Measuring
Broadband
New Zealand

Measuring Broadband New Zealand

Raw data and data dictionary

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. The first report provides an overview of the initial findings from the data collected during the early stages of the project.

Raw data

The raw data contains the raw measurement data and the associated database queries that were used to produce the Initial Findings Report.

Raw measurement data

Measurements are presented in their raw, unaggregated form. However, only measurements that were used within the report have been included in this raw data package. This means that tests other than download, upload, latency and loss have bene excluded (e.g. Netflix, YouTube). Additionally, metadata fields that were not used in the report have been excluded (e.g. RSP name and RSP's product).

Please see the enclosed data dictionary for details on how to interpret the fields contained in the raw data files.

Database queries

The database queries used to produce the charts in the report have been included in this package. These queries are slightly modified from the original queries, as the originals relied upon additional sharding keys that were not included in the raw data being published. This has no impact on the queries though and does not prevent reproduction of the charts.

Presto (https://prestodb.io) was the database used for data analysis. All queries are presented in Presto's SQL-dialect, which will not be directly compatible with other database servers (e.g. MySQL, Postgres, MS-SQL Server). However, with some minor query modifications it should be possible to make the queries run in any of these databases.

The raw data can also be analysed in Excel rather than a full RDBMS, but it is important to read the database query to understand the aggregations and filters being applied.

File listing

File	Description
./raw_data	Contains all of the raw measurement data files
./raw_data/curr_httpget.csv	Download speed test data
./raw_data/curr_httppost.csv	Upload speed test data
./raw_data/curr_udplatency.csv	Latency and packet loss data
./sql_queries_abbreviated	SQL queries (Presto dialect) for data analysis

Data dictionary

curr_httpget.csv (Download speed)

Field Name	Type	Description
unit_id	bigint	Unique identifier for an individual unit
dtime	datetime	The time of the test (local time)
target	string	Hostname of the test server
address	string	IP address of the test server
fetch_time	bigint	Time the test ran for in microseconds
bytes_total	bigint	Total bytes downloaded across all connections
bytes_sec	bigint	Running total of throughput, which is sum of speeds measured for each stream (in bytes/sec), from the start of the test to the current interval. Multiply by 0.000008 to convert to Mbit/s
bytes_sec_interval	bigint	Throughput at this interval. Will match bytes_sec when only a single interval is reported on. . Multiply by 0.000008 to convert to Mbit/s
warmup_time	bigint	Time consumed for all the TCP streams to reach a warmed-up state (Units: microseconds
warmup_bytes	bigint	Bytes transferred for all the TCP streams during the warm-up phase.
sequence	bigint	The interval that this row refers to. Will always be 0 when only reporting on a single interval
threads	bigint	The number of concurrent TCP connections used in the test
tcp_retransmissions	bigint	The number of TCP retransmissions encountered during the test
successes	int	Number of successes (always 1 or 0 for this test)
failures	int	Number of failures (always 1 or 0 for this test)
ip_version	int	IPv4 or IPv6
target_group	string	String representation of the target field
base	string	The Whitebox model number
max_available_download	int	The 'maximum available download speed', as provided by RSPs. Can be NULL.
max_available_upload	int	The 'maximum available upload speed', as provided by RSPs. Can be NULL.
access_technology	string	The access technology of the subscribed product.

Data dictionary

curr_httppost.csv (Upload speed)

Field Name	Type	Description
unit_id	bigint	Unique identifier for an individual unit
dtime	datetime	The time of the test (local time)
target	string	Hostname of the test server
address	string	IP address of the test server
fetch_time	bigint	Time the test ran for in microseconds
bytes_total	bigint	Total bytes downloaded across all connections
bytes_sec	bigint	Running total of throughput, which is sum of speeds measured for each stream (in bytes/sec), from the start of the test to the current interval. Multiply by 0.000008 to convert to Mbit/s
bytes_sec_interval	bigint	Throughput at this interval. Will match bytes_sec when only a single interval is reported on. Multiply by 0.000008 to convert to Mbit/s
warmup_time	bigint	Time consumed for all the TCP streams to reach a warmed-up state (Units: microseconds)
warmup_bytes	bigint	Bytes transferred for all the TCP streams during the warm-up phase.
sequence	bigint	The interval that this row refers to. Will always be 0 when only reporting on a single interval.
threads	bigint	The number of concurrent TCP connections used in the test
tcp_retransmissions	bigint	The number of TCP retransmissions encountered during the test
successes	int	Number of successes (always 1 or 0 for this test)
failures	int	Number of failures (always 1 or 0 for this test)
ip_version	int	IPv4 or IPv6
target_group	string	String representation of the target field
base	string	The Whitebox model number
max_available_download	int	The 'maximum available download speed', as provided by RSPs. Can be NULL.
max_available_upload	int	The 'maximum available upload speed', as provided by RSPs. Can be NULL.
access_technology	string	The access technology of the subscribed product.

Data dictionary

curr_udplatency.csv (Latency and packet loss)

Field Name	Type	Description
unit_id	bigint	Unique identifier for an individual unit
dtime	datetime	The time of the test (local time)
target	string	Hostname of the test server
rtt_avg	bigint	Average round-trip time in microseconds
rtt_min	bigint	Minimum round-trip time in microseconds
rtt_max	bigint	Maximum round-trip time in microseconds
rtt_std	bigint	Standard deviation round-trip time in microseconds
successes	int	Number of successful packets (note: use failures/(successes + failures)) for packet loss
failures	int	Number of packets lost
ip_version	int	IPv4 or IPv6
target_group	string	String representation of the target field
base	string	The Whitebox model number
max_available_download	int	The 'maximum available download speed', as provided by RSPs. Can be NULL
max_available_upload	int	The 'maximum available upload speed', as provided by RSPs. Can be NULL
access_technology	string	The access technology of the subscribed product.