Evaluating net-score as a measurement platform for broadband access performance

By
James Martin

Senior Honors Thesis
Computer Science
University of North Carolina at Chapel Hill

March 23rd 2015

Abstract

Net-score is a new platform for measuring broadband access connections for Internet users. In this paper, we outline the architecture of the net-score measurement platform and discuss its advantages over current state-of-the-art. We also describe how we evaluated net-score through a seven month long case study, describe the data we collected, and show how net-score reaches its goal of being a low-cost and low barrier to entry method of measuring broadband performance.

I. Introduction

Importance of measurement platforms

Measurement platforms for network performance are useful for three reasons. The first is that they are important for diagnosing problems with access connections. Second, measurement platforms encourage transparency so customers can ensure they are provided equal service for equal price. Third, measurement platforms can gather data useful for researchers. By recording test data, researchers can gain insight into the state of current network infrastructure.

Measuring the performance of broadband access connections is of interest to several groups. Consumers, content providers, and regulators are invested in transparent and accurate reports of network performance.

Consumers are increasingly using broadband Internet connections to replace dedicated services like telephone and cable television (Palazzo). New services such as Sling TV, which provide streaming select TV channels, as well as the meteoric rise of video streaming, indicate this consumer movement. As higher volumes of content become widely available via the Internet, users need to be aware of the quality of their access connections (Dobrian).

Content providers want their customers to be better connected to enable pushing of higher definition content. Higher definition and higher quality content through better connections retains customers and increases profits for providers. Slight increases in page load delays (100ms – 1s) can cause a noticeable drop in web searches and can result in revenue losses (Schurman and Brutlag). Consequently, content providers certainly hold a stake in promoting transparency for access connection performance.

Regulators and governing bodies also share a stake in making measuring broadband access connections more transparent and more easily understood for non-experts. A well-connected populace is vital to competing in an increasingly globalized economy. Access to online information and education allows citizens to train for new careers, purchase products from across the globe, and participate in important conversations regarding prominent social issues. By promoting transparency, regulators ensure ISPs are held accountable for the claims they market to consumers and can target regions lacking infrastructure for improvements and development.

Qualities of an effective measurement platform

An effective measurement platform will have a low barrier to entry for clients and will be scalable at low cost. To have a low barrier to entry, a measurement platform must not require additional hardware or software and must not take significant steps to initiate and complete. Navigating to a specific website, having to select a server to run a test to, or having to install a flash plugin are all examples of design decisions that increase barrier to entry. With a low barrier to entry, an effective measurement platform will be able to reach users in all corners of the

Internet. Increasing the audience of the measurement platform through a low barrier to entry will allow more insights to be made from sampled data.

Measurement platforms that are scalable at a low cost will be able to easily capture new populations. Being able to reach new audiences without incurring additional costs of maintaining the servers that host the platform will ensure continued growth and inclusiveness.

Lastly, a measurement platform should accurately measure and reflect the behaviors its test takers exhibit. Since we wish to capture the performance of Internet users, we attempt to emulate and measure the performance of network actions taken while browsing the web. Internet users also only have a choice of which ISP carries their traffic to the backbone of the web. A measurement platform measuring web performance should reflect its user's local network infrastructure, or last mile latencies. While limiting factors may exist in the backbone, last mile latency and local infrastructure play major roles in the speeds typical Internet users experience.

Our contribution

Considering these motivations, net-score is designed to be an easily scaled measurement platform that is a low cost alternative with a lower barrier to entry than the current state of the art. By using the browser as a platform, our tool can run HTTP based tests successfully on all browsers and operating systems, as discussed in later sections. Currently, net-score is deployed on several small blogs, including the personal websites of Eric Gavaletz (http://www.net-score.org/) and James Martin (http://cs.unc.edu/~jamesml). Visiting the net-score website will run a full "flock" test, which is a more expansive test than evaluated in this paper. However, the test embedded in James Martin's webpage is a "feather test," which is described in detail and evaluated in this paper. Through a seven-month long deployment in the field, we observed almost 200,000 tests taken by at least 7,000 unique user agent strings. These tests were taken all across the globe with 150 countries represented in our sample. The website and audience we reached is described in greater detail in later sections.

Our vision for net-score is to eliminate the necessity to drive new users to our platform, but rather actively include new populations by embedding the tool on popular third party websites.

II. Current State of the Art

Ookla Speedtests

The current most ubiquitous tool for measuring consumer broadband performance is the Speedtest, which at the time of writing has over seven billion tests. Taking a Speedtest has become a standard of network troubleshooting, and is employed by many ISP call centers. The user is presented with a graphical interface, including a button to begin the test and sometimes a map to select a test server. The test results display the user's latency, measured in milliseconds, and download/upload throughput values, measured in Megabits per second. Speedtests involve contacting a server, which is either selected based on geographic proximity or by the user, and a series of images are downloaded using HTTP requests. In addition to the dedicated Ookla Speedtest servers located around the globe, ISPs such as Time Warner Cable and Comcast also set up dedicated Speedtest servers to help users diagnose and troubleshoot problematic network connections. While Speedtests performed in the browser are similar to net-score tests, reporting average latency and average throughput between a specified dedicated test server is biased in three ways.

1. Speedtest's userbase may be weighted towards technically sophisticated, or "geeky," users who use the tool when they experience unsatisfactory network performance. These users may not be representative of the typical Internet user.

2. Since the tool requires users to navigate to Speedtest.net and install a flash plugin for their browsers, Ookla may be excluding "non-geeky" populations from their data.

3. The software that Ookla deploys on its test servers is closed-source and most of its development is funded by ISP licenses. ISPs also host Speedtests on their own networks, and direct customers to taking such tests when diagnosing problems with their connection. Here, a conflict of interest arises.

Speedtests, in summary, are a popular but blackbox method of measuring network performance. Another valuable tool for evaluating network performance are broadband reports.

Broadband reports

Published by large content providers such as YouTube and Netflix, broadband reports to a good job of providing contextual performance measurements for a particular application or service. For example, YouTube and Netflix speed indices display throughput in Megabits per second and provide statistics that are tabulated by service provider, region, or both. A major drawback to these reports is that they are only relevant in the context of their respective media players or applications. Since Youtube and Netflix are examples of service providers with large infrastructure, metrics are also only relevant for downloads from their servers. Despite these shortcomings, broadband reports provide key comparisons of service providers by controlling variables such as application type and user behavior. Since they report performance of the user behavior of the same application, disparities between regions or service providers may be more easily visible.

Measurement Lab

The work done by Measurement Lab (M-Lab) to progress network performance research is a step in the right direction. By creating accessible data repositories and hardware platform for research tools, M-Lab allows researchers to deploy their tools on well-provisioned infrastructure and gain access to a large user base of researchers and participating firms in the industry. Current tools deployed on M-Lab's infrastructure include:

1. NDT: a test providing detailed packet level information and kernel-level statistics on how a TCP connection performed in a given path.

2. NPAD: a tool using TCP to measure end-to-end throughput and information about the switch and router queues along a path.

3. Glasnost: a tool emulating a BitTorrent client to measure performance back to a server, detecting if application flows are being limited or disrupted by an ISP.

4. Pathload2: a tool using UDP packets to measure "available bandwidth" of an end-to-end path between a user's client and an M-Lab server. Pathload2 also stores its collected data in M-Lab's archive.

M-Lab also enables access to a wealth of network measurement data, "removing the need to for every research project to collect its own data and facilitating cross-sample analyses that would not otherwise be possible" (Dovrolis et al.). While M-Lab's growth and vision are valuable to advancing the current state of the art, the tools and platforms are nonetheless mystifying for non-experts. Many of the tests on M-Lab's platform require the use of Java applets on the client side and explicitly require users to visit the measurement site to run the test. These drawbacks retain the selection bias similar to that of Speedtests, which may favor "geeky" technologically knowledgeable users. Querying M-lab's data requires knowledge of SQL and relational databases, which is taught in higher-level computer science courses at the university level.

With the shortcomings of current state of the art in mind, it is imperative to design a measurement platform that seeks to sample and include as broad of an audience as possible. In order to have a measurement platform be inclusive, it must be low cost with a low barrier to entry. A measurement platform with a low barrier to entry must not require additional software or hardware. Thus, net-score seeks to satisfy these motivations and address such shortcomings.

III. Net-score Architecture

Test setup and procedure

While the net-score platform currently has several test tools developed, the focus of this paper spotlights the net-score feather test. The feather test consists of seven lines of JavaScript and is embedded at the end of an HTML file. The tool is also not visible to users and does not affect the user interface of the website on which it is embedded. Once started, the test runs in the background while a user browses, and test results is reported to a server maintained by the University of North Carolina at Chapel Hill. Secure UNC Computer Science department logins and knowledge of the address of the server are required to access the data recorded by net-score tests.

To run a net-score feather test, a client must first navigate to a third party website that hosts the tool. Since the W3C specification for HTML5 states that elements of HTML are loaded sequentially, the net-score tool is retrieved last, after all other elements in the page.

Once the client receives the test tool, the test begins. First, a HTTP1.1 GET request is issued for an image weighing only 362 Bytes located on YouTube's Content Distributed Network (CDN). HTTP GET requests is made possible through the XMLHttpRequest (XHR) API, which is available to many web scripting languages, including JavaScript. The round trip time of request to response is measured using wall clock time, to include any input and output handling on the server's end. This process is repeated several times and the resulting round trip times are averaged into a single latency value.

After latency is calculated, the tool then employs a similar process to estimate the client's throughput. A HTTP request is issued for a larger image weighing 252 Kilobytes. Again, we use the elapsed wall clock time to measure round trip time. Latency is subtracted from this value to account for any time needed for a handshake between the client and server. Finally, this elapsed trip time is divided by the size of the file and the client's throughput is calculated. Like the calculation for latency, this process is repeated several times and averaged to provide a single throughput value.

Deployment

Since net-score is an embedded tool, selecting a website for deployment is also an important design decision. To reach a large audience, we selected TarHeelReader.org for our seven-month case study. Tar Heel Reader is an educational website for people learning to read and is used frequently in the classroom. The website is also built to be accessible to students with disabilities and can be used on all types of platforms. Due to its global audience on a variety of platforms, Tar Heel Reader provided an excellent environment to test how net-score scales to capture performance of new audiences.

IV. Advantages of net-score

The architecture of net-score has several key advantages over other measurement platforms. First, using the browser allows us to keep costs low for users taking the test and providers who embed the tool on their website. Users do not see a significant change in web page load times since the net-score test does not begin until after all resources for the web page are loaded (HTML Standard). The feather test that is embedded on the web page is only seven lines of JavaScript, which retrieves a helper script of only 263 Bytes. Consequently, the setup for the net-score test is trivial.

For websites hosting the tool, net-score tool is not costly. Embedding the test on a website is a matter of adding an additional script block to the end of the page before the close of </body> HTML tag. Since no resources used for the test are hosted on the third party website, including the script block is the only cost to hosts. The feather test described in this paper also has no visible interface, so there are no intrusions to the design of the third party website.

For us, net-score costs very little to maintain. Net-score leverages existing infrastructure to minimize operation costs and to use a scalable platform to distribute the necessary code for running tests. By using YouTube's well-provisioned edge servers, we benefit from using the content distribution network's load balancers and edge caching. Current charges remain under $1 a day for supporting up to 4,000 daily responses. Also, as previously discussed, consumers of the Internet have few options to invoke change in their connection speed. The main choice consumers have is to change their ISP, which carries their network traffic to the backbone of the web. Using YouTube's CDN shortens the path the file transfers take, following the path of infrastructure maintained by local ISPs. Because we make use of these short paths, we can replicate speeds of typical web requests retrieved from popular websites, rather than contacting static, regional servers.

Embedding the tool in the browser also relegates the need for special hardware, special software, or even the necessity to visit a particular website to run tests. Since our tool runs in the background, net-score tests behave similarly to analytics engines and data collection does not require opt-in from the user. This streamlines data collection by lowering the barrier to entry for new net-score users and as a result, mitigates selection bias.

Ookla Speedtest's user base certainly must be global, but the fact that a user is required to navigate to their website, which holds no other content, to take the speed test suggests there may be large populations not captured by Ookla's measurement tool. Ookla has to drive users to their website, whereas net-score feather tests can be embedded in third party sites. In summary, net-score eliminates bias through a unique ability to reach out to new users rather than have to drive users to a specific website. By gathering the support of websites, net-score can scale to sample large new populations and significantly increase its baseline performance data rather than have to grow slowly through attracting individual users.

We observed this significant increase in users upon embedding our tool in a single popular website. We outline the data collected in the next section.

V. Evaluation Methodology and Data Collected

Over our seven-month long test deployment, we recorded almost 200,000 data points from embedding in a single website. As described in previous sections, net-score tests recorded latency and throughput values for users. Geographic data was collected during each test, as well. Country, region, and city of the user were recorded along with platform data such as browser render engine and operating system. This information is sent with the user agent string in the HTTP request headers. From these measurements, we were able to visualize the reach of our audience. We also evaluated if the net-score measurement platform achieves the level of reach and scalability motivated in earlier sections. The data we collected during our small deployment revealed that net-score reached a global audience in a short period of time.

Countries represented

Breaking our represented countries into five major regions, we observed high representation in each region (Figure 1). Tests were taken in all 36 countries in the Americas, as well as Bermuda and Puerto Rico, which were counted separately from their sovereign nation. 44 out of the 50 countries of Europe, 9 out of the 14 countries in Oceania, and 45 of 48 countries of Asia were represented. Africa was least represented, with only 19 of the 54 countries running tests.

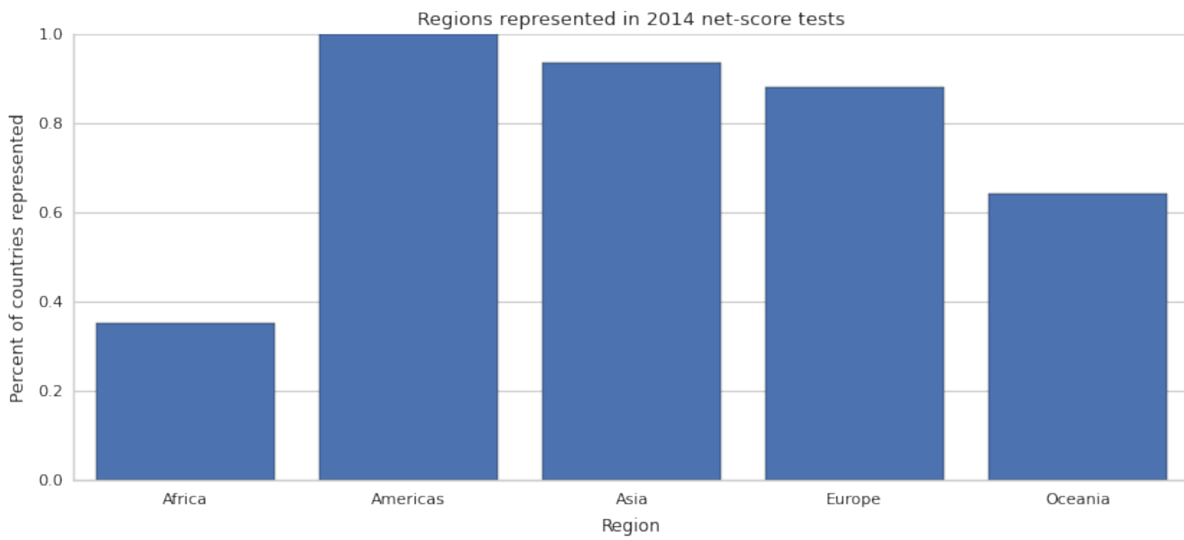

Figure 1: Percent of countries represented by region

To get an idea of the experience of our global audience, we generated a map of the average throughputs of each location that recorded a test. From this map, we can also see the global distribution of our audience. Central Africa is largely unrepresented. This could result from a lack of Internet infrastructure, population scarcity, climate, or a simple lack of knowledge of Tar Heel Reader. Also from our map, we see a concentration of users in coastal areas in developing countries, consistent with highly populated cities in those areas. A screenshot of our map is provided with a shortened URL to an interactive web version is provided in the caption.

We observed that faster speeds, the larger balloon icons on the map, are more prevalent in the West as well as East Asian countries such as Japan, Korea, Hong Kong, and some cities in

Southern India. Surprisingly, we observed near 3Gb/s speeds in Mongolia's capital as well as Colombo, Sri Lanka. These could possibly be military installations or anomalies in our data, but with 9 and 92 tests taken in those cities respectively, it is very possible these areas have state of the art fiber networks.

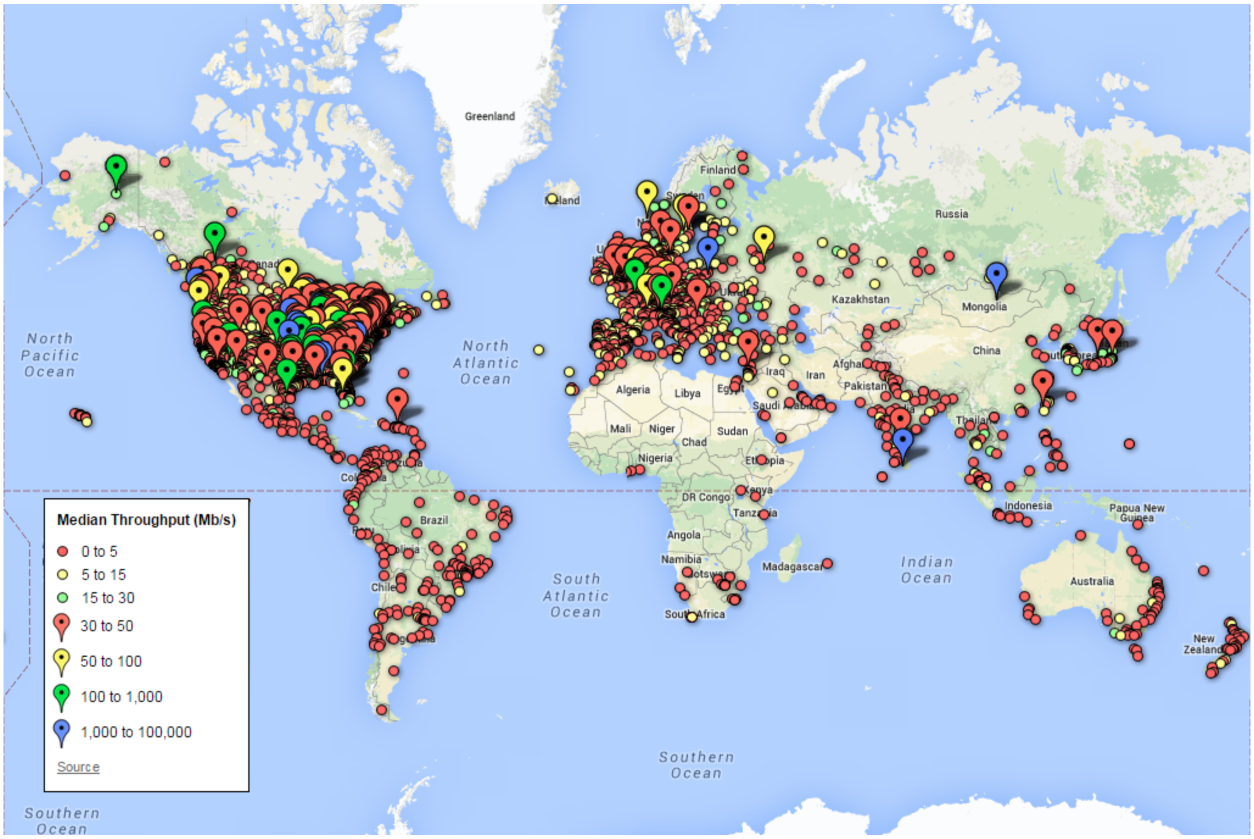

Figure 2: Shortened URL: http://goo.gl/rArhtJ

Additionally, we found that we reached many Internet users with connection speeds less than 500 Kilobits/second. By filtering the map shown in Figure 2, we can easily spot regions that may be lacking in Internet infrastructure.

In Figure 3, we display a breakdown of the number of tests taken in each region, plotted on a logarithmic scale. While a large number of the tests were taken in the Americas, Asia and Europe are well represented. Africa recorded between 100 and 1,000 tests taken.

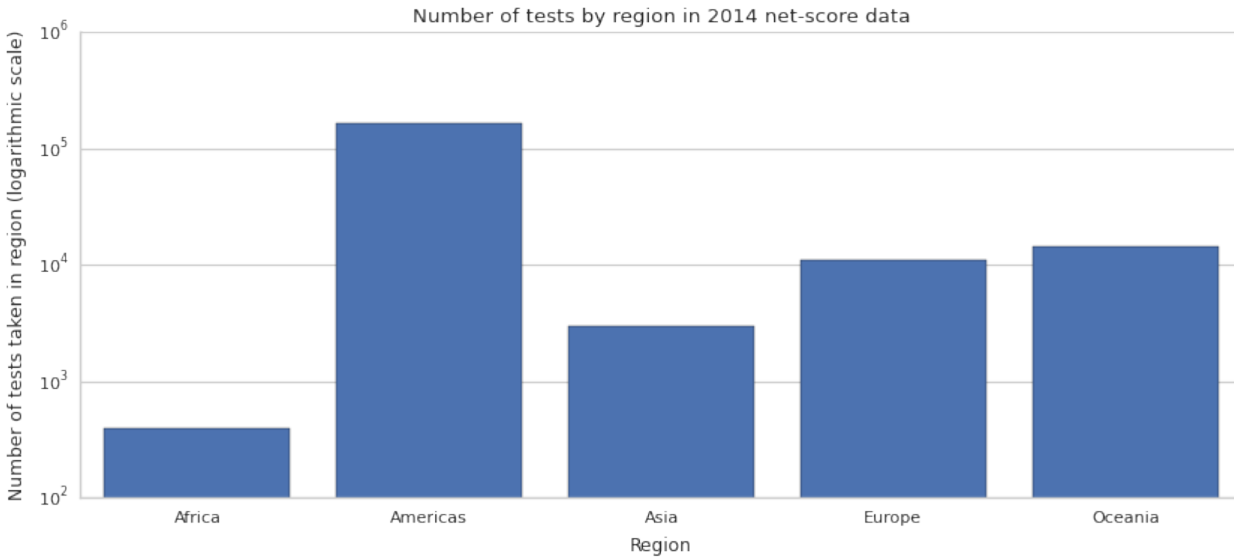

Figure 3

Platforms represented

Our tool recorded tests on a variety of platforms. Mobile technologies such as Android, BlackBerry, iPad, iPhone, iPod Touch, and Nokia devices running Symbian OS were represented in our data set. Traditional desktop and laptop platforms were expectedly well represented, with tests being run on Mac, Windows, and Linux. Non-traditional platforms such as Playstation 3 and Nintendo WiiU gaming consoles also successfully ran net-score feather tests. The breakdown of how many tests run on each platform is visualized in Figure 4.

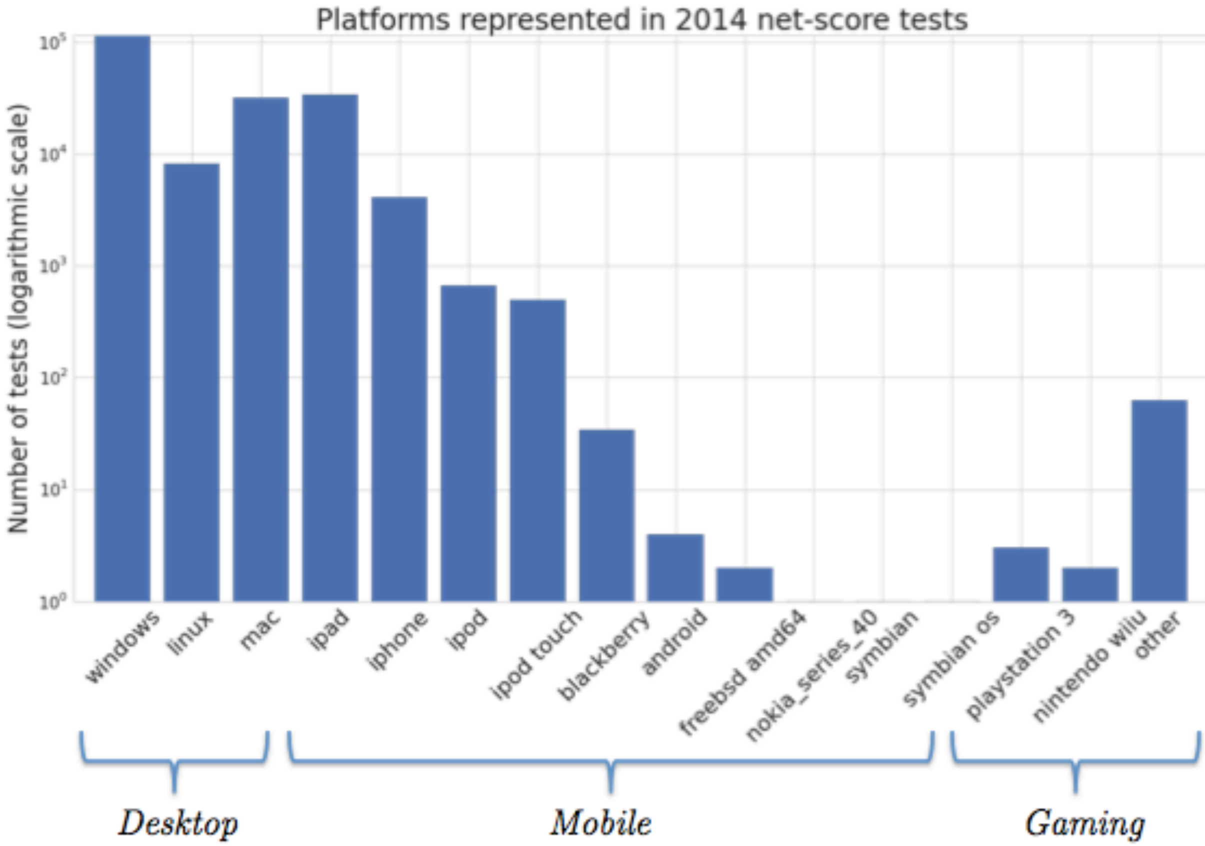

Figure 4

Browsers

The ability to run net-score tests on Symbian OS, Android, and gaming consoles is remarkable. A true testament to the ubiquity of the web, these tests exemplify net-score's ability to reach users not currently represented by current state of the art. To visualize the distribution of browsers represented in our tests, we looked at the render engine portion of the user agent string. We saw that three main render engines, Gecko, IE, and WebKit were about evenly represented (Figure 5). Following in fourth was the Opera rendering engine. Our findings correlate with StatCounter's tabulation of popular desktop browsers, as WebKit is used by both Chrome and Safari, and Gecko being used by Firefox.

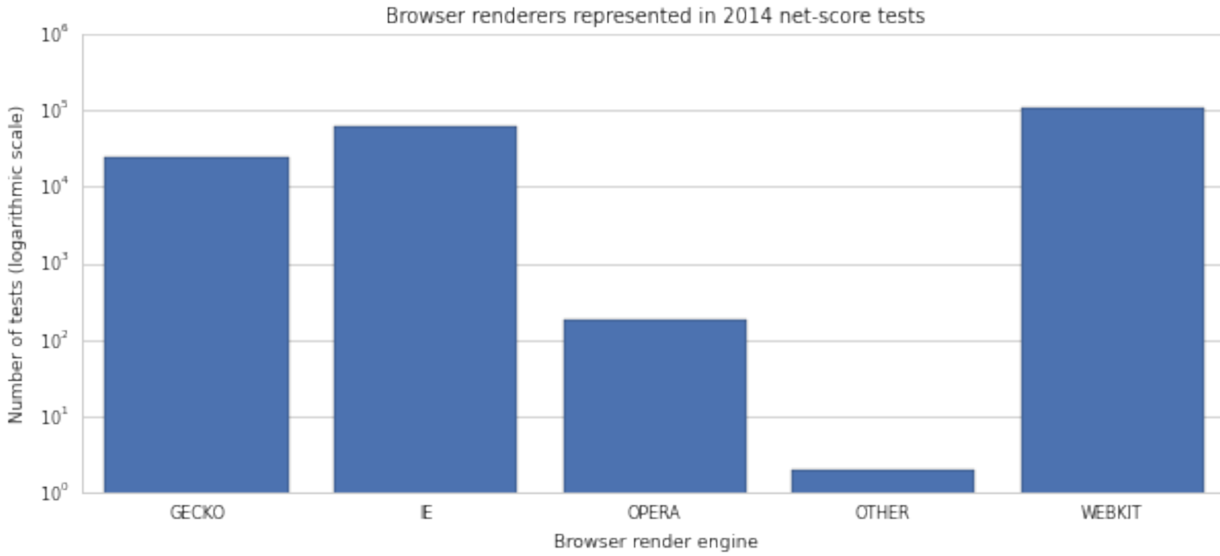

Figure 5

Autonomous systems

The reach of our tool was primarily educators and students using Tar Heel Reader in the classroom. Looking at the most popular Autonomous Systems, the New York City Board of Education and MCNC reside as the second and third most popular systems. MCNC maintains the connection between UNC-Chapel Hill, Duke, and Wake Forest universities. Though we only sampled a population from a single website, we reached 3,540 unique Autonomous Systems. According to Cisco, of the 34,827 assigned AS numbers, only 21,191 are advertised (Huston). This means we reached 16.71% of advertised and publically reachable Autonomous Systems.

While there is no current process for converting AS numbers to the ISP that carried the traffic, it is possible to research the name of the organization of which the AS belongs to. Our five most popular AS numbers were:

AS Number	*Name*	*Organization*	*Count*
7018	ATT-INTERNET4	AT&T Services, Inc., US	9799
81	NCREN	MCNC, US	6942
21704	NYCBOE-BGPNET	New York Board of Education, US	6433
701	UUNET	Verizon Business, US	5672
22773	ASN-CXA-ALL-CCI-22773-RDC	Cox Communications Inc., US	5112

Data collected

Maximum, minimum, median, mean, and standard deviations of both latency and throughput measurements were recorded for each test. In previous sections we detailed the importance of both measurements. Our tool recorded a wide range of measurements, with throughputs ranging from near zero to 2Gb/s. What follows is a series of density coded heat maps that help visualize what latencies and throughputs our clients experienced. The x-axis of the heat maps are the delay or latency experienced by the client, and the y-axis is the throughput. The redness of each region represents how many clients were clustered in that range of latency and throughput values. To begin, we apply a filter to the latencies of two minutes, which is the timeout for connections over TCP/IP.

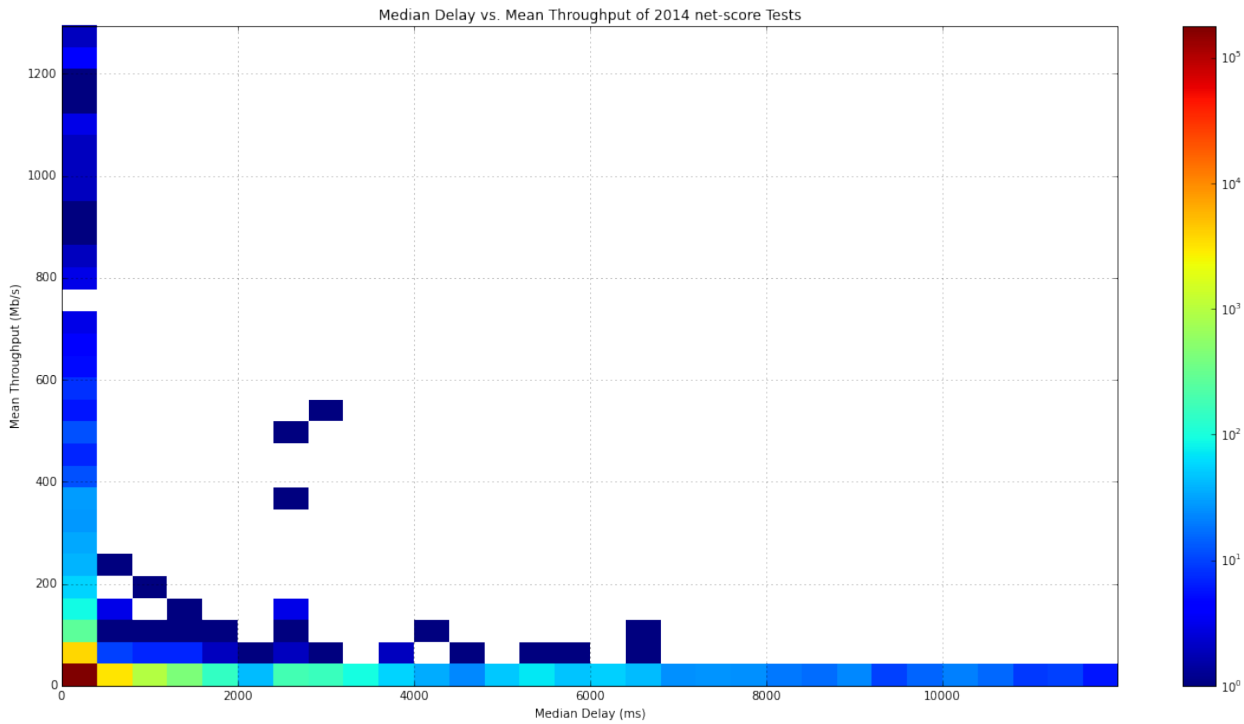

Figure 6: Filtered distribution of tests with latencies < 12,000ms

To get a better idea of the speeds experienced by users with our highest latencies, we apply a filter on the mean throughput values to only plot speeds up to 100 Mb/s. This level of granularity shows the distribution of speeds in the upper quartile of latencies (Figure 7). Users who experienced these long delays mostly experienced low throughput. However, we observed users with good throughput values despite long delays.

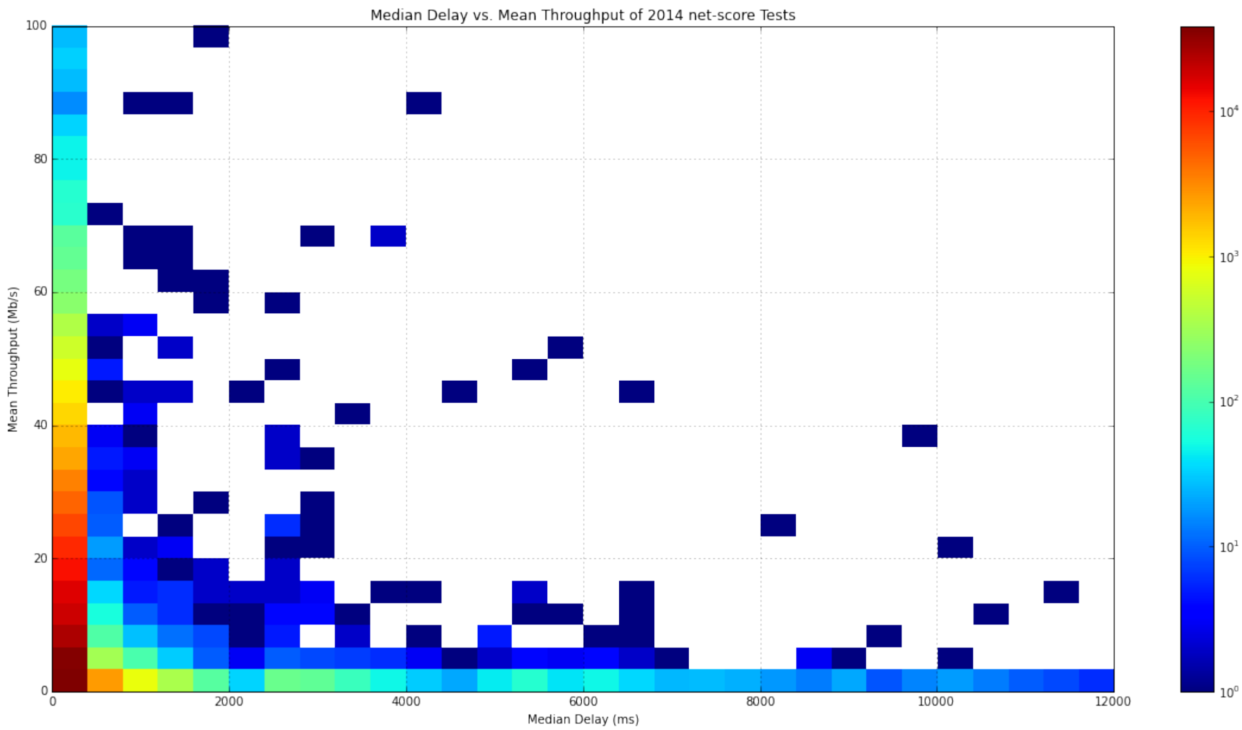

Figure 7: Filtered tests with throughputs < 100Mb/s

Finally, we examine the densest region of our data. Users in the densest regions shown in Figure 8 are well suited for web browsing, as web browsing is still possible with delays of 40-60ms.

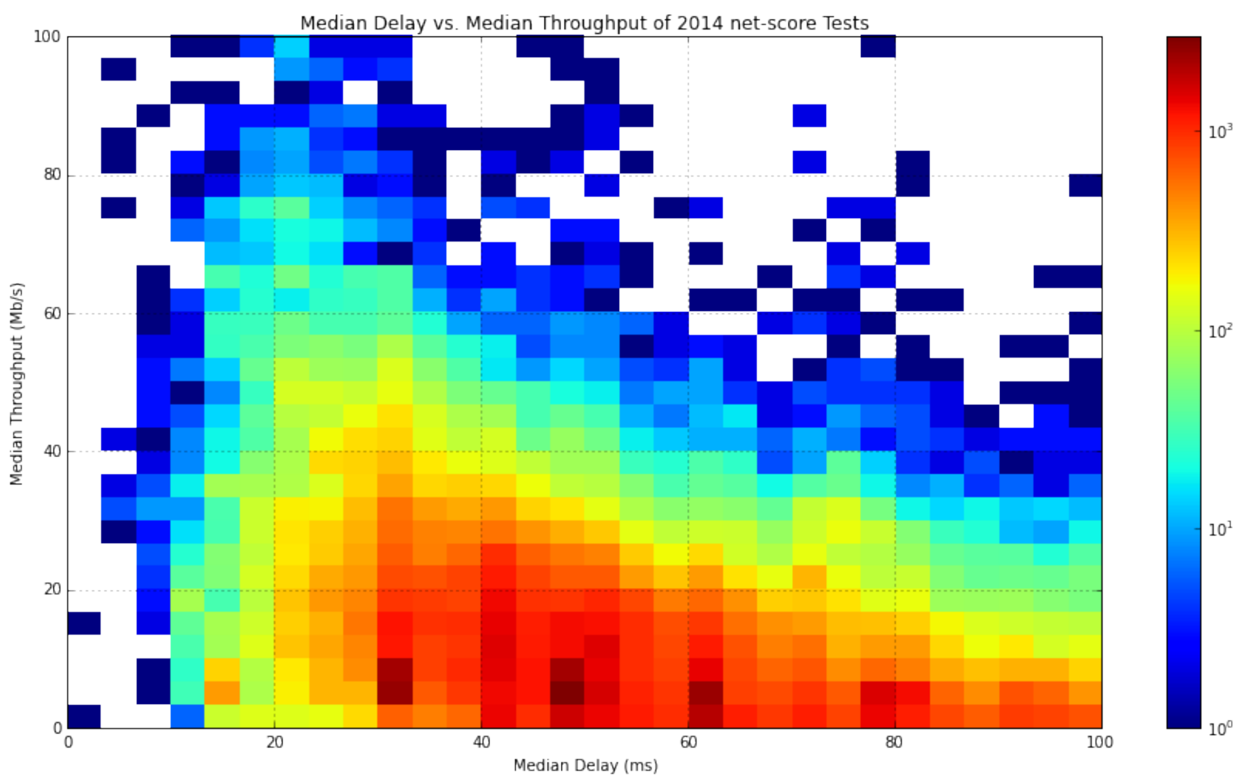

Figure 8: Densest region of our heat map

We also wished to visualize the cumulative distribution functions of each important measurement. These provided insights into what our median client experienced. The median latency experienced by our clients was 57ms, and the median throughput of our clients was a healthy speed of 8.03Mb/s.

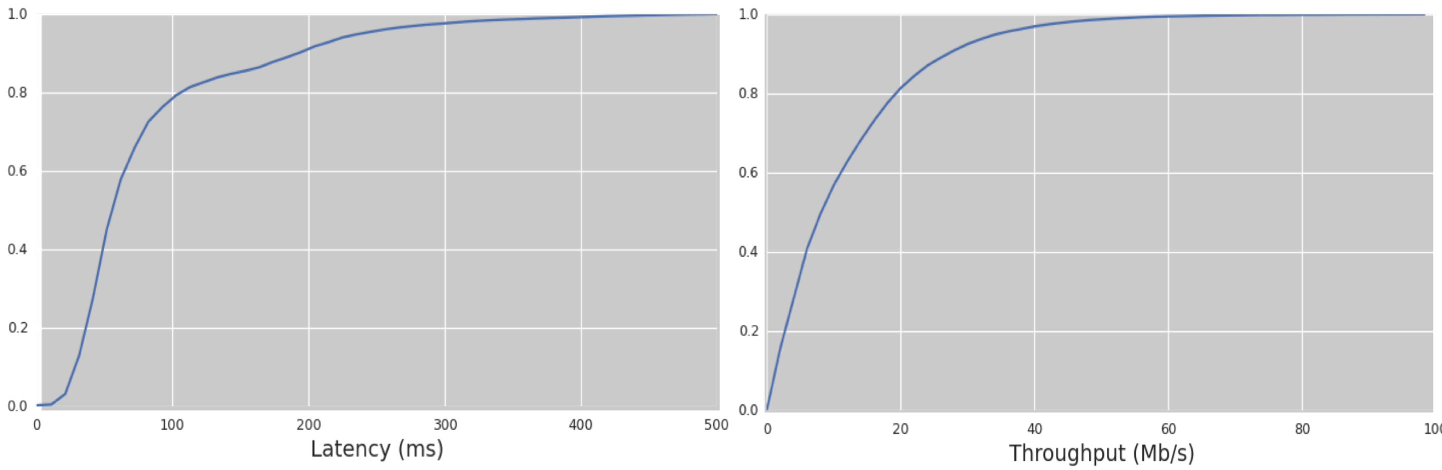

Figure 9: Cumulative distribution functions for latencies and throughputs

We believe that the reach we achieved in seven months on a single website is a testament to the cross-platform capabilities of using a browser-based tool. In the next section, we further comment on our data and discuss whether we accomplished the goals set forth for an effective and modern measurement platform.

VI. Concluding Remarks

Inclusiveness

By moving the measurement platform to the browser, we successfully reached an audience in 80% of the globe and in 16.72% of reachable networks worldwide. While net-score can be used to diagnose a connection the same way Speedtests can, we have proved that we can represent populations not just looking to diagnose their connections. While our audience may be biased towards educators and students using Tar Heel Reader in the classroom, we have shown it is possible to capture the network performance of users of any site. By embedding the tool in an educational website, we have captured network performance of schools across the globe, which may be good indicators of the publically available network infrastructure of smaller towns as well large cities. We also have the ability to find areas of the globe which do not have strong network infrastructure, positioning us well to recommend improvements. Expanding net-score to other third party websites will allow the inclusion of additional populations that may not have been reachable otherwise.

Consequently, instead of having to attract users one-by-one, as with Speedtest, we can include hundreds of thousands of new users in our data by embedding the tool into another third party website. We also eliminate the “geeky” bias present in current state of the art through this ability to be embedded in any website.

Low barrier to entry

Reaching this global population was possible through the ubiquity of browsers. Because our tests use a browser based API, there is no requirement to download any additional software or use of specific hardware. The tests taken on Nokia and Android mobile platforms reflect our ability to reach users with the most minimal of hardware. Running Speedtests on mobile platforms without flash is not possible; we have proved that even in our minimal deployment that we can reach mobile users with no additional apps or plugins. Since our platform runs in the background of third party websites and do not require user action, we streamline the necessary steps to run net-score tests. With these design choices, net-score has the most minimal barrier to entry. The only limiting factor to running a net-score test is enabling JavaScript, which is the default setting for all browsers.

From the map of our user base, many tests were taken in a diverse set of developing countries. By reaching Internet users in these countries, who may not have access to network infrastructure as robust as Western countries, we exemplify a measurement platform with a low barrier to entry.

Low cost to users and providers

As mentioned in previous sections, websites incur no cost by hosting the net-score tool. Because images are downloaded from a YouTube edge server, the user does not incur any costs and the speed of completion of the test is dependent on the user's connection. Net-score is also low cost to users because there are no noticeable differences in the functionality of the third-party website since the test does not begin until the rest of the webpage has loaded.

Operating costs for us remained around $1/day for the duration of the seven-month case study. We have not pushed the bounds of these costs yet and further study is needed to find a correlation between these low costs and the number of tests performed.

Based on the data collected, net-score accomplished its goal of reaching a broad audience by being a measurement platform with a low barrier to entry. Using the browser and running tests in the background proved to be a good design choice for unbiased data collection. We believe that the flexibility of embedding the tool into third party websites eliminates any selection bias created by driving users to visit a single website to specifically test their connections. While we do provide this functionality, the measurement platform of embedding the feather test in third party websites allows us to reach populations of Internet users of all types.

Our future work will be focused on using the data we collect through the feather test to provide a normalized score for test takers. We also hope to gain insights on how closely user's experiences match with the speeds for which they pay their service provider. These questions introduce new challenges, especially that of attempting to provide a single score from multi-dimensional measurements.

Works Cited

Dobrian, Florin, Vyas Sekar, Ion Stoica, and Hui Zhang. "Understandng the Impact of Video Quality on User Engagement." *SIGCOMM '11*. Toronto. ACM, Aug. 2011. Web. 5 Apr. 2015.

Dovrolis, Constantine, Krishna Gummadi, Aleksandar Kuzmanovic, and Sascha D. Meinrath. "Measurement Lab: Overview and Invitation to the Research Community." *SIGCOMM '10*. ACM, July 2010. Web. 5 Apr. 2015.

Huston, Geoff. "Exploring Autonomous System Numbers." *The Internet Protocol Journal* 9.1 (2006): 2-23. Web. 5 Apr. 2015.

Palazzo, Anthony. "Estimated Pay-TV Cord Cutters Topped 1 Million in 2012." *Bloomberg Business*. Bloomberg, 2 Apr. 2012. Web. 5 Apr. 2015.

Plunkett, John. "Decline of the Phone Call: Ofcom Shows Growing Trend for Text Communication." *The Guardian*. 17 July 2012. Web. 5 Apr. 2015.

Schurman, Eric, and Jake Brutlag. "Performance Related Changes and Their User Impact." Velocity 2009. The Fairmont, San Jose. 24 June 2009. *YouTube*. Web. 5 Apr. 2015.

"StatCounter Global Browser Statistics." *StatCounter Global Stats*. Mar. 2015. Web. 05 Apr. 2015.