



Measuring Broadband New Zealand

Autumn Report, May 2020

In 2018, the Commerce Commission appointed SamKnows to measure New Zealand’s internet performance. The programme, called Measuring Broadband New Zealand, gives internet users in New Zealand access to SamKnows Whiteboxes to measure the quality of their fixed-line internet. The aim of the programme is to increase transparency about actual in-home broadband performance and provide consumers with independent information about internet performance across different providers, plans, and technologies, to help them choose the best broadband for their homes. It will also encourage providers to improve and compete on their performance. This report provides an overview of the findings from data collected during February 2020 with some additional analysis on March and April 2020.



Raw Data

Alongside the report ComCom also release the raw data and summary unit information used to produce the Autumn 2020 report.

Two levels of data are included in this publication: **Raw Measurement Data** and **Per-Whitebox Summary Data**. More information on what is included in these files is outlined below.

Raw Measurement Data

This is the measurement data in its raw, unaggregated form. Only measurements that were used within the report have been included in this raw data package. Additionally, metadata fields which were not used in the report have been excluded (e.g. RSP name and product in specific instances).

The raw data is available in the ‘./raw_data’ directory, and a data dictionary describing the fields is included later in this document.

Per-Whitebox Summary Data

The measurements in the raw data are aggregated by Whitebox ID (also known as unit_id) as part of the data analysis process. The per-Whitebox data is far smaller, and therefore more accessible to third parties, than the raw data. It also includes additional derived fields which are used later in the analysis (e.g. the fraction of Netflix videos that were delivered at HD or better quality).

This summary data is calculated from the raw data using the statistical analysis tool R. The eventual aim is to release the R script used to create the charts along with the raw data so that interested parties can recreate the results. Due to the fact that certain metadata fields are excluded in the raw data (e.g. the advertised speeds for each RSP’s Fibre Max product is not shared, as this could allow analysis on Fibre Max by RSP which is not currently reported on), the R script is not able to run.

A data dictionary describing the fields is included later in this document.

File listing

File	Description
./raw_data	
raw_download_tests.csv	Download speed test data.
raw_upload_tests.csv	Upload speed test data.
raw_latency_tests.csv	Latency and packet loss data.
raw_netflix_tests.csv	Netflix data.
raw_gaming_tests_exc_fortnite.csv	Gaming data excluding Fortnite.
raw_fortnite_tests.csv	Fortnite data.
raw_social_media_tests.csv	Social Media data.
./output	
report_charts.csv	Data behind the graphs which appear in the Autumn report.
unit_summary_statistics_download_upload_latency.csv	One line per Whitebox with download, upload and latency results
unit_summary_statistics_gaming.csv	One line per Whitebox per game.
unit_summary_statistics_netflix.csv	One line per Whitebox with Netflix results.
unit_summary_statistics_social_media.csv	One line per Whitebox per service per media type.

Raw Data

raw_download_tests.csv (Download speed)

Field Name	Type	Description
unit_id	int	Unique identifier for an individual Whitebox.
dtype	datetime	The time of the test (UTC).
ddate	date	The date of the test.
target	string	Hostname of the test server.
download_mbps	decimal	Test speed in Mbps.
did_test_complete_successfully	boolean	Did the speed test complete successfully?
target_server_country	string	The country in which the test server is located.
is_during_peak_hour	boolean	Is the test in peak hour (7-11pm Mon - Fri)?
is_during_reporting_period	boolean	Is the test during the reporting period?
time_period	string	The time period the test was run in.

raw_upload_tests.csv (Upload speed)

Field Name	Type	Description
unit_id	int	Unique identifier for an individual Whitebox.
dtype	datetime	The time of the test (UTC).
ddate	date	The date of the test.
target	string	Hostname of the test server.
upload_mbps	decimal	Test speed in Mbps.
did_test_complete_successfull	boolean	Did the speed test complete successfully?
target_server_country	string	The country in which the test server is located.
is_during_peak_hour	boolean	Is the test in peak hour (7-11pm Mon - Fri)?
is_during_reporting_period	boolean	Is the test during the reporting period?
time_period	string	The time period the test was run in.

raw_latency_tests.csv (Latency & Packet Loss)

Field Name	Type	Description
unit_id	int	Unique identifier for an individual Whitebox.
dtype	datetime	The time of the test (UTC).
ddate	date	The date of the test.
target	string	Hostname of the test server.
latency_ms	decimal	The time for a round trip from Whitebox -> Server -> Whitebox.
num_successes	int	Number of packets which made a successful round trip.
num_failures	int	Number of packets which failed to make a round trip.
packet_loss_pct	decimal	Ratio of packets which did not make a successful round trip: failures
target_server_country	string	The country in which the test server is located.
is_during_peak_hour	boolean	Is the test in peak hour (7-11pm Mon - Fri)?
is_during_reporting_period	boolean	Is the test during the reporting period?
time_period	string	The time period the test was run in.

Raw Data

raw_netflix_tests.csv (Netflix)

Field Name	Type	Description
unit_id	int	Unique identifier for an individual Whitebox.
dtime	datetime	The time of the test (UTC).
ddate	date	The date of the test.
target	string	Hostname of the server assigned by Netflix to stream content.
bitrate_mbps	decimal	The bitrate that can be reliably streamed without stalls (in Mbps).
video_quality	string	Quality of video that can be achieved during test.
download_mbps	decimal	The download speed when downloading content from Netflix (in Mbps).
latency_ms	decimal	The time taken to establish a connection with Netflix (in milliseconds). Used as a proxy for the latency between Whitebox and Netflix server.
did_test_encounter_stall	boolean	Did the test encounter a stall event?
did_test_complete_successfully	boolean	Did the test complete successfully?
is_during_peak_hour	boolean	Is the test in peak hour (7-11pm Mon - Fri)?
is_during_reporting_period	boolean	Is the test during the reporting period?
time_period	string	The time period the test was run in.

raw_social_media_tests.csv (Social Media)

Field Name	Type	Description
unit_id	int	Unique identifier for an individual Whitebox.
dtime	datetime	The time of the test (UTC).
ddate	date	The date of the test.
service	string	Currently one of Facebook-app, Facebook-messenger, Instagram-app, Instagram-messenger, Snapchat, Whatsapp, Twitter
media	string	Currently one of Image, Text, Audio, Video
direction	string	One of "Downlink" or "Uplink". Indicates whether we're measuring receiving or uploading content.
hop_count	int	The number of hops to the server
latency_ms	decimal	The time for a round trip from Whitebox -> Server -> Whitebox
did_test_complete_successfully	boolean	TRUE if any packets made a successful round trip, FALSE if not.
num_successes	int	Number of packets which made a successful round trip.
num_failures	int	Number of packets which failed to make a successful round trip.
is_during_peak_hour	boolean	Is the test in peak hour (7-11pm Mon - Fri)?
is_during_reporting_period	boolean	Is the test during the reporting period?
time_period	string	The time period the test was run in.

Raw Data

raw_fortnite_tests.csv

Field Name	Type	Description
unit_id	int	Unique identifier for an individual Whitebox.
dtime	datetime	The time of the test (UTC).
ddate	date	The date of the test.
region	string	The region of the Fortnite server, as defined by the game author
datacenter	string	The datacenter code (e.g. "SYD") of the Fortnite server, as defined by the game author
selected_server	int	The IP address of the server that we measured against
hop_count	int	The number of hops to the server
latency_ms	decimal	The time for a round trip from Whitebox -> Server -> Whitebox
did_test_complete_successfully	boolean	TRUE if any packets made a successful round trip, FALSE if not.
num_successes	int	Number of packets which made a successful round trip.
num_failures	int	Number of packets which failed to make a successful round trip.
is_during_peak_hour	boolean	Is the test in peak hour (7-11pm Mon - Fri)?
is_during_reporting_period	boolean	Is the test during the reporting period?
time_period	string	The time period the test was run in.

raw_gaming_test_exc_fortnite.csv

Field Name	Type	Description
unit_id	int	Unique identifier for an individual Whitebox.
dtime	datetime	The time of the test (UTC).
ddate	date	The date of the test.
target	string	Target hostname or IP address
ping_latency_ms	decimal	The time for a round trip from Whitebox -> Server -> Whitebox
num_successes	int	The number of pings successful sent and received.
num_failures	int	The number of pings which failed.
packet_loss_pct	decimal	The proportion of pings which failed.
target_server_grouping	string	The gaming service being tested (e.g League of Legends, Playstation)
is_during_peak_hour	boolean	Is the test in peak hour (7-11pm Mon - Fri)?
is_during_reporting_period	boolean	Is the test during the reporting period?
time_period	string	The time period the test was run in.

Raw Data

unit_summary_statistics_download_upload_latency.csv

Field Name	Type	Description
unit_id	int	Unique identifier for an individual Whitebox.
target_server_country	string	The country in which the test server is located.
trimmed_mean_download_mbps_24h	decimal	The 1% trimmed mean (average of the middle 98% of data) of download_mbps. Results where download_samples_24h is less than 5 are removed from the final dataset.
trimmed_mean_download_mbps_peak	decimal	The 1% trimmed mean (average of the middle 98% of data) of download_mbps - only considering tests during peak hours i.e. where is_during_peak_hour is TRUE. Results where download_samples_peak is less than 5 are removed from the final dataset.
download_samples_24h	int	The number of download tests (count of rows in raw_download_tests.csv).
download_samples_peak	int	The number of download tests (count of rows in raw_download_tests.csv) - only considering tests during peak hours i.e. where is_during_peak_hour is TRUE.
trimmed_mean_upload_mbps_24h	decimal	The 1% trimmed mean (average of the middle 98% of data) of upload_mbps. Results where upload_samples_24h is less than 5 are removed from the final dataset.
trimmed_mean_upload_mbps_peak	decimal	The 1% trimmed mean (average of the middle 98% of data) of upload_mbps - only considering tests during peak hours i.e. where is_during_peak_hour is TRUE. Results where upload_samples_peak is less than 5 are removed from the final dataset.
upload_samples_24h	int	The number of upload tests (count of rows in raw_upload_tests.csv).
upload_samples_peak	int	The number of upload tests (count of rows in raw_upload_tests.csv) - only considering tests during peak hours i.e. where is_during_peak_hour is TRUE.
trimmed_mean_latency_ms_24h	decimal	The 1% trimmed mean (average of the middle 98% of data) of latency_ms. Results where latency_samples_24h is less than 5 are removed from the final dataset.
trimmed_mean_latency_ms_peak	decimal	The 1% trimmed mean (average of the middle 98% of data) of latency_ms - only considering tests during peak hours i.e. where is_during_peak_hour is TRUE. Results where latency_samples_peak is less than 5 are removed from the final dataset.
latency_samples_24h	int	The number of upload tests (count of rows in raw_latency_tests.csv).
latency_samples_peak	int	The number of upload tests (count of rows in raw_latency_tests.csv) - only considering tests during peak hours i.e. where is_during_peak_hour is TRUE.

Raw Data

unit_summary_statistics_netflix.csv

Field Name	Type	Description
unit_id	int	Unique identifier for an individual Whitebox.
netflix_uhd_fraction	decimal	The percentage of successful Netflix tests where netflix_bitrate_mbps was greater than 6 and successes was equal to 1.
netflix_hd_fraction	decimal	The percentage of successful Netflix tests where netflix_bitrate_mbps was between 2.35 and 6, and successes was equal to 1.
netflix_sd_fraction	decimal	The percentage of successful Netflix tests where netflix_bitrate_mbps was greater than less than 2.35, and successes was equal to 1.
mean_netflix_download_mbps	decimal	The (untrimmed) mean of netflix_download_mbps - results where there were less than 5 tests were excluded from the final dataset.
netflix_samples	int	The number of Netflix tests.
check_percentages_add_to_1	boolean	TRUE/FALSE to check that percentages add to 1, with a tolerance of 0.01
max_concurrent_uhd_streams	int	The greatest integer smaller than mean_netflix_download_mbps divided by 15.6 (i.e. floor(mean_netflix_download_mbps / 15.6). This field is used to as an estimate of the number of simultaneous users who could stream Netflix in UHD; because this field is estimated based on download speed rather than bitrate, the threshold is different to that used in netflix_uhd_fraction.

unit_summary_statistics_social_media.csv

Field Name	Type	Description
unit_id	int	Unique identifier for an individual Whitebox.
service	string	Currently one of Facebook-app, Facebook-messenger, Instagram-app, Instagram-messenger, Snapchat, Whatsapp, Twitter
media	string	Currently one of Image, Text, Audio, Video
direction	string	One of "Downlink" or "Uplink". Indicates whether we're measuring receiving or uploading content.
trimmed_mean_latency_ms	decimal	The 1% trimmed mean (average of the middle 98% of data) of latency_ms. Results where social_media_samples is less than 5 are removed from the final data set.
median_hop_count	int	The median number of hops to the server
trimmed_mean_hop_count	int	The 1% trimmed mean (average of the middle 98% of data) of hop_count. Results where social_media_samples is less than 5 are removed from the final dataset.
social_media_samples	int	The number of successful tests.

unit_summary_statistics_gaming.csv

Field Name	Type	Description
unit_id	int	Unique identifier for an individual Whitebox.
game	string	The name of the game tested.
target	string	The target server tested against.
trimmed_mean_latency_ms	int	The 1% trimmed mean (average of the middle 98% of data) of latency_ms. Results where gaming_samples is less than 5 are removed from the final dataset.
gaming_samples	int	The number of successful tests.