Summary Review Documentation for

"Revisiting Broadband Performance"

Authors: I. Canadi, P. Barford, J. Sommers

Reviewer #1

Summary: This paper reports the analysis of the broadband performance using data from Speedtest.net. The data analyzed consists of 54M tests collected from 59 metropolitan areas over 6 months. The results confirms the findings from prior broadband performance studies. The results also includes new findings about clusters for low-performing clients and correlation analysis on distance and latency. Finally, the paper compares the US results against other 35 metropolitan areas worldwide.

Strengths: The main strength of the paper comes from the data sets from a popular web-based Speedtest service covering much broader users than previous studies. The results shows that data from this service is mostly consistent with more accurate but smaller-scale methods, which opens up the possibility to use these crowd-sourcing measurements with a higher confidence.

Weaknesses: Most of the results are on confirming the previous reports, and naturally they are not surprising.

The overall analyses are cursory, only with summary statistics and scatter plots of upload/download performance. The authors raised several seemingly interesting questions from their observations, but these questions were not explored much further in the paper.

The authors state that the advantage of this data set is the size, diversity, and details, compared to the data sets used in previous studies. However, the analyses do not make full use of these advantages.

The Speedtest has data since 2006, but a longitudinal study is not performed in this paper.

Comments to authors: The main contribution of this paper is to confirm the previous findings especially by Sundaresan et al. using much broader data sets from Speedtest.net covering much larger user base. The results indicate that we can use data from these services to quantify broadband performance with a higher confidence.

The issues with WiFi are identified in the experiment in Sec 4.2, but no effort was made to reduce the bias caused by the use of WiFi at home. It seems to me that it is possible to filter low-quality results to some extent by observing variations in a test result.

The analyses use only simple summary statistics (mean and standard deviation), but they are not enough to capture the behaviors of diverse broadband users. The scatter plots indicates the distributions of upload and download performance are multimodal (with different service plans, the use of WiFi, etc) so that I'd like to see deeper analysis on distributions.

What are implications from the new findings presented in Sec 4.3? The authors speculate a few possible causes for the low-performing clusters, but further analysis is left for future research. They also observed non-linear relationship between distance and latency for some ISPs but it was not investigated further.

For the international comparison, it is interesting to see the differences in the upload/download scatter plots (Fig. 20-22) among different countries. I'd like to see further analysis on these high-performing fiber users.

You should probably add Korea that has the highest broadband penetration in the world.

Reviewer #2

Summary: This paper presents a large-scale study of broadband Internet performance, as observed from the Ookla's Speedtest tool. In contrast to previous studies (e.g., SamKnows), the study has a much larger set of vantage points, since the measurements can be run from end hosts, rather than requiring a router deployment. The paper confirms many of the findings from the SamKnows study and also presents some new findings about broadband performance in markets outside of the United States.

Strengths: The paper offers a nice re-appraisal and independent confirmation of the findings from a previous study. To me, this is exactly the kind of paper that belongs at IMC: a thorough confirmation of previous findings using an independent dataset. The findings are interesting, and the paper adds some new findings, such as persistent low performance and the effect of distance on latency. The paper also offers some findings on the performance of various fiber deployments, which were not included in previous work.

Weaknesses: The paper doesn't offer any new insights. The paper makes much ado about "crowdsourcing" measurements to measure access link performance, but this is not the first study to have done so. In particular, Netalyzer and Dasu effectively take the same type of approach, and both also characterize broadband performance from the end host.

Some of the methods used to "validate" the Speedtest measurements against the SamKnows tests don't seem particularly valid. The sample sizes for these tests are too small, and these two tests shouldn't be expected to produce the same set of results anyway.

Comments to authors: Although the paper offers confirmation of previous results and adds a few of it's own, I must admit that this is not the most exciting paper in the world: it doesn't really offer any new insights or discoveries. It evaluates the question of broadband access performance, as previous work has done, and finds largely some of the same conclusions. For an IMC paper, I

think this is certainly an acceptable paper, but I would have appreciated much more insights, trends, and so forth, as opposed to simply "mere data reporting".

Many of the new findings are unsurprising, such as the observation that latency increases with distance. In other cases, the scatter plots appear somewhat inconclusive. Figures 13, 17, and 20 don't really seem to show any meaningful trends.

Overall, I would have appreciated a slightly better treatment of the issue that the Speedtest measurements are being performed from a host, rather than the router itself. The paper makes some effort to compare the two sets of measurements, but there do appear to be fundamental shortcomings of host-based measurements (wifi interference, cross traffic, load on the host, etc.). The fact that the results generally match the router-level measurements would indicate that these factors are not an issue, but it would be nice to have a more careful treatment of this issue somehow, perhaps with some more controlled experiments. This, alone, could make for a great paper.

Reviewer #3

Summary: This paper provides an analysis of 54M bandwidth tests conducted by people in 59 metropolitan locations (24 US, 35 outside) around the world via speedtest.net between June 1 and November 30, 2011. The authors explain speedtest.net's test approach, and present high-level evaluation method (e.g., only measurements within a 200-mile radius to the server).

Next is a comparison between SamKnows and speedtest.net: SamKnows, due to its next-to-modem position, consistently performs more accurately than the potentially WiFi-affected Speedtest.net. Concurrent transfers hurt both.

Then they compare per-ISP findings to the Sundaresan paper and can mostly confirm their findings. They investigate cases of suspiciously poor performance for some customers, and suspect exceeded quotas and old equipment. Distance to the test server matters, but sometimes non-linearities exist.

Finally, they switch focus to international areas and compare to the Netalyzr results, finding they are nearly twice as high, and study fiber deployments.

Strengths: Fantastic dataset, nice analysis and (mostly) reaffirmation of previous findings. I learned a few things. Good stuff.

Weaknesses: It would have been *great* to see a more longitudinal analysis of the Speedtest.net data, given that they've been around for 6 years. More analyses would help, as the paper feels a bit padded to make it up to full size.

Comments to authors: I feel your intro overdoes it a bit. There's no doubt the Speedtest.net dataset is fantastic. Do you really need more than that, such as "provide a crowd-sourced perspective on broadband measurement"? I don't see any novelty in that regard.

What motivation does Speedtest.net have for offering TCP-based tests and no UDP-driven ones? The usual argument is that these produce "what the user experiences", and it would be great to know their thinking. Seeing the differences at this scale of a dataset would be terrific.

Section 2.3: You only have half a year's worth of data. You *could* do a longitudinal study if you actually had the years of data that speedtest.net has gathered. You don't, so you might as well save the space.

Your comparison to Netalyzr's results in 4.4 is interesting – would be great to know how much difference the passing of two years have made.

The paper feels a bit superficial and figure-heavy. It would be nice if you could balance the latter with deeper analysis of the more interesting findings.

Reviewer #4

Summary: This paper presents an analysis of broadband performance using speedtest.net and SamKnows tool. The paper has two major contributions: first, it reconfirms some of the prior results in broadband performance, and second it compares the broadband performance around the world.

The paper is very well written and was a pleasure to read. Even though the paper does not have any particularly novel insights, its thoroughness in analysis and its well written style make it worth the reading.

Strengths:

- Looks at a pertinent problem of broadband penetration and its statistics in US and across the world.
- Very well written with great clarity in results and analysis. Enjoyed reading the paper.

Weaknesses: The paper does not have any particularly novel insights beyond what's already been published in parts and pieces before.

Comments to authors: The paper would be even better if you delved deeper into some of the questions such as what exactly is causing the clusters of consistently low performing broadband services in Fig. 11 and 12.

Reviewer #5

Summary: Reappraisal of previous findings on broadband performance data/analysis, using 6-month window of speedtest data. Unfortunately they do not add much to previous findings.

Strengths: Uses large set of Internet measurement data (although the methodology of this data has been questioned, including by these authors in this paper) to study a timely and relevant question.

Weaknesses: For a paper that is reviewing previously analyzed data and comparing it to other related studies in the field, the related work section is pretty weak. It seems to me the authors could have undertaken deeper comparisons among the results of different studies presented in Section 5. e.g. there is a lack of insightful comparisons between the Speedtest data and other data sets, not just the Sundaresen but also the Kreibich.

There is a tendency to substitute figures for insight and explanations. I rather drowned in figures half-way through the paper. It added to the overall sense that this paper was more about reporting data/numbers than providing insight/understanding into network behavior/evolution.

Comments to authors: "Nevertheless, many of our conclusions described below are consistent with prior work, suggesting that any biases are fairly limited in nature." – which prior work do you mean, and how do you know it doesn't suggest that the works are merely biased in the same way?

Figure 4 could use some traffic flow to show how the tests are working, in particular how they work differently. Figure 9 – can you compare the data in this figure with the data from the SamKnows study (don't need IP addresses for that, right?)

'We hypothesize that these artifacts represent "cheap and slow" service plan offerings. While operators may no longer offer these low performance plans, it is clear that some of the customers have not yet switched to new service plans" – isn't this something you can verify using one of these cheap and slow plans and running a test from one of them, or you literally cannot find a subscriber?

Response from the Authors

We thank the reviewers for their comments and suggestions. In response to the reviews, we have significantly enhanced the figures in the paper and addressed several concerns in the text. Specifically, the figures and font sizes have all been increased and Figures 14, 15, and 23 are re-formatted as bar plots. We have also added additional text regarding other measurement platforms, comparing host-based and router-based measurements, and have highlighted implications of the new findings in our study.

To address the concern of the reviewer who felt that we should expand our commentary on related studies, we have expanded the related work section in several ways. First, we added text that compares our findings with prior work by Dischinger et al. In particular we note that the differences between upload and download speeds have narrowed. Next, we restate the differences in measurement data collected from SamKnows vs. Speedtest, highlighting the gateway-based deployment versus client-based deployment. Finally, we significantly expand the explanation of the study by Kreibich et al., which employs the Netalyzer tool. While there are indeed similarities between Netalyzer and Speedtest, they key differences are now highlighted, including the fact that measured upload/download bandwidths have increased between the two studies.

A common criticism by the reviewers was that the results in our paper are not "surprising", or that the conclusions we draw are well known. First, it's not true that all the conclusions in our paper are known (see, for example, Section 4.3). Second, our data are significantly different than other data that have been used in prior work. As a result, our work adds a broader and important perspective on broadband performance. Moreover, several results in our work support earlier claims, despite the fact that our data and results are derived from a completely different measurement method. So, even though our work replicates prior results, the *quality* of replication is much more powerful than if we had used a similar measurement method or data. Although science "reserves its highest honours for those who do things first" (Harry Collins, "Changing Order"), a critical aspect of a scientific finding is that it should be able to be replicated. We think that the network measurement community should be doing *more* and more powerful types replication studies to ensure that our results meet this critical bar of being scientific.