and high-traffic areas). The costs involved in collecting this data are large on a per-area-covered basis and so are not available in as many locations. Moreover, the area that can be covered in a given period of time (which is important since measurements that are too long separated in time may be less easily comparable) is necessarily limited by the walk speed of the individuals carrying the test devices.



Figure 5: Carrying case with mountings for multiple cell phones

According to the Federal Highway Administration there are over 4 million miles of road in the United States. While this is a large number, it is tiny compared to the potential paths a walk tester might traverse in the United States. Crafting walk testing plans that correspond to locations that are useful to gather data therefore requires consideration of the measurement objectives. Recognizing the importance of mobile broadband to the happiness of a public transportation riders might motivate incurring the costs to assess the performance of mobile broadband in subway and train stations. Unfortunately, in the literature we have reviewed so far, we have not found any detailed documentation detailing how walk test paths and plans are selected.

In the BEREC survey results we mentioned previously, only one of the responding countries, Ireland noted that they employed any form of walk testing of mobile networks.

4.3. App Based Panels for Mobile Testing

App based panel testing of mobile broadband is a means of testing mobile broadband using applications installed on regular users' devices that automatically conduct repeated network performance tests. These measurements are typically crowd-sourced since the tests typically rely on voluntary participation by the testers, who have to proactively download and install an application on their user device. Moreover, they have to be willing to allow the application to use

device and plan resources to conduct the tests.²⁹ The apps are typically designed to minimize the likelihood that they will cause congestion or otherwise interfere with the tester's own usage, but even the use of device battery power is a potential deterrent for users willingness to participate in such measurements.

Relative to drive testing, such crowd-sourced measurements render it more difficult to control for many of the factors that drive testing can address (e.g., the precise configuration of the user devices or the locations and timing of when tests are done). The need to protect the privacy of the test takers is also a significant challenge that does not arise (at least in the same way) for drive test strategies. All of these differences can make it more difficult to collect a systematic sample of measurements over a target coverage area that are comparable in time. That is, crowd-sourced data may eventually provide measurements in as many locations as drive-testing, but the time period over which those measurements are collected is likely to be longer. On the other hand, crowd-sourced data may provide more data at different times of day and over a longer time periods in those locations where the crowd-sourced test takers allow measurements to be taken the drive tests that are typically undertaken and completed during a defined time window. Also, collecting crowd-sourced data that relies on the voluntary participation of testers (who provide the test platform and other key components required to undertake tests) is much less expensive to implement on a per-measurement basis which may make it much easier to collect large volumes of data, especially if a large number of testers can be convinced to participate.

The FCC's Speed Test app for mobile devices falls into this category. The users with the app installed and running constituted the fixed panel of testers that contributed the FCC's mobile broadband measurement initiative. In the case of the FCC, the panel is open to any users that elect to install the app. See Figure 6.

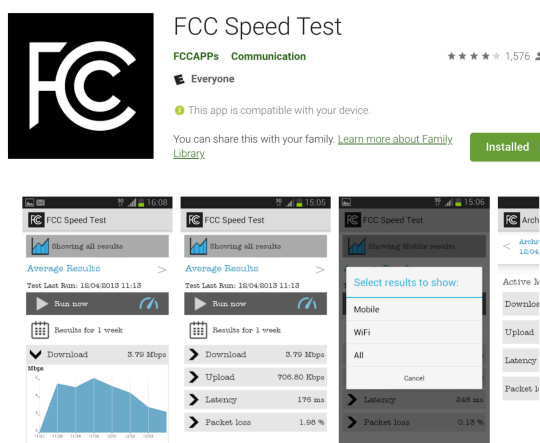


Figure 6: Screenshots of the FCC in the Google Play Store. (Captured July 11, 2018)

²⁹ For example, the mobile FCC speed test app discussed below is configured to limit test traffic to 100MB per month to avoid a user's monthly mobile data quota being consumed.

The FCC Speed Test app was built by the same organization, Samknows, that operates the FCC's fixed broadband measurement program. Arguably, choosing the same organization made sense since it might be seen as increasing the likelihood that many of the metrics and related-institutional issues associated with implementing a regulatory measurement program would be more easily resolved, and that the results would be more readily comparable to and potentially integrable with the FCCs' fixed broadband assessment efforts.³⁰ Given that it is unsurprising that the measurement methodology itself shares a lot in common with the measurements taken by the fixed broadband tests.³¹ Tests are conducted of throughput, loss and latency. As with the approach used in the fixed broadband measurement effort, the FCC mobile app is designed to conduct measurements to nearby servers, and thereby reduce the potential effect of network performance issues arising further upstream.³²

While this methodology provides repeated measurements from the same device and often in the same locations (work and home, etc), it does not itself guarantee nearly simultaneous measurements of multiple different networks in the same location like drive and walk testing. The lack of simultaneity in location/time of multiple operators had lead to criticisms about the inability to use this data to compare performance across operators, locations, and time.

4.4. Crowd sourced mobile measurements

The most popular example of this category is Ookla's Speedtest. We have been testing networks with Ookla's speedtest since the original web-based versions of their test. Client-side implementations of the Ookla test have expanded from Flash-based to Javascript-based web clients, Speedtest apps have been introduced for every major mobile device type, and finally native desktop implementations of the speedtest have been introduced. The fact that Ookla testing has been around for so long and so widely deployed means that there is a very rich source of test data available, which is a significant benefit of this data source. A problem with any of the testing approaches (whether drive test or crowd sourced) is the fact that many of the testers have only limited sets of observations (few locations and isolated in time) and many of those are not capable of being pooled or aggregated.

A key differentiating feature of the Ookla test approach is that the users initiate the actual tests, which further complicates the ability of this testing approach to provide comprehensive and consistent coverage area tests. This is in contrast to the approach taken by the FCC's speed test app that relied on regular automated testing. While the FCC approach may yield the testers' a greater degree of control over the test panel, the Ookla tests are conducted in locations that reflect the pattern of life of individual users and their interests and desires to understand mobile

³⁰ Policymakers have struggled with challenge of asymmetric regulation for mobile and fixed broadband services, especially in light of the convergence of these services at both the market and network levels. Having a common understanding of performance measurement would make it easier to move toward service/technology neutral regulations, which is a common aspiration for regulatory authorities.

³¹ The code itself is open source and available at <https://github.com/SamKnows/skandroid-fcc>.

³² That is, the FCC did not seek to conduct measurement tests of the end-to-end path to popular content and services, although for some users and questions, those results might be more relevant. Other mobile testing applications do provide such end-to-end test results.

performance in different locations. For some questions, this may render the Ookla approach a more appropriate sample of data (e.g., to evaluate the impact of performance on user behavior).

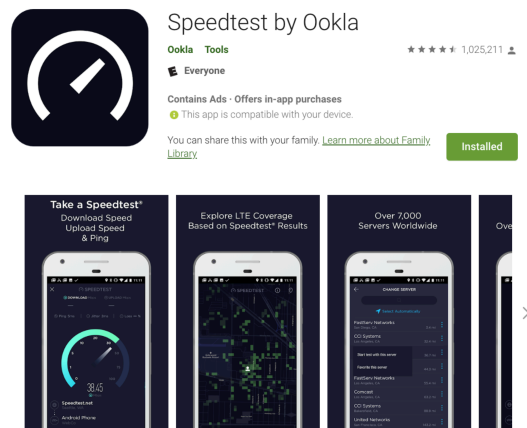


Figure 7: Ookla's Speedtest App in the Google Play store for Android.

5. Summing Up and Further Thoughts

The goal of this paper was to provide a check-point on the current status of mobile broadband testing efforts and to highlight the challenges confronting the measurement research community, which spans efforts in academia, private consultancies, network operators, and regulatory authorities.

When considering mobile broadband performance measurement, we confront all of the difficulties that have made fixed broadband performance measurement challenging, and that we have discussed in earlier papers. With mobile broadband performance measurement, the same challenges are often worse (test samples that fail to capture the entire population and may be limited with respect to time and location) and new challenges arise (fundamental relative differences between core methodologies). We have identified a number of these in this paper.

We believe the measurement community needs to work toward building a healthy measurement ecosystem and that will require accepting the existence of multiple measurement sources (for each methodology) and multiple methods (drive testing, crowd-sourcing and other approaches). Because the challenges of measuring the performance for a complex but important service such as mobile broadband raises so many real scientific and technical issues that are difficult even to understand even within the technical community, we need to continue to work to bridge multidisciplinary efforts. For example, previously the Internet performance and spectrum measurement communities have been largely separate, addressing different concerns and engaging different communities of engineers and researchers. Much of the debate about spectrum measurement has revolved around the occupancy of different frequencies and the ability to share spectrum in non-interfering/non-congesting ways. This has pitted advocates for unlicensed and licensed spectrum against each other with the former pointing to the fact that much of the RF is not occupied in many locations and much of the time (which is used to support arguments for allowing shared unlicensed access to those spectrum bands). In contrast, much of

the discussion over Internet performance has focused on the realized speeds of (fixed) broadband data services and how those compare across countries and with operators' advertised speeds. With mobile broadband, these issues are increasingly coming together as fixed and mobile services continue to converge in the looming world of 5G.

Key messages from this paper include that there is no single best measurement approach for all situations and that having multiple methods offers benefits in terms of enabling a wider-set of questions to be answered (some approaches shine light on questions that others leave unanswered), provide multiple perspectives on questions that both can inform (enabling cross-validation), and may enhance the emergence of greater trust in the measurement ecosystem (if users can select which performance measurement sources to consider).

We hope that this paper and the associated research will contribute to an active and better informed discussion among all industry participants about speed metrics. We believe competitive broadband markets will work better with more information and better data -- and better understood data -- about mobile broadband speeds.

6. References

Bauer, S., W. Lehr and M. Mou (2016), "Improving the measurement and analysis of Gigabit Broadband Networks," TPRC44, September 2016, Alexandria, VA, available at SSRN: <https://ssrn.com/abstract=2757050> or <http://dx.doi.org/10.2139/ssrn.2757050>

Bauer, Steve, William Lehr, and Shirley Hung (2015), "Gigabit Broadband, Interconnection Propositions, and the Challenge of Managing Expectations," TPRC2015, Alexandria, VA, September 2015, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2586805.

D.D. Clark, Bauer, Steven, Lehr, W., Claffy, K.C., Dhamdhere, A.D., Huffaker, B., Luckie, M. "Measurement and Analysis of Internet Interconnection and Congestion" Journal of Information Policy, Vol 4 (2014).

Bauer, Steven and Clark, David D. and Lehr, William, Understanding Broadband Speed Measurements (August 15, 2010). TPRC 2010. Available at SSRN: <http://ssrn.com/abstract=1988332>.

Bauer, Stephen, David Clark, and William Lehr (2009), "Broadband Micro Foundations: the Need for Traffic Data," published in Richard Taylor and Amit Schejter (eds), Beyond Broadband Access: Developing Data-based Information, Fordham University Press, 2013.