Technical Annex

Technical methodology

This report is Ofcom's seventeenth fixed-line residential broadband speeds report and the fourteenth in which we have published broadband provider package-specific data and comparisons between broadband providers. The technical methodology chosen is the same as that used in Ofcom's previous reports and is based on that created by broadband performance company SamKnows Limited, Ofcom's technical partner in this research project.

SamKnows recruited a panel of UK residential broadband users and supplied monitoring units, known as 'Whiteboxes' to each panellist. SamKnows also managed the collection and aggregation of the performance data and made a major contribution in assisting Ofcom in the analysis of the data.

All panellists were sent a hardware monitoring unit which they were instructed to connect to their router. The monitoring unit sits between the panellist's router and the rest of their network, thereby allowing the unit to determine when the network is free to run tests (the device operates in a bridging mode, rather than routing).

The measurement units are connected to panellists' routers using an Ethernet cable in order that the test results accurately reflect the performance of their connections. Where people use Wi-Fi (or other technologies such as powerline) to connect devices to their router, it is possible that the actual speeds received will be lower than those delivered over an Ethernet connection because of the limitations of these technologies (although recent mass market Wi-Fi technologies can theoretically support speeds in excess of 300 Mbit/s). The potential for this difference is greater for higher-speed broadband connections, where the speeds delivered may be higher than the maximum bandwidth that the in-home network technology can support.

This report also used broadband provider-provided data where the SamKnows Router SDK had been embedded directly into the customers' CPE. The Router SDK includes all the SamKnows quality of service (QoS) and quality of experience (QoE) tests, as well as cross-traffic detection capabilities. The test methodology used is the same as employed in the 'Whitebox' and as described below.

SamKnows developed a customised OpenWRT firmware image which is installed on the units. At the point of delivery to the panellists, this is all that is present on the device; the physical unit contains no additional software, apart from a single script that checks for the availability of the software component at boot-up. This is beneficial both from a security perspective (everything is destroyed when the power is lost) and from a support perspective (any problems with a unit's configuration can be undone simply by power-cycling it). New versions of the software can be delivered remotely without requiring a reboot.

Software within the unit then performed a range of tests to a set schedule, with individual panellists running up to 75,000 (generally around 62,000) separate tests over the course of a day. The software was configured to identify other network activity and not to run tests when such activity was detected. This avoided compromising results by running tests at a time when bandwidth was

being used by other internet-connected devices in the household (including those using a wireless connection).

The software uses a combination of standard UNIX tools and custom code developed in C and C++.

All monitoring units maintain accurate time using ntp.

We believe that this technical methodology is robust as it does not rely on monitoring solutions that do not account for the impact on speed of PC set-up, or for having more than one computer using a broadband connection.

Speed tests

The project uses speed tests with multiple concurrent TCP connections, to assess the capacity of the user's broadband connection.

Speed tests run for a fixed duration of 10 seconds if the user's broadband connection is not subject to a data cap or has had it lifted for the purposes of this project. Fixed-duration speed tests ensure comparability across broadband connections regardless of their access speed.

On connections slower than 30Mbit/s, units download 6MB using separate TCP sessions (in parallel), or transfer for 10 seconds (whichever is reached first). Connections faster than 30 Mbit/s but slower than 50Mbit/s will transfer an increased amount during the downstream throughput test. This amount is up to 12MB or 10 seconds (whichever is reached first). Connections of 50 Mbit/s or faster are all without data caps and therefore employ the full 10-second speed test.

The nature of the protocols used on the internet means that during a file download the speed at which data is sent is gradually increased until a stable speed is achieved. To measure this stable speed, our tests exclude the period of the speed ramp-up. The exact way in which the speed ramp-up occurs on different networks may lead to slight variations in the accuracy with which the stable speed can be measured.

An initial lead-in period is used to ensure that TCP window sizes are increased before measurements are made. Multi-thread tests were run twelve times per day, once every three hours in off-peak periods and once every hour at peak times. We found that, typically, the download speeds achieved using the multi-thread tests in the early hours of the day determine the maximum speed the line can support.

Additionally, it is understood that some broadband providers operate transparent HTTP proxy servers on their networks. To overcome this, the web servers are configured to respond with the following headers, which should disable caching in standards-compliant proxy servers:

Cache-Control: "private, pre-check=0, post-check=0, max-age=0"

Expires: 0

Pragma: no-cache

Similar to the download tests, connections with upload speeds slower than 10Mbit/s upload 3MB or transfer data for 10 seconds (whichever is reached first). Connections with upload speeds higher than 10Mbit/s and lower than 20Mbit/s upload 6MB or transfer data for 10 seconds (whichever is

reached first). Connections with upload speeds higher than 20Mbit/s employ the full 10-second speed test. Four speed-test servers are deployed in a data centres in London to handle the traffic. 40Gbit/s of capacity is shared between these servers. Each server is monitored for excessive network load and for CPU, disk and memory load.

The test results gathered by each server are compared against one another daily, to ensure that there is no significant variation in the speed attainable per server. Units cycle through the speed-test servers in a round-robin fashion when testing.

Testing web page loading times

The test downloaded the HTML and media assets of a simple web page hosted on a SamKnows-managed server. This makes use of up to eight concurrent TCP connections to fetch the assets. Both tests make use of libcurl.

The time in milliseconds to receive the complete response from the web server is recorded, as well as any failed attempts. A failed attempt is deemed to be one where the web server cannot be reached, or where a HTTP status code of something other than 213 is encountered.

Tests were run every hour.

Testing latency, packet loss and jitter

A bespoke application was used to test latency and packet loss. The application was designed to run continuously to get a statistically robust set of data. The test used UDP rather than ICMP and sent approximately 2000 packets every hour.

The test also records instances of contiguous packet loss events. These are termed 'disconnections'. The duration of the disconnection event will vary by its cause – a minor routing issue may only cause a few seconds' disconnection, whereas a modem losing synchronisation with the telephone exchange may result in a 30 second disconnection.

Another bespoke application was used to test jitter. This application sends 500 packets upstream and 500 packets downstream via UDP, every hour, and records the number of packets it sent and received (thus providing a loss rate), and the jitter observed for packets it received from the server. The server does the same, but with the reverse traffic flow, thus providing bi-directional loss and jitter.

Testing recursive DNS resolver responsiveness and failures

Testing a broadband provider's recursive DNS resolution can be accomplished using many tools, such as nslookup, dnsip and dig. A custom DNS measurement client is used so that we may support regular UDP-based DNS queries, as well as emerging standards such as DNS-over-HTTPS and DNS-over-TLS.

For regular UDP-based DNS queries, as used in this report, the Whitebox will honour the DNS servers offered by the home router over DHCP. This router may provide the DNS servers of the broadband

provider directly, or it may provide its own IP address and then proxy the DNS requests upstream. In all cases, we record the DNS server IP address that the router provided over DHCP.

The tests record the number of milliseconds for a successful result to be returned. A successful result is deemed to be one when an IP address is returned (the validity of the IP address is not checked). A failure is recorded whenever the DNS server could not be reached, or an IP address was not returned. The hostnames of four popular websites were queried every hour.

Testing Netflix video streaming performance

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 in 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 from which Netflix server this user's IP address would normally be served content. This logic will consider the broadband provider and geographic location of the requesting IP address. Where the broadband provider 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 then establishes a HTTP connection to the returned server and attempts to fetch the 25MB binary file. This runs for a fixed 20 seconds of real time. HTTP pipelining is used to request multiple copies of the 25MB binary, ensuring that if the payload is exhausted before the 10 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 is 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 can then infer when a stall occurred (by examining when our simulated video stream has fallen behind real-time). The test currently simulates videos at bitrates of 235Kbps, 375Kbps, 560Kbps, 750Kbps, 1,050Kbps, 1,750Kbps, 2,350Kbps, 3,000Kbps, 4,500Kbps, 6,000Kbps and 15,600Kbps.

The test captures the 'bitrate reliably streamed' (the highest quality video that can be streamed without rebuffering events), the download speed from the Netflix Open Connect Appliance and the video start-up delay.

Testing social media platforms

This test measures the round-trip latency and reachability of a selection of major social media services, taking into account different endpoints that the social media service may use for different content types.

A single social media site can use a variety of endpoints for different content types (e.g. audio, video) and different activities (downloading and uploading). For example, Facebook use a different set of servers when allowing users to download videos versus upload videos. This test captures round-trip latency to all of the supported combinations.

Figure A1 summarises the different content types and social media services for which we measure latency.

Lack of support for a particular combination is due to the social media service itself not supporting certain content types. For example, the main Facebook app does not support downloading or uploading audio clips (but the Facebook Messenger app does).

Figure A1: Support for download and upload of various media types by major social media and messaging services

	Download: Text	Download: Image	Download: Video	Download: Audio	Upload: Text	Upload: Image	Upload: Video	Upload: Audio
Facebook App	Yes	Yes	Yes		Yes	Yes	Yes	
Facebook Messenger	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Instagram App		Yes	Yes			Yes	Yes	
Instagram Messenger	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
WhatsApp	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Snapchat	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Twitter	Yes	Yes	Yes		Yes	Yes	Yes	

Source: SamKnows

Note that we have split Instagram into two separate apps, even though they are delivered to the user as a single smartphone app. This is due to the private messaging feature within the Instagram app supporting different functionality to the main part of the app. Moreover, the private messaging feature of the Instagram app uses different endpoints to the main part of the app.

For each social media service we determined the endpoints to test against by performing a traffic analysis of how their Android and iOS apps behaved. All apps, with the exception of Facebook and Instagram, use a static set of endpoints. For example, Twitter uses api.twitter.com for most operations. Of course, this does not prevent Twitter from geographically load balancing api.twitter.com via any cast or DNS based load balancing, but all clients use this single hostname. Facebook and Instagram make use of the Facebook "FNA" caches for retrieval of image and video content. Facebook FNA caches are the on-premises caches that Facebook provides to large broadband providers, much like Google do with GGCs or Netflix does with OCAs. The Facebook FNA cache to be used is determined dynamically by the latency measurement client.

Additionally, the latency measurement mechanism can vary for some social media sites too. All services, with the exception of Snapchat, currently use ICMP to measure round trip latency. For Snapchat we instead use HTTP time-to-first-byte for Snapchat because they front all of their API servers (currently hosted in the US) by Amazon's CloudFront reverse proxy, which are distributed

globally. To just measure round trip time to their CloudFront hostname would misrepresent the end-to-end latency that a user really experiences.

The social media test fully supports IPv4 and IPv6. It may optionally be run with DNS resolution performed over DNS-over-HTTPS or DNS-over-TLS, instead of using the default system resolver. For each social media service and endpoint, we measure the following:

- Average round trip latency for 10 packets (by default)
- Minimum, maximum, median and standard deviations for the latency measurements
- The number of sent and received packets
- The number of hops to the endpoint (performed using a traceroute)
- The resolver IP address of the endpoint

Connections with usage caps

Some of the test units were deployed on broadband connections with relatively low usage caps. To avoid using a significant proportion of the available download limit each month, the test schedule for the test units on these connections was reduced.

Research methodology

The performance data in this report are taken from a base of 3,207. This includes data from 2,844 panellists on SamKnows' independent global platform, SamKnows One, who had a broadband monitoring unit connected to their routers in November 2019, and 363 panellists who had the SamKnows SDK embedded into their CPE. Ofcom's definitions of geographic broadband markets use the definitions for the wholesale broadband access (WBA) market review. These were an important consideration in recruiting our panel and applying statistical analysis, because they enabled us to ensure that our panel was representative of the UK residential broadband market overall, and facilitated like-for-like comparison between broadband provider packages:

- Each panellist was assigned to one of the geographic markets, and we weighted the analysis accordingly to ensure that our overall findings were representative of UK residential broadband performance overall (for example, as Market A represents around 1% of UK premises, we ensured that performance data from panellists in Market A contributed 1% towards the overall computation of UK residential broadband performance).
- For comparisons of broadband provider package performance, we used only panellists who live within geographic markets B. This means that all panellists used for the broadband provider package comparisons live in areas served by a local telephone exchange in which at least one operator other than BT is present, i.e. there is at least one local loop unbundling (LLU) operator. This avoids any potential distortions of the data by broadband providers using BT wholesale services (BT Retail, EE and Plusnet), caused by the inclusion of panellists who live in (typically less densely populated) 'Market A' areas, and to whom LLU services are not available.

We have used statistical techniques to adjust our results to ensure that they are representative of the overall UK broadband population. This includes weighting the results from our panel by rural/urban, distance from exchange, geographic market definition and broadband provider. For the provider-specific comparisons we have also 'normalised' the data for ADSL operators by distance from exchange (using the straight-line distance from the panellist's location to the exchange), which we believe is necessary to provide like-for-like comparisons of broadband providers which have different customer profiles.

David Saville of Saville Rossiter-Base has assessed the research methodology and panel and helped ensure its suitability for purpose. Checks were also applied to ensure that straight-line distance was an appropriate metric to carry out normalisation, including comparing this distance with the line attenuation. Details of the statistical methodology used are provided in Annex 2. The methods of analysis for the provider-specific comparison are based on those used in the July 2009 report which had expert review by econometrician Professor Andrew Chesher of University College London.