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 have 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 general 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. 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.

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 facilitates customer retention and 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 [Eric 21, 26, 30]. 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 user friendly. 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 must possess qualities to minimize bias and capture as wide of an audience as possible. A measurement platform that is accessible to all will be low cost and have a low barrier to entry. In the context of measurement platforms, low cost takes on two meanings. The first is being low in monetary costs. Expensive infrastructure may limit the scalability of the tool, meaning it may not be able to handle serving tests to a large population without incurring costs to the developers of the platform. The second meaning of low cost is in terms of the platform user’s time. The test should be proportionally fast to the user’s connection. If the test takes too long to run, even on fast connections, error can be introduced or the test may be disrupted. In the ideal scenario, the user would not recognize that the measurement platform is running and they could continue using their broadband connection without interruption.

Furthermore, a measurement platform with low barrier to entry will not require significant steps to use. Downloading additional software or requiring additional hardware increases the barrier to entry of using the platform and limit the reach of a measurement platform. If such steps are costly in time, money, or both to a platform’s users, then selection biases can be introduced into the gathered data. Only users with the time or money to use the platform will be represented.

A measurement platform possessing these two qualities must also be able to scale to reach new populations. If a platform is truly low cost with a low barrier to entry, the measurement platform will be collecting a large amount of data. The platform itself must be able to handle serving the tests and be able to store the information securely for later use. The platform also must not actively or passively exclude populations based on geographic location or socioeconomic status.

Lastly, but certainly not least important, 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. 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, which we will discuss 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 on in the field, we observed almost 200,000 tests taken by at least 7,000 unique users. These 7,000 users came from across the globe, and 150 countries were 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 “geeky” users who use the tool when they experience unsatisfactory performance. By “geeky” users, we mean users who are familiar with metrics that explain network performance, such as throughput and latency.

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. Since Ookla reports rankings for ISPs and certain ISPs set up dedicated regional servers, a conflict of interest arises. Previous studies have determined that ISPs have the ability to prioritize traffic to these dedicated servers.

Speedtests, in summary, are a robust but blackbox method of measuring network performance. Another valuable tool for evaluating network performance are 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 easily 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 new, wide user base. Current tools deployed on M-Lab’s infrastructure include:

1. NDT: a test providing detailed packet level information and kernel-level statistics on how the 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” (M-Lab Overview). 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.

Hence, it is imperative to design a measurement platform that is not biased towards any populations and does not exclude any population from data collection. 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 the shortcomings of the current state of the art.

III. Net-score architecture

As previously described, a measurement platform must be capable of accurately recording metrics that are relevant to the behavior of the users participating in the platform. Net-score runs completely in the browser and works to measure the performance of typical Internet users.

We chose TarHeelReader.org, an educational website for students learning English, to embed our tool to test its feasibility. Tar Heel Reader sees voluminous traffic from all over the globe. The website is also developed for devices which facilitate tactile movement of students with disabilities, meaning that it can be accessed from platforms of all types. We discuss the platforms represented in our data collection in following sections. Due to this broad audience, Tar Heel Reader provided an excellent website for an initial deployment of the net-score feather test. This test is the lightest weight tool on the net-score platform and is run in the background of a third party website, similar to an analytics engine.

The two core functionalities of the feather test are measuring a user’s latency, also referred to as delay, and estimating a user’s experienced throughput. These two measurements are generated using similar methods and aim to capture typical user application level behavior. To measure latency, the net-score tool issues a HTTP1.1 GET request for an image weighing only 362 Bytes located on YouTube’s Content Distributed Network (CDN). Using YouTube’s CDN allows us to localize the path to the file for each user, providing a better idea of last mile latencies. The round trip time is recorded by measuring the wall clock starting when the request is sent, and ending when the response is received. Net-score uses the wall clock to account for input and output handling on both the server and client side. To provide a more accurate estimate of a user’s experienced delay, the request for the lightweight image is issues many times and round trip times are recorded. After the individual round trip times have been calculated, an average is calculated and reported.

The estimation of the connection’s download throughput employs a similar process. A larger image of size 252 Kilobytes is retrieved from a YouTube edge server repeatedly. Wall clock time is recorded for each request. We correct for delay from the server handshake by subtracting the previously averaged round trip time from this recorded throughput time. Finally, we compute the throughput value by dividing the recorded time by the size of the file. This process is repeated, and an average throughput value is calculated and reported.

Our platform uses the XMLHttpRequest (XHR) API to generate HTTP requests, which is available to many web scripting languages. Because this API is so popular and integrates well into JavaScript, net-score achieves a high level of browser compatibility. Tests have not only been run on platforms as universal as Windows, Linux, and Mac, but on PlayStation, WiiU gaming consoles, and iPads. Mobile platforms such as Symbian, BlackBerry, Android, and iPod Touches, are compatible as well. Though HTTP is not used by every Internet application, net-score measures the behavior of a broad audience by capturing performance of a widely implemented protocol.

Test results from feather tests are stored securely on a server maintained by the University of North Carolina at Chapel Hill. Students and faculty possessing secure Computer Science department logins and the address of the server can access this data. With these design choices, we believe we hold several key advantages over current state of the art.

IV. Advantages of net-score

The architecture of net-score has several key advantages over other measurement platforms. First, the browser is a low cost, low barrier-to-entry platform that makes growing a user base easy. The browser is low cost to both users who take 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. The speed of retrieving images for the test is a function of the user’s connection, so it cannot be accounted for when evaluating the speed of the test.

For providers, the net-score tool has little cost. 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 providers. Providers can also choose to deploy the feather test to their website, which does not provide a visible interface for users taking the test. The feather test discussed in this paper consists of a script block with only six lines of JavaScript, which pales in comparison to entire websites dedicated for performing similar tests.

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 a static, regional server.

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 similar 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. While Speedtest.net boasts seven billion tests taken, the demographics of their audience is not clear. In addition to providing their web test, Ookla provides a framework for ISPs to help customers troubleshoot their connections. Providers such as Time Warner Cable and Comcast have dedicated regional Speedtest servers that users contact when testing their connections. At the surface, this appears to be a conflict of interest since it is possible for providers to prioritize traffic to and from these servers.

Ookla’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 tools 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

*Tar Heel Reader*

To evaluate whether net-score accomplishes the motivations outlined in previous sections, we deployed the feather test in the field for seven months on a frequently trafficked website. TarHeelReader.org, or Tar Heel Reader, is a collection of easy to read books available for free, maintained by Dr. Gary Bishop of UNC Chapel Hill. Tar Heel Reader is widely used by educators in classrooms as well as global citizens looking to learn English. We deployed the feather test, which as described, runs in the background and is hidden from the user’s view.

As observed by Dr. Bishop, the tool did not affect overall server load since the tool retrieves images from YouTube’s CDN. The tool is loaded last on the webpage when embedded just before the closing body tag of the HTML, so load time for the content of the webpage is not affected (HTML standard). During the seven month long case study, the tool collected 194,239 data points between January 1st 2014 to August 1st 2014.

*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. Latency figures were widely ranging as well, and an unfiltered distribution of the data is shown in Figure 1. We apply a filter to the TCP/IP timeout of two minutes or 12,000ms in Figure 2.

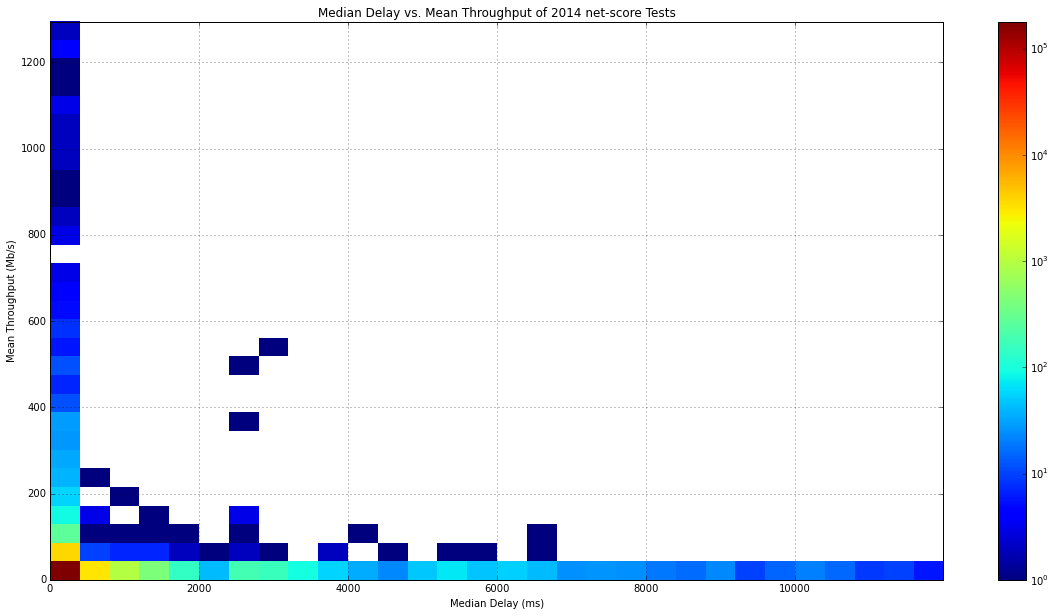


Figure 1: Unfiltered distribution of tests

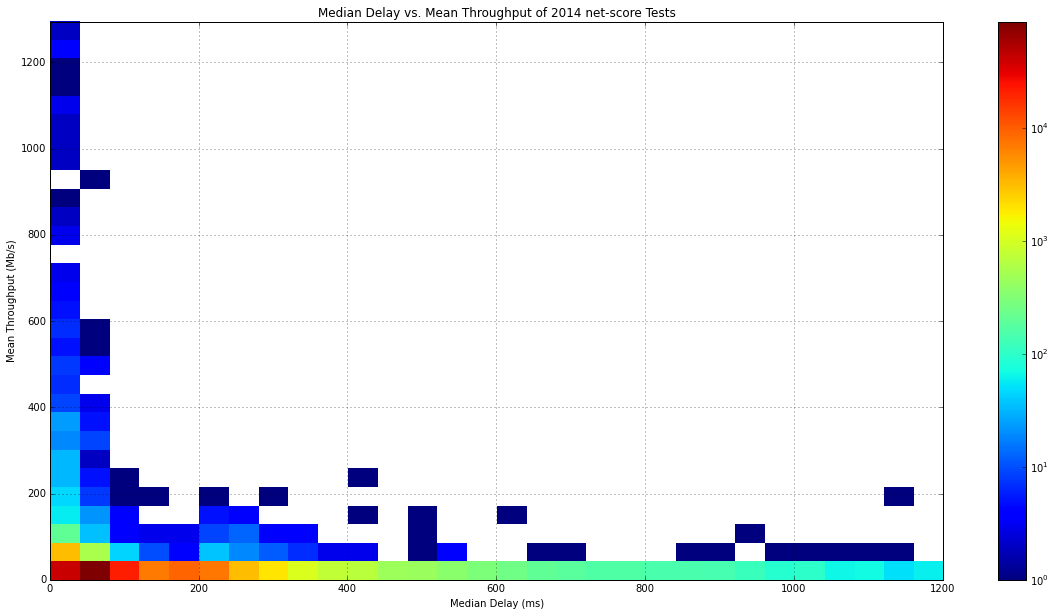


Figure 2: Filtered tests with latencies < 12,000ms

Figure 3 zooms in on tests with median latencies under 300ms to better visualize the distribution of throughputs given a fast latency. From our unfiltered distributions, we observed that a significant number of tests saw latencies between 0ms and 300ms, so we plot them here.

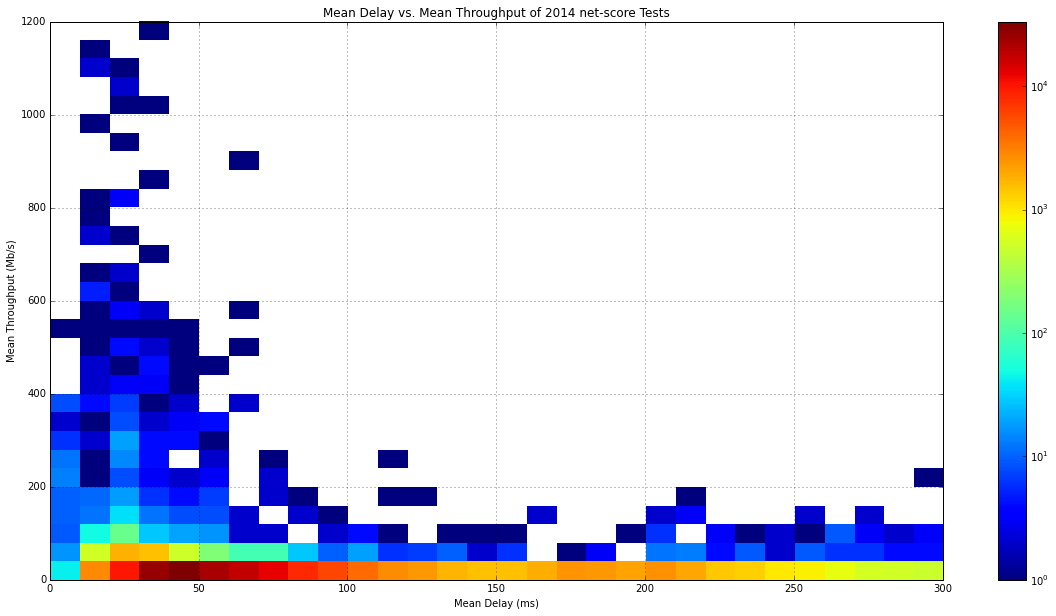


Figure 3: Filtered tests with latencies < 300ms

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 4).

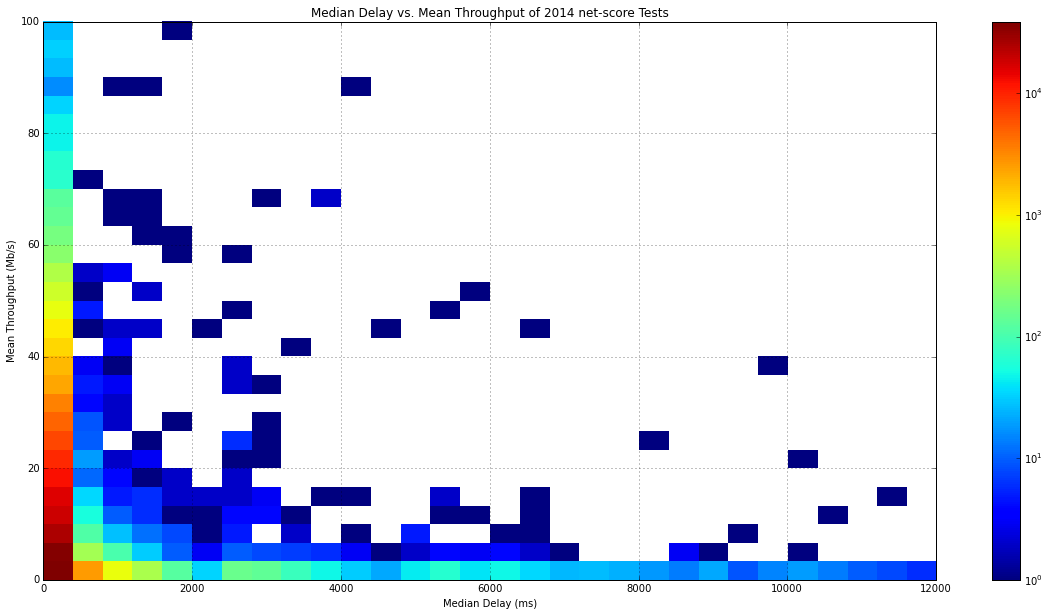


Figure 4: Filtered tests with throughputs < 100Mb/s

We also wished to visualize the cumulative distribution functions of each important measurement. They are plotted succinctly below.

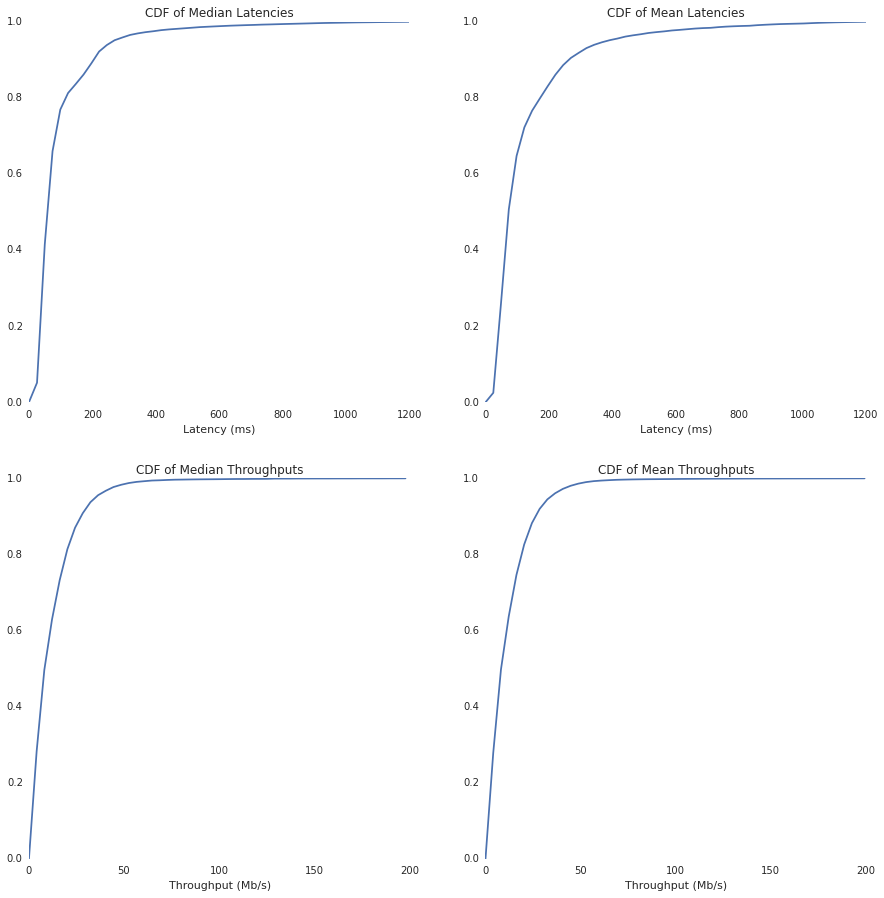


Figure 5: Cumulative distribution functions for latencies and throughputs

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. Embedded in a single website, we captured tests from 150 different countries with over 7,000 unique user agent strings.

*Countries represented*

Breaking our represented countries into five major regions, we observed high representation in each region (Figure 6). 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.

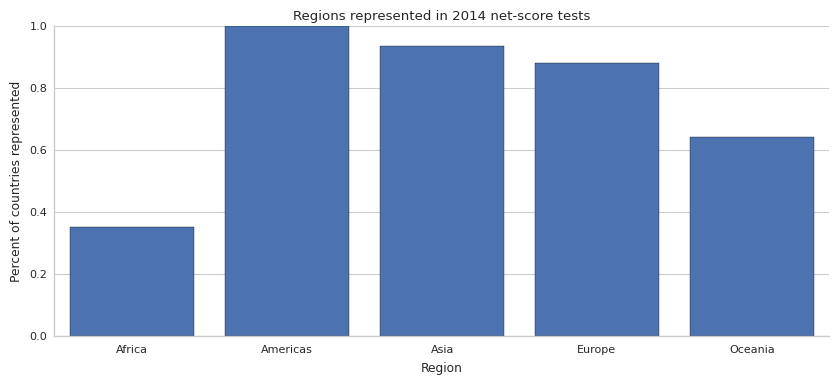


Figure 6: Percent of countries represented by region

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.

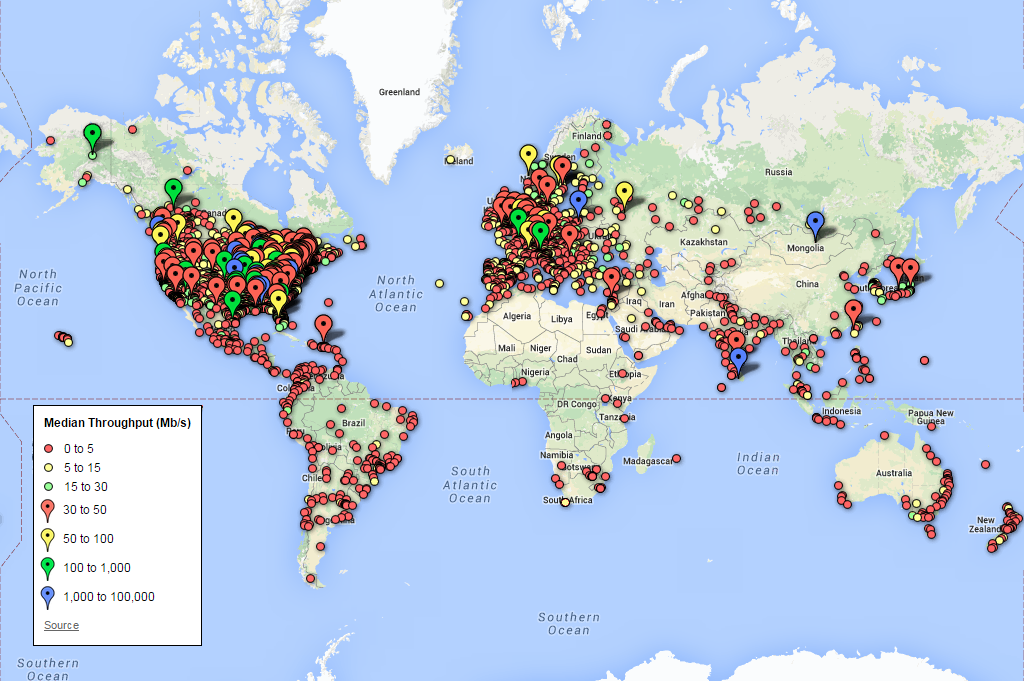


Figure 7: Shortened URL: <http://goo.gl/rArhtJ>

In Figure 8, 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 saw between 100 and 1,000 tests taken.

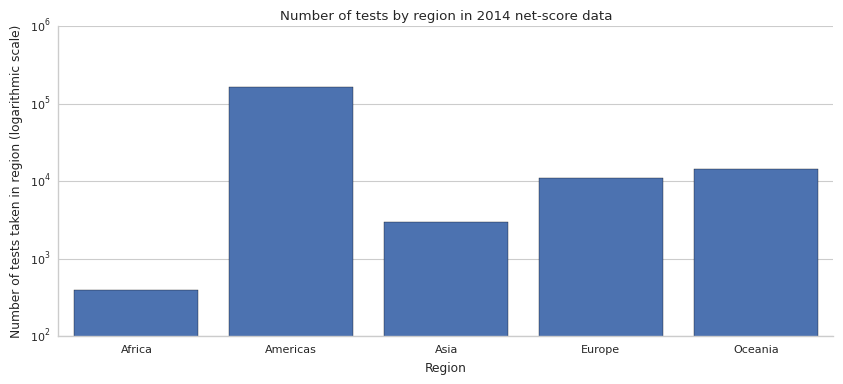


Figure 8

*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 9.

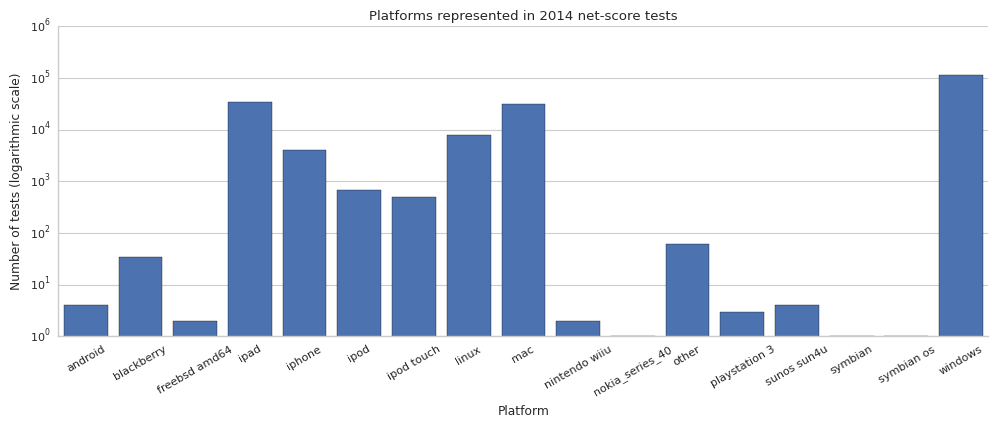


Figure 9

*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 10). 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 (http://gs.statcounter.com/#desktop-browser-ww-yearly-2014-2014-bar).

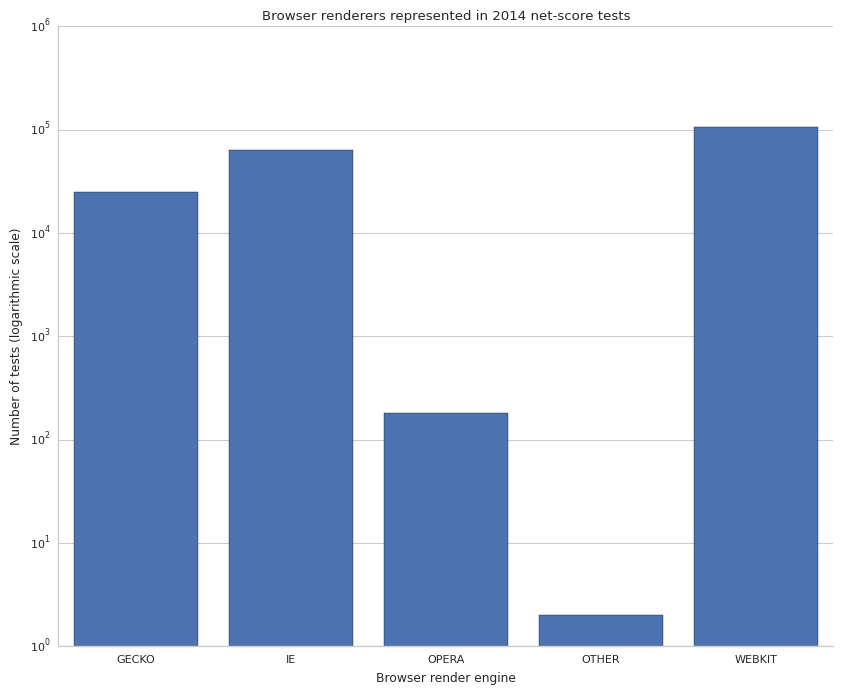
****

Figure 10

*Autonomous systems*

The reach of our tool was primarily educators and students using Tar Heel Reader in the classroom (Gary’s Google Analytics). Though we only sampled a population from a single website, we reached 3540 unique Autonomous Systems. According to Cisco, of the 34,827 assigned AS numbers, only 21,191 are advertised (Exploring Autonomous System Numbers). 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 City Board of Education,US | 6433 |
| 701 | UUNET | Verizon Business,US | 5672 |
| 22773 | ASN-CXA-ALL-CCI-22773-RDC | Cox Communications Inc.,US | 5112 |

Table 1

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. Remarks

*Inclusiveness*

The net-score feather test uses the XMLHttpRequest (XHR) application programming interface to issue HTTP requests for images to measure latency and throughput. Since XHR is available to a multitude of programming languages, such as JavaScript, it can be utilized in many web page rendering engines. As our data shows, we saw large numbers of tests run on Gecko, IE, Opera, and Webkit renderers. Furthermore, net-score tests were run on mobile, desktop, and gaming platforms. Also, as mentioned, 155 countries were represented along with 3540 autonomous systems.

By moving the measurement platform to the browser, we successfully reached an audience in 80% of the globe and in 7.5% of networks worldwide. Since the tool was embedded in an educational site, users that visited the site and took the test were not using net-score to diagnose their connections. While net-score can be used to diagnose a connection the same way Speedtests can, we have proved that reaching new populations is possible. 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. However, 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 network infrastructure of smaller towns as well large cities. 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 such minimal hardware are not possible; we have proved that even in our minimal deployment that we have reached such users. 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 the west, we exemplify a measurement platform with a low barrier to entry.

*Low cost to users and providers*

As mentioned in previous sections, net-score incurs no cost to the websites that host the 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. However, test completion times are comparable to Speedtest completion times (need source). 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. (Reminder to use some data from Eric)

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 “geeky” 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.