Improved Internet speed tests can enhance QoS and QoE

Al Morton

Network Technologies, AT&T Labs, Middletown, New Jersey, USA
acmorton@att.com

Abstract

Many Internet users spend time away from their typical activities to make measurements of their current performance, using one or more of today's widespread "Speed Test" web sites and apps. Their motivations to test are varied, often prompted by new service subscription or performance problems, and the objective results they measure have become an aspect of their perception of service quality and a growing part of their experience. There are two key issues impeding the progress of this growing crowd of measurement "experts": 1, the accuracy of the measurements and 2, how few systems help the users find and resolve the problems they perceive. Therefore, there is room for improvement beginning with the design of the measurements themselves and the network scope where measurements are most needed. For users with fixed broadband access and Wi-Fi networks, the main cause of impairments may be in their home.

Index Terms: performance metrics, capacity, throughput

1. Introduction

Since today's widespread "Speed Test" web sites and apps are often conducted by those who subscribe to a faster service or who are experiencing performance problems, the objective results they measure have become an aspect of their perception of quality of service (QoS) and a growing part of their experience (QoE). But the accuracy of the measurements is questionable, and only a tiny minority of the tool suppliers really helps the user find and resolve the problems they perceive. Therefore, there is space for improvement beginning with the accuracy of the measurements themselves and the network scope where measurements are most needed.

This paper is organized to briefly describe the consumer performance measurement landscape, and then expose some of the issues associated with various types of measurements. We then look at ways to improve the situation by addressing the two key problems of measurement accuracy and measurement scope. A new approach to bulk transfer capacity estimation is introduced with two different methods of measurement, so that users who measure obtain valid and useful results. After looking at some early measurements and recognizing the presence of performance targets in this work, we discuss the application of a classic manufacturing statistical test to the problem, the Sequential Probability Ratio Test, which allows us to control error in ways not possible before. Finally, we investigate the benefits of standardized large-scale measurement system deployment, where the scope of sub-path measurements (e.g., home network only, access network only) can easily be chosen to isolate performance issues when they exist, and thereby improve the customer's overall experience with the service. Recent studies indicate that home networks contribute significant packet impairments to user applications, and may therefore set an upper limit on their QoE.

No personally identifiable information was gathered or used in conducting this study.

2. Measurement tools

According to the Broadband Forum press release (April 17, 2013), there were over 643 million broadband subscribers around the world at the end of year 2012. Many make measurements and there are many tools they can choose from, or they can easily view the throughput their browser reports during a download. Every tool will produce results, but useful and accurate measurement can be an elusive goal.

2.1. Many tools, many users

Motivated users choose a particular testing method using information from a variety of sources. A Google search of "test my speed" yields page after page of test sites, and they indicate a lifetime of testing possibilities (158 million, as of April 2013). Advertising-supported sites and paid placements dominate, followed by a long list of network operator-sponsored sites. Each "crowd" has its say on support web-sites for games and streaming video services, recommending one or another test site.

Sometimes, helpful networking advice accompanies test site recommendations (e.g., Xbox embedded Wi-Fi connectivity problems are solved by using wired Ethernet instead). This is particularly true of Netalyzr, where results include relevant advice, and the authors have studied how people arrive at their site by examining the HTTP "Referer" headers [1]. The initial review of Netalyzr in German technology magazine "heise.de" and a follow-up article resulted in one third of visits (over 2 years beginning June 2009). One-fifth came from a link in the tech support area of a single online game (League of Legends). Less than 10% arrive from Google searches. Other sources include an IPv6 trial and technical blogs around the world, mostly causing a spike in usage that abates fairly quickly.

Another popular web-based measurement system is Speedtest.net, and their results for 6 months in late 2011 have been analyzed [2]. The authors speculate that one source of sample bias is self-selection: testing "when users believe that their network performance is poor or otherwise problematic". The Speedtest.net web page reinforces this position: "Are you getting the speed you pay for? Our Promise Index™ and Speedtest.net help keep your ISP honest!" They also claim over 4.2 billion measurements since they went on-line in 2006, that's more than 7 per broadband subscriber based on the 2012 BBF estimate above (but the Speedtest.net total includes mobile testers, BBF's population does not).

There are too many measurement systems to name here, but the SamKnows[3] whitebox must be mentioned. SamKnows has played a critical role in regulator characterization studies by providing the hardware, measurement design, and logistics to support multiple assessments in the US, UK, Singapore, Brazil, and an on-going European Commission campaign. Although subscribers volunteer to host a SamKnows whitebox, they are subsequently selected to balance the sample among ISPs, for

This paper was published in the Proceedings of the 4th International Workshop on Perceptual Quality of Systems (PQS 2013), held in Vienna, Austria, 2-4 September, 2013. http://pqs.ftw.at/

example. But at some point they must be sufficiently interested in their Broadband performance to volunteer.

2.2. Measurement system issues

The results of the current crop of on-line tests do not always correlate with actual QoS/QoE, or produce information that improves trouble resolution when a problem exists. First, a web-based speed test host is not typically co-located with the host containing content that the user intends to access, and therefore measures a path with some common sub-sections and some unique sub-sections. The test path includes a remote access link which may be limiting the measurement due to the link capacity and presence of other test traffic on that link or remote host. Second, measurement methods for capacity testing have long been a topic of research, and a wide range of techniques are deployed (as far as we know). Also, techniques such as dispersion measurement methods work passably in some circumstances but fail to provide an accurate assessment in others. Most techniques do not simultaneously evaluate all fundamental network properties, and often exhibit significant variation among measurements owing to use of short measurement intervals stemming from the (more or less correct) assumption that users desire results quickly.

2.3. Measurement challenges

Wireless home networks can be a significant source of performance degradation. Netalyzr [1] results indicate that WLANs had lower average download speeds and more degradation (loss ratio and round-trip time, RTT) than wired Ethernet LANs. The Home Network Profiler [4] authors reported 28% of tests with loss on the LAN, but cannot distinguish wired from wireless in their study. The authors of [2] conducted their own comparison of wired-wireless performance using a single instance of the SamKnows whitebox and the Speedtest.net measurements, finding a Download rate reduction of >40% with the 802.11g WLAN in all but one measurement. Whitebox and Speedtest results are in close agreement on the wired LAN.

Without much public hotspot data to point to (Netalyzr [1] reports “in public” average download performance and degradation (loss, delay) worse than total WLAN and Ethernet populations), we conducted a WLAN scan at a public coffee shop. There were 55 visible access points, 20 of them using Channel 1, the same as the coffee shop. Performance was about 1.3 Mbps up and down measured with Speedtest.net and Netalyzr, probably limited by the access link and not the WLAN.

Various WLAN stress tests have been conducted, and some compare a “best-in-class” home networking product with the performance of various Enterprise-class access points. One such study was conducted by an organization called Wireless LAN Professionals [5], where the Linksys EA4500 (their high-end product in early 2012) was tested along with 15 other industrial-strength products in a classroom arrangement. A single pc conducting a large FTP file transfer represents the best-case aggregate throughput for each product. Here the Linksys delivered the lowest throughput, but a value within a factor of 1.5 of many other models. Adding 5 iPad users in the classroom, streaming HD video, resulted in approximately 50% reduction in aggregate throughput for all access points (less than 50% for some, more than 50% for Linksys). This study used 20MHz channels, and 40MHz has clear capacity benefits but also brings the potential for interference due to crowding (fewer wide channels are available).

2.3.1. TCP-based measurements

We take as a given that when the capacity of the user-to-content path is stressed, TCP flow control is involved. When UDP would have traditionally sufficed, TCP is called into service instead to traverse the user’s firewall(s) and NAPT.

It is our understanding that Speedtest.net (2 or 4 TCP connections), Samknows (3 TCP connections), and others employ multiple simultaneous TCP connections to achieve a view of maximum transfer capacity with each measurement. However, the longest among download tasks usually involve a single connection between the user host and remote host to deliver a single stream.

Speed measurements using conventional TCP connections can be misleading because of many factors summarized by Mathis in [6], who claims, “TCP has zero predictive value because of its equilibrium behavior”. The TCP congestion-control may be one of several types (AIMD, BIC, CUBIC, CTCP to name a few, see [7] for a comparative summary of algorithm features and individual algorithm references), and the options available/used for Initial Window, Window Scaling, and Selective Acknowledgement may differ on the connection with the remote test host. Further, it is a challenge for independent instances of TCP flow control to pass the repeatability and continuity tests of a standard metric and method of measurement in RFC 2330 [8]. One reason is the classic Additive Increase Multiplicative Decrease (AIMD) TCP window control is non-linear by nature. The round-trip time (RTT) on the tested path is likely longer or shorter than the path between a user and their desired content source, and RTT directly affects the control loop.

Most active measurement systems appear to consider their measurements as the “ground truth”. Perhaps some calibrations are performed in development, but these are not published with the general descriptions. This appears to be a unique aspect of our measurement design effort, benchmarking the transfer capacity estimate against a known TCP transfer conducted in close time proximity. However, we expect considerable variability due to the factors above, and account for the variability in analysis.

2.4. Home network management

The task of managing a home network is significant, and has obvious implications on user QoS and QoE.

The evidence mounts that wireless LANs often degrade Internet access performance in terms of data transfer speeds (the title of [9] says it all, while early results in [10] indicate the WLAN may be the bottleneck when combined with broadband access speeds higher than 10 to 15 Mbps), and many devices (tablets and smartphones) have no Ethernet/wired option. These devices have small and somewhat limited antennas, relegating them at the low end of WLAN performers.

There are only three 40MHz channels in the 2.4 GHz band in the US, and uncoordinated band choices will usually result in overlap in neighbor WLANs, with likely degradation to transfer capacity and latency. The “new” 5 GHz band has only 4 non-overlapping 40MHz channels.

With small portable devices, users can usually move to find a better signal (which may be a higher signal to interference ratio, related to both desired signal power and the co-channel interference), but the simple 4-bar displays are not sufficiently accurate, and may not allow them to find the best location or distinguish between signal and co-channel interference.

In summary, if users are not currently aware of their Home WLAN performance and making measurements, it seems that many more will be in the near future. A wide range of WLAN “sniffers and stumblers” are freely-available, and the Home Network Profiler [4] provides a solution that combines a WLAN survey tool with Netalyzr [1]. The crowding and contention is the predictable outcome of mature radio system deployment using unlicensed frequencies.

3. Measurement methods

The goal of full-service and support of user experience is enhanced by having standardized and accurate measurement methods that help the user and network provider locate a performance issue when one exists. This “holy grail” of measurements and standards work now seems approachable with techniques described below.

3.1. Measurement algebra

The new Model-based Metrics devised by Mathis [11] offer potential ways to avoid the many methodological issues described above. These metrics assess fundamental properties of network performance, which are then used with any available model of the various generations of TCP in order to estimate throughput of a single connection. Further, the fundamental properties can be measured on sub-sections of the end-to-end path, and therefore support the identification of the section limiting the performance.

The Model-based metrics take the Delay-Bandwidth product equation as a starting point to calculate a hypothetical transmission “pipe_size”

$$\text{pipe\_size} = \frac{rate \ \times RTT}{MTU - overhead} \qquad (1)$$

The pipe_size is in units of packets, and MTU is the Maximum Transmission Unit at the IP layer, often 1500 octets or less. Next, Mathis applies the Macroscopic model of TCP [12] to calculate the minimum packet run length between losses needed to support a specific target rate with target round-trip time (RTT) on the measured path.

$$\text{ref\_run\_length} = \left(\frac{3}{2}\right) \times \text{pipe\_size}^2 \qquad (2)$$

We can set target values for rate, RTT, and MTU, and calculate the target loss-free run_length in packets (losses can come no more frequently than 1 in run_length packets, or the multiplicative TCP window decrease will cause the pipe to be under-utilized [11]).

It is believed that to a reasonable approximation, all models of TCP have the same input parameters (primarily packet loss probability, but also RTT and MTU), thus the equilibrium throughput of any of the TCP congestion control algorithms should be predictable.

3.2. Direct loss measurement methods

Two categories of measurement methods use the Delay-Bandwidth product equation and macroscopic TCP behavior algebra described above. The first is a direct measurement of the loss–free run length in MTU-sized packets under sending conditions related to the target RTT. The TCP sender does not use the conventional window flow control. Instead, the sender operates “open-loop” and the receiver acknowledges as many packets (segments of bytes) as it receives them (possibly using the selective ACK, or SACK option, or the sender function must include enough of a control loop so that packet retransmissions are sent when needed).

[11] currently describes a large number of candidate methods in this category. Vetting this list and recommending a best method for different measurement circumstances is future work. It is intended that these methods be applied when Active Queue Management (AQM) is deployed on the tested path, so that there is deliberate notification from the network when congestion is encountered. Lost (discarded) or Explicit Congestion Notification (ECN) marked packets are treated equally, as “defects” in the sample.

3.3. Direct delay-bandwidth product method

The second method involves direct assessment of the Delay-Bandwidth product by measuring the arrival rate of a flight of packets (determined from the target rate and target RTT), and a statistic summarizing overall RTT experienced by the packets during transfer. The implied loss-free run length can be estimated from measured rate, RTT, and the limiting MTU, again using the Macroscopic model of TCP [11]:

$$\text{est\_pipe\_size} = \frac{\text{meas\_rate} \ \times \text{meas\_RTT}}{MTU - overhead} \qquad (3)$$

$$\text{implied\_run\_length} = \left(\frac{3}{2}\right) \times \text{est\_pipe\_size}^2 \qquad (4)$$

$$\text{est\_prob(loss)} = p = \frac{1}{\text{implied\_run\_length}} \qquad (5)$$

We compared the estimates using rate and RTT measurements with a real benchmark, a FTP/TCP throughput measurement conducted in a closely adjacent time over several days, with both the estimate and FTP/TCP transfer conducted periodically throughout each day. The daily averages compare very closely, as shown below, where B(p) is the throughput estimate using the TCP Reno model in [13], with (the empirically estimated probability of loss) p, meas_RTT, and MTU as inputs.

Table 1. *Comparison of throughput measurement (Mbps) and delay-bw method estimate using model-based metrics.*

Day	Ave FTP	Ave B(p)	Ratio,%	delta,dB	%<FTP
Tues	0.830	0.807	97.2%	-0.12	2.8%
Wed	0.831	0.815	98.1%	-0.08	1.9%
Thurs	0.839	0.825	98.4%	-0.07	1.6%
Fri	0.838	0.828	98.8%	-0.05	1.2%
Sat	0.905	0.899	99.3%	-0.03	0.7%

In this experiment, we followed a sending discipline like that used in Netalyzr [14], but simplified. Netalyzer assesses capacity with UDP packets sent in a pattern like TCP follows during Slow-Start with an exponentially increasing burst size (send 1, when “ack’ed” send 2, when “ack’ed” send 4, etc.). Netalyzr uses asymmetrical packet sizes: large packets in the direction being characterized and small packets in the other, again like a TCP octet stream being answered by 40 octet ACKs.

Although the method described above produces actual capacity estimates rather than a comparison with a target for loss-free run length (as described in the measurement algebra), it can be used to assess compliance with a target value of RTT since RTT is measured on every successful round-trip packet transfer. When the measured rate is below the corresponding target rate value, the measured RTT may also be higher than the RTT target for some packets, and these packets can be designated as “defects” (similar to lost or ECN marked packets). We note that this method appears to be complimentary to the direct loss measurement (described above) when there is no AQM deployed on the measured path, and that RTT defects will not be reliable with AQM deployed.

4. Statistical analysis

We also introduce a new application of traditional statistical methods, now applicable because Model-based Metrics evaluate whether specific packet transfer rates (throughput) are supported on the tested path or not. Sequential Probability Ratio Testing (SPRT) allows us to examine the empirical packet defect ratio as the tests are in-progress, where a defect could be a loss, ECN indication, excessive RTT, or other defect as defined. Based on the target defect ratio, a second defect ratio is used as the failure threshold. Choosing Type I and Type II error probabilities, the tester can determine when the results support the target ratio with desired confidence (or conversely, when to stop testing and declare failure). This is a non-parametric test, and robust to unexpected changes in the underlying probability distribution (as sometimes happen in manufacturing, or when network conditions change radically due to weather or failures).

Below, we provide some details on SPRT as it pertains to the sample size needed to make a decision and terminate the session. These factors are critical for any efficient “speed test” and they are not discussed by other system authors to our knowledge.

[15] provides an accessible description of SPRT calculations, originally addressed by Wald [16]. We have a target_defect_probability, 1 defect per target_run_length, where a "defect" is defined as a lost packet, ECN marked packet, or other impairment. This constitutes the null Hypothesis:

H0: no more than one defect in target_run_length = p_0

We can stop sending flights of packets (of size equal to the “pipe”) if measurements support accepting H0 with the specified Type I error = alpha (= 0.05 for example). We choose the alternative Hypothesis “failure” probability at four times the target:

H1: one or more defects in target_run_length/4= p_1

We can stop sending flights of packets if measurements support rejecting H0 with the specified Type II error = beta, thus preferring the alternate H1.

As flights are sent and measurements collected, the tester evaluates the cumulative defect count against two boundaries of acceptance and rejection:

$$X_A = -h_1 + sn \quad \text{(acceptance line)} \tag{6}$$

$$X_R = h_2 + sn \quad \text{(rejection line)} \tag{7}$$

where n increases linearly over all flights of packets and

$$h_1 = \left(\log\frac{1-\text{alpha}}{\text{beta}}\right)k^{-1} \tag{8}$$

$$h_2 = \left(\log\frac{1-\text{beta}}{\text{alpha}}\right)k^{-1} \tag{9}$$

$$k = \log\frac{p_1(1-p_0)}{p_0(1-p_1)} \tag{10}$$

$$s = \left(\log\frac{(1-p_0)}{(1-p_1)}\right)k^{-1} \tag{11}$$

for p_0 and p_1 as defined in the null and alternative Hypotheses, above.

The calculations above are implemented in the R-tool for Statistical Analysis [17], in the add-on package for Cross-Validation via Sequential Testing (CVST) [18].

Armed with equations above, we can calculate the minimum number of packets needed to accept H0 when x defects are observed, for example x=0.

$$X_A = 0 = -h_1 + sn \tag{12}$$

$$n = \frac{h_1}{s} \tag{13}$$

For example, with values of p_0=0.00167, p_1 = 4p_0, and alpha= beta = 0.05 corresponding to a target rate of 4.67Mbps at 50ms RTT, pipe = 20 packets and (MTU-overhead) = 1460 octets, then to accept H0, a minimum of n = 586 packets need to be sent successfully while there is no defect observed to support the targets. Figure 1 illustrates these values with an example.

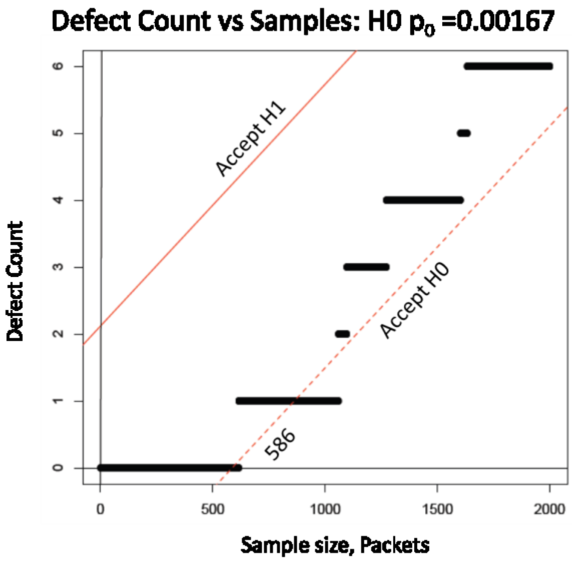

Figure 1: *Graph of Sequential Probability Ratio Testing – As each packet arrives, the defect count is evaluated against acceptance thresholds for H0 and H1, and in this case H0 is accepted after 586 packets, the minimum. The defect counts are simulated using a binomial distribution with target probability =* p_0 *,thus the defect count remains near the H0 acceptance threshold as the sample size increases. The R-tool [17] and CVST add-on package [18] produced this graph.*

5. Large-scale measurement of broadband

As mentioned earlier, the end-to-end scope of many measurements today is a partial cause of inaccuracy and variability, making path segments with degraded performance difficult for the user to locate. The Large-scale Measurement of Broadband Performance (LMAP) project in the IETF intends to help address the need to measure segments of the end-to-end path in isolation.

Although the project is in the formative stages at the time of writing, LMAP intends to provide measurement control and results collection protocols that would support measurement of IETF standard performance metrics between a host on a home network and their Residential Gateway, for example. Preliminary results from the BISMark project [10] and other home measurement efforts indicate that home networks may be the cause of poor performance in many cases. The capability to test key segments of the user's path should improve trouble isolation and with it, the future quality of service and quality of experience.

The proposed reference path and measurement points are work-in-progress [19], but they give a view to the scale and penetration of the standardized measurement infrastructure.

Figure 2 shows the reference path with designated measurement points.

In order to isolate performance degradation to one or more segments of the end-to-end path, LMAP measurement agents would be located at various measurement points, designated mpNNN. Measurement points under a single organization's control are indicated with the same hundreds digit (mp1NN). An ISP could qualify its access infrastructure by performing measurements between mp100 and mp190. Another key example is where a service Subscriber could measure their private network performance by conducting tests between a designated host (mp000) and the Access Demarcation point (mp100).

Segmented measurements of the end-to-end path can be combined to estimate the performance of the complete path, but only under well-planned circumstances. Fortunately, this topic has already been examined and the results standardized in the IETF IPPM working group. [20] provides a general framework for composition and aggregation of measurements, defining "spatial composition" as the form most applicable here. In [21], we defined the deterministic functions that yield the complete path metrics using metrics of the sub-paths for loss, delay, and delay variation. It remains to be shown that the Model-based metrics [11] can be combined to assert complete path support for a target rate and RTT, but the design using fundamental metrics that are already addressed in [21] bodes well. Including the subscriber in the measurement framework, especially when they suspect or experience degradation, is a keystone of credibility and success for LMAP.

6. Discussion

To be successful and contribute to improved QoS and QoE for users, the many efforts described above will need to progress steadily and achieve their stated goals. Good performance metric definitions without a supporting infrastructure leave users on the path to unsatisfactory performance through uncoordinated use of WLAN equipment and "SSID-naming wars" (to communicate indirectly with neighbors). WLAN QoS mechanisms may offer some relief, but they require some expertise on the part of the home network administrator.

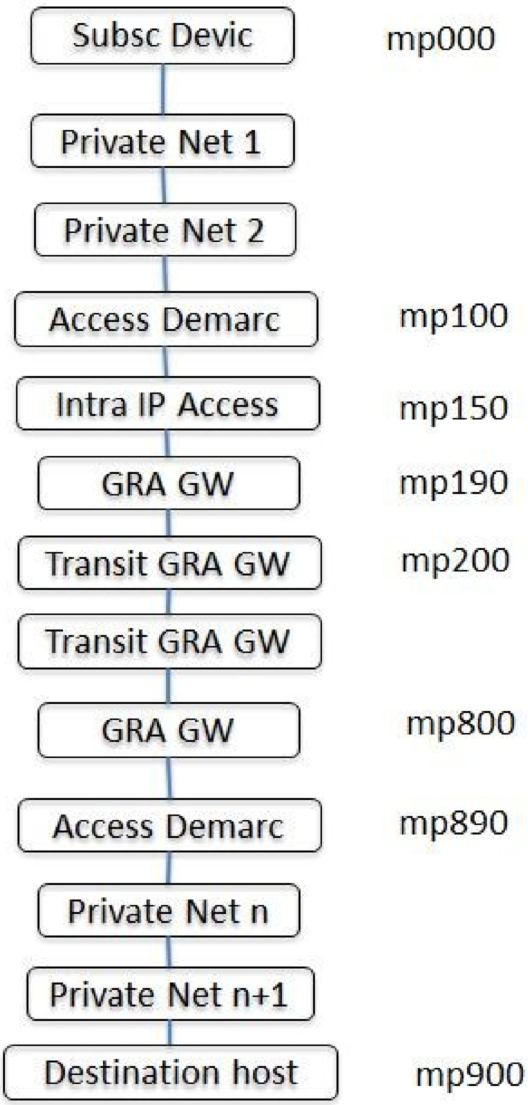

Figure 2: *LMAP reference path and measurement points (GRA GW = Gateway with a globally routable address).*

7. Conclusions

From the rapid growth of user measurement tools for Internet performance, we infer that user measurement activity is a key contributor to QoS and QoE for many. The currently available tools have limitations, due to measuring the limiting sections of the path from distant locations, thus making conclusions more suspect than if they were measured ideally – section by section. At the same time, evidence mounts that the home WLAN can contribute significant performance degradation because of co-channel interference on the unlicensed frequency bands. The mature deployment of Wi-Fi and broadband Internet access, with user applications that intend to use radio resources on a nearly continuous basis, as well as the widespread adoption of tablets and smart phones with no wired connectivity option, paint a crowded picture for the future. The first step toward satisfaction after a disappointing experience is always to isolate the problem. We have described new Model-based metrics that lend themselves to this task and introduced a standardized measurement framework to support the users at their current large scale. We have briefly introduced two measurement methods for the

Model-based metrics and show that they each have promise in different circumstances: categorized by network features for congestion control. With the measurement approach that validates targets for transmission rate or RTT, we can apply a classic statistical test and understand when we have performed sufficient measurements as well as the error associated with our outcome. We briefly summarized how the IETF LMAP project can support the needed measurements at the extreme network edge.

8. Acknowledgements

Len Ciavattone always contributes his development and measurement expertise when called on, thanks Len. Matt Mathis has patiently explained some networking nuances as part of our collaboration on the Model-based metrics, and it’s a privilege to work with him in IETF. In an extremely efficient manner, Ganga Magulri suggested the SPRT statistical test, a perfect match to this problem. We all say thanks to Ganga.

9. References

[1] Kreibich, C., Weaver, N., Maier, G., Nechaev, B., and Paxson, V., “Experiences from Netalyzr with Engaging Users in End-System Measurement”, ACM SIGCOMM Workshop on Measurements Up the Stack (W-MUST), 2011, Toronto, Canada.

[2] Canadi, I., Barford, P., and Sommers, J., “Revisiting Broadband Performance”, Internet Measurements Conference IMC ’12, Nov 14-16, Boston, MA.

[3] Crawford, S., et. al, “SamKnows: Global Leaders in Broadband Measurement”, http://www.samknows.com/broadband/index.php

[4] DiCioccio, L., Teixeira, R., May, M., and Kreibich, C., “Probe and Pray: Using UPnP for Home Network Measurements”, Passive and Active Measurement Conference (PAM), 2012, Vienna, Austria.

[5] Wireless LAN Professionals, “Wi-Fi Stress Test: A Vendor-Independent Access Point Analysis”, Feb 5, 2013, http://www.wlanpros.com/wp-content/uploads/2013/02/Wi-Fi%20Stress%20Test%20Report.pdf

[6] Mathis, M., “Why BTC is so hard”, presented at IETF IPPM working group session, March 14, 2013. http://www.ietf.org/proceedings/86/slides/slides-86-ippm-7.pdf

[7] Yang, P., Luo, W., Xu, L., Deogun, J., and Lu Y., “TCP Congestion Avoidance Algorithm Identification”, ICDCS '11, Proceedings of the 2011 31st International Conference on Distributed Computing Systems, Pages 310-321.

[8] Paxson, V., Almes, G., Mahdavi, J., and Mathis, M., “Framework for IP Performance Metrics”, IETF RFC 2330, May 1988, (section 6.2).

[9] Tom’s Hardware, “Why your W-Fi s***s and how it can be helped, Part 1”, http://www.tomshardware.com/picturestory/571-wi-fi-beamforming-networking.html

[10] Sundareson, S., “State of Project BISmark”, presented at ISMA 2013 AIMS-5 - Workshop on Active Internet Measurements, February 6-8, 2013, slide 10. http://www.caida.org/workshops/isma/1302/

[11] Mathis, M. and Morton, A., “Model-Based Internet Performance Metrics”, IETF Work-in-progress, Feb 25, 2013. http://tools.ietf.org/html/draft-mathis-ippm-model-based-metrics-01.txt

[12] Mathis, M., Semke, J., Mahdavi, J., and Ott, T., "The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm", Computer Communications Review, volume 27, number 3, July 1997.

[13] Padhye, J., Firoiu, V., Towsley, D. and Kurose, J., “Modelling TCP Throughput: a Simple Model and its Empirical Validation”, SIGCOMM 1998.

[14] Kreibich, C., Weaver, N., Nechaev, B., and Paxson, V., “Netalyzr: Illuminating The Edge Network”, IMC’10, November 1–3, 2010, Melbourne, Australia.

[15] Montgomery, D., "Introduction to Statistical Quality Control - 2nd ed.", ISBN 0-471-51988-X, 1990.

[16] Wald, A., “Sequential Analysis”, Wiley, 1947.

[17] R Development Core Team, "R: A language and environment for statistical computing”, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/ 2013.

[18] Krueger, T. and M. Braun, "R package: Fast Cross- Validation via Sequential Testing", version 0.1, Nov 2012.

[19] Bagnulo, M., Burbridge, T., Crawford, S., Eardley, P., and Morton, A., “A Reference Path and Measurement Points for LMAP”, IETF Work-in-progress, Feb 24, 2013, http://tools.ietf.org/id/draft-morton-ippm-lmap-path-01.txt

[20] Morton, A. and Van den Berghe, S., "Framework for Metric Composition", RFC 5835, April 2010.

[21] Morton, A. and Stephan, E., "Spatial Composition of Metrics", RFC 6049, January 2011.