

Revised collection entitled: Consumer Fixed Broadband Services Testing and Measurement

Part B: Collections of Information Employing Statistical Methods:

1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.

The target population for this study is the American consumer with broadband. The survey delivers reliable national estimates of broadband speed performance for the US Internet Service Providers (ISP) delivering broadband service to the majority of fixed and mobile consumers. The program current supports data collection from a set of roughly 15,000 fixed broadband hardware devices and a crowd-sourced mobile broadband measurement using iOS and Android Application clients.

Fixed Broadband Measurement Statistical Methods

In response to RFQ-10-000013 issued on March 24, 2010, a contract award was made to SamKnows to conduct the FCC's fixed broadband performance data collection, and has grown to include both mobile broadband performance and satellite broadband. SamKnows developed a panel of over 15,000 fixed and satellite broadband subscribers, geographically dispersed across the United States, to conduct measurements to determine the national estimates of actual broadband speed performance per ISP for 54 services tiers, and a reliable estimation of performance by ISP packages within geographical region was conducted. Ten annual broadband speed measurements have been collected through hardware devices placed in consumer panelists' homes in March 2011, April 2012, September 2012, and September 2013, September 2014, September 2015 and September 2016, September 2017, September 2018 and September 2019.

ISPs included in the study are:

- Optimum
- CenturyLink
- Charter
- Cincinnati Bell
- Comcast
- Cox
- Frontier DSL
- Frontier Fiber

- Hawaiian Telecom
- Mediacom
- Verizon DSL
- Verizon Fiber
- Windstream

A significant proportion of volunteers were recruited via an initial public relations and social media campaign led by the FCC. This included discussion on the FCC website and on technology blogs, as well as articles in the press regarding the study. The demographics of this initial panel were reviewed to identify any deficiencies with regard to the sample plan described above. These goals were set to produce statistically valid sets of volunteers for demographics based on ISP, speed tier, technology type, and region. This initial pool of volunteers was then supplemented by the participating ISPs, who sent out an email to customers in desired demographics that were underrepresented in the pool of publicly solicited volunteers. Emails directed interested volunteers to contact SamKnows regarding participation in the trial. At no time during this recruitment process did the ISPs have any knowledge regarding which of their customers might be participating in the trial. In almost all cases, ISP engagement in soliciting volunteers enabled us to meet desired demographic targets.

Note that broadband speed performance may be considered to correlate with:

- Distance between premises and exchange,
- Contention in the ISP's own network particularly at peak time,
- Location density, such as rural or urban; typically in urban areas there is greater availability of higher speed services. Additionally, in rural areas the average line length from local exchange to premises is longer than in urban areas.
- The technology used. There are different technologies available to delivered broadband services.

Therefore, although this sample is based on pool of volunteers, it is unlikely that this particular sample would be different from a random sample since SamKnows control the region, U/S/R split and services tiers so as to get a broad representation of consumer experience. The individual panelist behavior has no impact on the measured broadband performance since the methodology used autonomously performs testing at random times but only when panelist activity is de minimis.

Strata Definition and Allocation

The panel of U.S. broadband subscribers was drawn from a pool of over 200,000 volunteers following an ongoing recruitment campaign that ran from May 2010 through October 2019.

The volunteer sample was organized with a goal of covering major ISPs in the 48 contiguous states across five broadband technologies: DSL, cable, fiber-to-the-home, fixed terrestrial wireless, and satellite.¹

Target numbers for volunteers were also set across the four Census Regions—Northeast, Midwest, South, and West—to help ensure geographic diversity in the volunteer panel and compensate for network variances across the United States.

Region breakdowns are:

Northeast Region (including the New England division): Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont; and the Middle Atlantic division: New Jersey, New York, and Pennsylvania.

Midwest Region (including the East North Central division): Illinois, Indiana, Michigan, Ohio, and Wisconsin; and the West North Central division: Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota.

South Region (including the South Atlantic division): Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia; the East South Central division: Alabama, Kentucky, Mississippi, and Tennessee; and the West South Central division: Arkansas, Louisiana, Oklahoma, and Texas.

West Region (including the Mountain division): Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming; and the Pacific division: Alaska, California, Hawaii, Oregon, and Washington.

Data collected in September 2019

For the fixed line project only, a total of 14,428,136,399 measurements were taken across 155,520,975 unique tests from the Whiteboxes deployed.

¹ Recently, at the request of the state of Hawaii, panelists from Hawaii have been included in the program and we expect the 2018 report to include these panelists.

2. Describe the procedures for the collection of information including:

- *Statistical methodology for stratification and sample selection,*
- *Estimation procedure,*
- *Degree of accuracy needed for the purpose described in the justification,*
- *Unusual problems requiring specialized sampling procedures, and*
- *Any use of periodic (less frequent than annual) data collection cycles to reduce burden.*

Building on its experience completing a similar fixed broadband performance project for the United Kingdom telecommunications regulator, Ofcom, SamKnows originally recruited a US-based panel and deployed its technology throughout all 48 states to cover all regions. Sample quotas were set for 4 regions, West, Midwest, Northeast and South. The quotas, defined by geography, technology and service level, have been named buckets. Within each bucket, a soft quota was defined by population density using Census Bureau definitions.

SamKnows best practices in panel recruitment developed in its global efforts have been reviewed by 3rd parties during the course of the project². The panel recruitment consisted of two steps, using a multi mode recruitment effort to build a large pool of volunteers. Volunteers were encouraged to sign-up to a project specific website (www.measuringbroadbandamerica.com), providing a panel of volunteers. This was supplemented by more targeted participants who were recruited with the support of the ISPs, who targeted specific customer groups with mailings. The approach provides flexibility in building the panel to ensure geographical representation, excluding outliers and, as much as possible, maximizing the multi mode recruitment selection to avoid pitfalls such as selection bias, coverage error, and panel attrition. Maintenance of panelists is provided on an ongoing basis through email recruitment of volunteers with targeted demographics.

SamKnows provides report data when deemed accurate enough to be useful. Accuracy is reflected by sample size and variance. In order to reflect the accuracy of the sample SamKnows shows the speed results as an interval using the 95% interval around the mean. With a sample of approximately 50 national wide panelists for a tier, the associated sample error has been found to typically be within $\pm 4\%$.

Estimates for sub groups – including buckets – are subject to higher sampling error margins based strictly on the smaller number of panelists in a specific subgroup.

Collation of Results and Outlier Control

All measurement data were collated and stored for analysis purposes as monthly trimmed averages during three time intervals (24 hours, 7:00 pm to 11:00 pm local time Monday through Friday, 12:00 am to 12:00 am local time Saturday and Sunday). Only participants who provided a minimum of 5 days of valid measurements and had valid data in each of the three time intervals were included in the September 2019 test results. In addition, the top and bottom 1 percent of measurements were dropped to control for outliers that may have been anomalous or

² Statistical experts from both the FCC and the surveyed ISPs were involved in the initial establishment of the survey methodology. In 2013 the entire program was reviewed by the Institute for Advanced Analytics at North Carolina State University as to methodology and accuracy.

otherwise not representative of actual broadband performance. All statistics were computed on the trimmed data. The resulting final sample of data for September/October 2019 was collected from 6,006 participants.

Consistency of Speed Measurements

In addition to reporting on the weighted mean of medians speed of panelists, the MBA Report also provides a measure of the consistency of speed that panelists experience in each tier³. For purposes of discussion we use the term “80/80 consistent speed” to refer to the minimum speed that was experienced by at least 80% of panelists for at least 80% of the time. The process used in defining this metric for a specific ISP tier is to take each panelist’s set of download or upload speed data during the peak period across all the days of the validated measurement period and arrange it in increasing order. The speed that corresponds to the 20th percentile represents the minimum speed that the panelist experienced at least 80% of the time. The 20 percentile values of all the panelists on a specific tier are then arranged in an increasing order. The speed that corresponds to the 20th percentile now represents the minimum speed that at least 80% of panelists experienced 80% of the time. This is the value reported as the 80/80 consistent speed for that ISP’s tier. We also report on the 70/70 consistent speed for an ISP’s tier, which is the minimum speed that at least 70% of the panelists experience at least 70% of the time. We typically report the 70/70 and the 80/80 consistent speeds as a percentage of the advertised speed.

When reporting on these values for an ISP, we effectively weigh the 80/80 or 70/70 consistent speed results (as a percentage of the advertised speed) of each of the ISP’s tier based on the number of subscribers to that tier; so as to get a weighted average across all the tiers for that ISP.

Mobile Broadband Measurement Statistical Methods

In 2012, the FCC announced its intent to expand the existing program to include measurement of mobile broadband performance using a crowd-sourced application deployed on Android smartphone devices. In a four-month privacy by design process, the FCC developed a privacy policy to guide its data collection and Open Data policies. The process brought together privacy experts, government agencies, academics, manufacturers, carriers, public interest and other interested stakeholders to address technical, legal and other policy concerns resulting in a completely anonymized data collection process that minimized risk of any individual being able to be identified from the pool of anonymous performance data.

The FCC instituted a policy to collect anonymous data and make no public use without first undergoing a technical privacy review to delete or process datum to minimize re-identification risks to participant volunteers submitting data. The technical privacy review will employ clustering and other cutting edge statistical techniques in analyzing the bulk raw collected data. The privacy review will produce recommended business rules for the deletion of particular

³ This statistic was implemented upon a recommendation by the Institute for Advanced Analytics.

results or the coarsening of data elements located in areas limited total number of samples for the defined geographic area. The review will also provide recommendations for allowable combinations of data elements for data sets to be released to the public. The technical privacy review will be performed after data collection and the statistical analysis will make recommendations that take into account the total number of samples collected as well as the variance found in the collected data.

The FCC released the FCC Speed Test App for Android in November 2013, and the FCC Speed Test App for iPhone in February 2014. The smartphone measurement application is designed to be installed on a user's smartphone either directly or via an app store, such as Google Play or iTunes. The application is free to download, but carrier charges may apply. The Android and iOS application will periodically perform measurements, either when initiated by the user or in the case of Android via automated testing. The automated testing function for Android can be disabled and the app can be configured to start a test only when manually executed, but due to limitation of the operating system iOS devices do not have automated testing capability and can only execute the speed test manually. Android devices also support collecting more information about cellular performance that is not supported by iOS devices.

3. *Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.*

Maximizing Response Rates in Fixed Test Panel

Once the fixed volunteer panel was recruited and the SamKnows ‘Whiteboxes’ deployed, the tests run on a pre-configured timetable, subject only to changes in the schedule as agreed upon by the Measuring Broadband America collaborative (which includes FCC members, SamKnows, ISPs, researchers and other interested parties) and the volunteers own use of their broadband connection. Therefore, ‘response rates’ will always be at a ‘maximum’. Details on the proprietary framework are provided below:

Panel Recruitment

A significant proportion of volunteers were recruited via an initial public relations and social media campaign led by the FCC. This included discussion on the FCC website and on technology blogs, as well as articles in the press regarding the study. The demographics of this initial panel were reviewed to identify any deficiencies with regard to the sample plan described above. These goals were set to produce statistically valid sets of volunteers for demographics based on ISP, speed tier, technology type, and region. This initial pool of volunteers was then supplemented by the participating ISPs, who sent out an email to customers in desired demographics that were underrepresented in the pool of publicly solicited volunteers. Emails directed interested volunteers to contact SamKnows regarding participation in the trial. At no time during this recruitment process did the ISPs have any knowledge regarding which of their customers might be participating in the trial. In almost all cases, ISP engagement in soliciting volunteers enabled us to meet desired demographic targets.

The mix of panelists recruited using the above methodologies varied by ISP.

A multi-mode strategy was used to qualify volunteers for this trial. The key stages of this process were as follows:

1. Volunteers were directed to complete an online form, which provided information on the study and required volunteers to submit a small amount of information, which was used to track subsequent submissions by these volunteers.
2. Volunteers were selected from respondents to this follow-up email based on the statistical requirements of the panel. Selected volunteers were then asked to complete the User Terms and Conditions that outlined the permissions to be granted by the volunteer in key areas such as privacy.
3. Of those volunteers that completed the User Terms and Conditions, SamKnows selected the first panel of 13,000 participants, each of whom received a Whitebox for self-installation.

SamKnows provide full support during the Whitebox installation phase.

Test Scheduling

Tests for download and upload speeds will be run every hour during peak time and once every 6 hours outside of this, on a randomized timing in the hour - 24 hours a day, 7 days a week. Many other tests run on the MBA panel including latency, packet loss, jitter, DNS resolution and web browsing amongst others. Within a month period, the software within the Whitebox unit will perform over 700 separate speed tests. The unit is installed immediately in front of their home internet connection. This ensures that tests can be run at any time, even if all home computers are switched off.

Testing Locking

No two tests may operate simultaneously. The first test to begin will have priority, and the other will block until the first finishes (or a timeout is reached).

Threshold Manager

A key benefit to the Whitebox / CPE approach is the fact that 'cross-traffic' (other traffic in the End User's home) can be accounted for. This means we can avoid running tests when the user is using their connection, resulting in (a) cleaner results for us and (b) a happy End User (because their use of the Internet is not being interrupted).

End Users are instructed to connect their wired devices via the Whitebox and leave their wireless devices unchanged.

Prior to and between tests, a threshold manager service monitors the inbound and outbound traffic across the WAN interface of the Whitebox to calculate if a panelist is actively using the Internet connection. The threshold for traffic is set to 64 kbps downstream and 32 kbps upstream. If these thresholds are breached prior to the test starting or between tests, the test will be delayed for a minute and the process repeated. If the connection is being actively used throughout, this pause and retry process will occur up to 5 times before the entire test cycle is abandoned.

A similar process is performed for wireless clients. Wireless users are not asked to make any changes. As with the wired approach, measurements are not conducted when there is wireless activity detected. Wireless activity is determined by passively monitoring the traffic from the user's wireless SSID(s). There are two techniques used to determine the user's wireless SSID:

Perform a scan for wireless networks near the Whitebox. Search for an access point that has a MAC address adjacent to the MAC of the LAN interface on the volunteer's CPE. In a user's home environment, this is typically a combined modem/router/WAP. This takes advantage of the fact that most CPE use similar MACs on their Ethernet and WiFi interfaces. This provides significantly improved confidence in high density wireless environments (like apartment blocks).

Where no adjacent wireless MAC is found, the Whitebox falls back to the old approach of choosing the device with the strongest signal, whereby the SamKnows Whitebox passively monitors the strongest nearby wireless network for traffic.

Once an SSID has been identified, the Whitebox passively monitors all traffic that the SSID exchanges and records volume information. Note that it does not matter if the wireless network is encrypted; the Whitebox does not need to join the wireless network, it simply cares about volumes of data (it makes a conservative assumption that all wireless traffic is destined for the Internet).

If a wireless AP is broadcasting multiple SSIDs on the same channel, then the Whitebox will catch traffic from all, because they will use the same or adjacent MAC addresses. It does not matter if the user has hidden their SSID or encrypted their wireless network; the Whiteboxes are simply passively monitoring packet volume and do not need access to the data contained within the packets.

The wireless monitoring process described above is repeated in both the 2.4GHz and 5GHz channels, for applicable Whitebox models.

Order of Tests

The following tests are run:

- Download speed
- Upload speed
- Web browsing
- UDP latency
- UDP packet loss
- Video streaming (Netflix / YouTube / Hulu)
- Voice over IP - Upstream packet loss, downstream packet loss, upstream jitter, downstream jitter, round trip latency
- DNS resolution Time
- DNS failures
- ICMP latency
- ICMP packet loss
- Latency under load

Required number of data points

For reporting purposes, the data will be aggregated first per hour for the reporting period and then overall to minimize the effect of missing data. If less than 5 data points are recorded for one time slot within a month, the data will be discarded.

Tests

Download and upload speed

Measures the download and upload speed of the broadband connection in bits per second. The transfer is conducted over one or more concurrent HTTP connections (using the GET verb of download and the POST verb for uploads).

In the download speed test the client will fetch a portion of a 1GB binary (non-zero, randomly generated) payload hosted on an HTTP server on the target test server. The content is discarded as soon as it is received.

In the upload test the client will generate the payload itself (using /dev/urandom as a non-blocking source of random content) to send to the server. The measure of throughput may be optionally carried out on the server side (the receiver) in the upload test.

The speed tests (both download and upload) operate for either a fixed-duration (specified in seconds) or a fixed-volume (specified in MB). Where possible, a fixed-duration test is preferred as it will cater well for all broadband access speeds. However, a fixed-volume test may be necessary where predictability of bandwidth usage is desired.

Four separate variations of the test are supported:

- Single TCP connection download speed test
- Multiple TCP connection download speed test
- Single TCP connection upload speed test
- Multiple TCP connection upload speed test

For multiple TCP connection tests we typically recommend that three concurrent connections are used. In some cases (e.g. where the round-trip time between client and server is very high) it may be necessary to increase this.

Factors such as TCP slow start are accounted for through the use of a “warm-up” period. This period begins as soon as the test starts and seeks to establish that the throughput has reached stable rate before starting the real test (which will continue over the same TCP connection(s)). It is important to note that the data transferred in the warm-up period is excluded from the main test results, but it is still recorded separately as a supplementary metric.

The speed test client will record the throughput, bytes transferred and time taken at the end of the test. It may also record these values at multiple intervals during the test. This is commonly used to help characterise the difference between ‘burst’ and ‘sustained’ throughput (where transfer speeds may be inflated at the start of a TCP connection).

Web-browsing test

Measures the time taken to fetch the HTML and referenced resources from a page of a popular website. This test does not test against centralised testing nodes; instead it tests against real websites, allowing for content distribution networks and other performance enhancing factors to be considered.

Each Whitebox will test ten common websites on every test run. The time taken to download the resources, the number of bytes transferred and the calculated rate per second will be recorded. The primary measure for this test is the total time taken to download the HTML page and all associated images, Javascript and stylesheet resources.

The results include the time taken for DNS resolution. The test uses up to eight concurrent TCP connections to fetch resources from targets. The test pools TCP connections and utilises persistent connections where the remote HTTP server supports them.

The test may optionally run with or without HTTP headers advertising cache support (through the inclusion or exclusion of the “Cache-Control: no-cache” request header). The client advertises the user agent of Microsoft Internet Explorer 10.

VoIP

The Voice over IP (VoIP) measures the quality of a voice call between the client and a nearby test server. It is intended to characterise how suitable the user’s broadband connection is for placing and receiving VoIP calls.

The test uses the SIP protocol for signalling and can support a variety of codecs for the call itself. By default, the test uses the G.711 codec, which uses a bi-directional stream of 64kbps UDP traffic. Other codecs are supported, including G.722. The duration of the test is configurable, but will run for 10 seconds by default.

The test may use a SamKnows-provided SIP server for signalling, or may alternatively be configured to use a third-party SIP server. This allows for third-party SIP services to be tested on an end-to-end basis, assessing the entire VoIP infrastructure, not just the quality of the end user broadband network.

The test captures the following metrics:

- Round trip latency
- Downstream packet loss
- Upstream packet loss
- Downstream jitter
- Upstream jitter
- MOS (Mean Opinion Score)

UDP latency and packet loss

Measures the round trip time of small UDP packets between the Whitebox and a target test server. Each packet consists of an 8-byte sequence number and an 8-byte timestamp. If a packet is not received back within two seconds of sending, it is treated as lost. The test records the number of packets sent each hour, the average round trip time of these and the total number of packets lost. The test will use the 99th percentile when calculating the summarised minimum, maximum and average results on the Whitebox.

As with the availability test, the test operates continuously in the background. It is configured to randomly distribute the sending of the echo requests over a fixed interval, reporting the summarised results once the interval has elapsed.

Typically 600 samples are taken per hour, distributed throughout the hour. If the line is busy then fewer samples will be taken. A higher sampling rate may be used if desired.

UDP contiguous loss (disconnections)

This test is an optional extension to the UDP Latency/Loss test. It records instances when two or more consecutive packets are lost to the same test server. Alongside each event we record the timestamp, the number of packets lost and the duration of the event.

By executing the test against multiple diverse servers, a user can begin to observe server outages (when multiple probes see disconnection events to the same server simultaneously) and disconnections of the user's home connection (when a single probe loses connectivity to all servers simultaneously).

Typically, this test is accompanied by a significant increase in the sampling frequency of the UDP Latency/Loss client to ~2000 packets per hour (providing a resolution of 2-4 seconds for disconnection events).

DNS resolution and failure rate

This test measures the DNS resolution time of a selection of common website domain names. These tests will be targeted directly at the ISPs recursive resolvers. A list of appropriate servers will be sought from each ISP in advance of the tests.

YouTube measure

The YouTube test is an application-specific test, supporting the streaming of video and audio content from YouTube using their protocols and codecs.

The test begins by seeking out the most popular video in the user's country. This is achieved by fetching a list of the most popular YouTube videos from a central SamKnows server. The central list of videos is refreshed once every 12 hours using the YouTube API. We filter for videos that are at least 60 seconds in length and have an HD quality variant. Note that by interacting with the YouTube API from a central location we can ensure that every probe is delivered the same list of videos.

The test running on the probe will now fetch the YouTube web page for the most popular video, and parse the Javascript contained within the page. Within this Javascript is held a list of all the encodings of the video in question and the content server hostname. By making this request from the probe we ensure that the test is receiving the same content server as the user would if they were using a desktop computer on the same connection.

The test will then connect to the content server (using whatever server YouTube would normally direct a real client on the same connection to) and begins streaming the video and audio. MPEG4, WebM, Dash (adaptive) and Flash video codecs are supported. Although the adaptive codec is supported, the test does not actually adapt its rate; we stream at full rate all the time, which provides for reproducibility.

The test parses video frames as it goes, capturing the timestamp contained within each video frame. After each frame, we sample how much realtime has elapsed versus video time. If video time > realtime at a sample period, then an underrun has not occurred. Otherwise, one has occurred.

The test downloads 10 seconds of audio and video at a time, with a buffer of 40 seconds. So, on startup, the test will immediately download (at full speed) 40 seconds of video and audio, and will then download more as required, keeping the 40 second playback buffer full. By default, the test will run for a fixed duration of 20 seconds of realtime.

In its default mode of operation, the test will capture the ‘bitrate that can be reliably streamed’ on the user’s connection. This is achieved through the following process:

1. Find the fastest recent speedtest result that the probe has completed.
2. As described above, fetch the list of YouTube videos, find the most popular one, and then select the highest bitrate encoding which is less than the fastest speedtest result found in step 1.
3. Attempt to stream this video, for a fixed duration of 20 seconds of realtime. If successful, then the “bitrate reliably streamed” for this instance is the bitrate that we just fetched.
4. However, if a stall event occurs, then we immediately abort the test and retry at the next lower bitrate.
5. If we find a bitrate that we can stream without a stall event occurring, then that bitrate is our “bitrate reliably streamed” for this instance.
6. However, if we encounter stalls for every bitrate, then the “bitrate reliably streamed” is zero.

The key outputs from this metric are:

- The bitrate reliably streamed
- The startup delay (the time taken to download two seconds of video)
- The TCP connection time
- The number of stalls and their duration (this is only applicable if the test is not running in the ‘bitrate reliably streamed’ mode)

Netflix measure

The Netflix test is an application-specific test, supporting the streaming of binary data from Netflix’s servers using the same CDN selection logic as their real client uses. The test has been developed with direct cooperation with Netflix.

The test begins by calling a Netflix hosted web-based API. This API examines the client’s source IP address and uses the existing proprietary internal Netflix logic to determine which Netflix server this user’s IP address would normally be served content from. This logic will consider the ISP and geographic location of the requesting IP address. Where the ISP participates in Netflix’s Open Connect programme, it is likely that one of these servers will be used. The API will return to the client a HTTP 302 redirect to a 25MB binary file hosted on the applicable content server.

The test will then establish an HTTP connection to the returned server and attempt to fetch the 25MB binary file. This runs for a fixed 20 seconds of realtime. HTTP pipelining is used to request multiple copies of the 25MB binary, ensuring that if the payload is exhausted before the 20 seconds are complete, we can continue receiving more data. The client downloads data at full rate throughout; there is no client-side throttling taking place.

It's important to note that this 25MB binary content does not contain video or audio; it is just random binary data. However, with knowledge of the bitrates that Netflix streams content at, we can treat the binary as if it were video/audio content operating at a fixed rate. This allows us to determine the amount of data consumed for each frame of video (at a set bitrate) and the duration that it represents. Using this, we then can infer when a stall occurred (by examining when our simulated video stream has fallen behind realtime). The test currently simulates videos at bitrates of 235Kbps, 375Kbps, 560Kbps, 750Kbps, 1050Kbps, 1750Kbps, 2350Kbps, 3000Kbps, 4500Kbps, 6000Kbps and 15600Kbps.

This approach also allows us to derive the 'bitrate reliably streamed', using the same methodology as the YouTube test described above. A small difference here is that we do not need to restart the download at a lower bitrate if a stall is encountered; because the incoming stream of binary data is decoded at a simulated bitrate, we can simply recompute the playback characteristics of the same network stream at a different bitrate entirely on the client side. This simply means that the test uses a predictable amount of bandwidth, even in cases where stalls occur.

The key outputs from this metric are:

- The bitrate reliably streamed
- The startup delay (the time taken to download two seconds of video)
- The TCP connection time
- The number of stalls and their duration
- The downstream throughput achieved

Hulu Measure

Supported clients: Whiteboxes, Routers

The Hulu test is another application specific test that makes use of a third parties live infrastructure, ensuring that the test is mimicking the real behavior of applications. The streams a reference video from Hulu's CDN and measures key metrics related to its performance. SamKnows has partnered with Hulu to facilitate these measurements; Hulu has placed a reference video on its infrastructure and provided SamKnows with the ability to stream this video from any source.

At the time of writing, Hulu makes use of two CDNs – Level3 and Edgecast. The test uses these same CDNs and will randomly choose between one of these two CDNs at startup. The test makes use of the same DNS names that the live Hulu service uses, ensuring that any CDN redirection logic put in place by the CDN operator, Hulu or the ISP is followed, just like a real client would use.

Hulu has placed a reference video on their two CDNs, transcoded at the same bitrates that Hulu clients typically use. This video is encoded using H264 at the following bitrates: 64kbps, 200kbps, 400kbps, 650kbps, 1Mbps, 1.5Mbps, 2Mbps and 3.2Mbps. The URL of the video encoded at each bitrate, on each CDN, is hard-coded into the Hulu test client.

At startup, the test randomly chooses the CDN to use, and begins to fetch video content using the list of URLs stored in the client. The test parses video frames as it goes, capturing the timestamp contained within each frame. After each frame we sample how much real-time has elapsed versus video time. If video time > real-time at a sample period, then an underrun has not occurred. Otherwise, one has occurred.

As with the other video tests, the client is configured to start testing at the highest supportable bitrate and then step down if and when stalls occur. This allows us to identify the 'bitrate reliably streamed'. Optionally the test may be instructed to continue at a fixed bitrate and simply record the number and duration of underrun events.

The key outputs from this metric are:

- The bitrate reliably streamed
- The startup delay (the time taken to download two seconds of video)
- The TCP connection time
- The number of stalls and their duration (this is only applicable if the test is not running in the 'bitrate reliably streamed' mode)
- The downstream throughput achieved

4. Describe any tests or procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more

respondents. A proposed test or set of test may be submitted for approval separately or in combination with the main collection of information.

The SamKnows tests and methodology have been evaluated and approved by Government Regulators, Internet Service Providers, leading academics and researchers.

The stated service level name/identifier, headline speed, price and download limit will be used to check and assign panelist to the correct broadband packages. Different tests and checks will be held to confirm ISP and flag incorrect packages in the following manner:

At the recruitment stage

1. The ISP allocation will be validated using customer's IP against those held in ISP IP libraries. If inconsistent the record is flagged
2. Actual speed will be validated in relation to package and distance from premises to exchange distribution. Outliers will be excluded
3. Exclusion of household when distance to exchange is untypical to maximize the effective sample size after normalization

Test on continuous basis

1. Panelists have the ability to update their information through SamKnows web interface
2. Automatic flag of change of ISP and geographical location through the reporting system
3. Automatic flag of untypical maximum download speed

Data Integrity

As the Whiteboxes ran tests consistently from homes across the United States, it was important to check the data to ensure that any anomalies were removed. To ensure the integrity of the large amount of data collected, the following protocols were developed:

1. Change of ISP intra-month: found units that changed ISP intra-month (determined by performing daily WHOIS query using the panelist's IP address), and removed data for the ISP on which they spent less time over the course of that month.
2. Change of service tier intra-month: found units that changed service tier intra-month by isolating the difference between the average sustained throughput observed for the first three days in the reporting period from the average sustained throughput observed for the final three days in the reporting period. If a unit was not online at the start or end of that period, then the first/final three days that they were actually online were taken. If this difference was over 50 percent, the downstream and upstream charts for this unit were individually reviewed. Where an obvious step change was observed (e.g., from 1 Mbps to 3 Mbps), the data for the shorter period was flagged for removal.
3. Removal of any failed or irrelevant tests: removed any failed or irrelevant tests by removing measurements against any non-M-Lab or Level 3 servers (to catch tests to ISP

test nodes). Removed measurements against any M-Lab or Level 3 server outside of the United States. Removed measurements against any M-Lab or Level 3 server that exhibited greater than or equal to 10 percent failures in a specific one hour period (the purpose was to remove periods where M-Lab or Level 3 servers were unavailable).

4. Removal of any problem units: removed measurements for any unit that exhibited greater than or equal to 10 percent failures in a particular one hour period (the purpose was to remove periods where units were unable to reach the Internet).

Independent Review: As there are a large number of stakeholders in this project, SamKnows has dedicated resource to liaising with interested 3rd parties who wish to participate in validating/enhancing the SamKnows methodology. These discussions are open to anyone who wishes to participate.

The mobile program has employed the same collaborative review process with diverse mobile broadband stakeholders in securing evaluation and approval of the collection approaches.

The data collected in this program are made available as OpenData for review and use by the public. Raw and processed data sets, and the methodologies used to process and analyze data are freely and publicly available. Researchers and developers interested in working with measurement data in raw form will need skills in database management, SQL programming, and statistics, depending on the analysis. A develop FAQ for database configuration and data importing instructions for mysql and postgresql are available. <http://www.fcc.gov/measuring-broadband-america/2012/database-setup-and-importingmeasuring-broadband-america-data-april-2012>

5. Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

Martin Doczkat
Federal Communications Commission
Email: martin.doczkat@fcc.gov
Phone: 202-418-2435

Rajender Razdan
Federal Communications Commission
Email: rajender.razdan@fcc.gov
Phone: 202-418-0879

Roxanne Robinson

Director, Government Projects, SamKnows
Email: roxanne@samknows.com
Phone: +44 (0) 203 111 4343