making communications work
for everyone

Connected Nations 2022

Annex: Mobile performance

Publication date: 15 December 2022

How crowdsourcing works

Mobile crowdsourcing works by collecting anonymous measurement information from consumer mobile devices to provide a view of network performance from a user perspective. To achieve significant scale and a broad distribution of samples, the mobile crowdsourced data is collected through partnerships with popular mobile applications with a considerable number of users. These partners embed the crowdsourcing software (known as a software development kit, or SDK) in their own apps to collect performance measurements.

We are working with a mobile crowdsourcing data provider named Opensignal, who have provided us with access to a crowdsourced dataset. This allows us to assess the latest view of mobile network performance from the consumer's perspective.

No personal identifiable data is passed on to Ofcom and throughout the crowdsourced data collection process. Opensignal have assured us that no personal identifiable data is collected and that all collected data are further anonymised in a way such as not to be able to identify any individual or individual usage patterns.

Measurement methodology

With Opensignal's crowdsourcing platform, data is collected in various locations all around the UK covering both rural and urban areas, as well as indoor and outdoor locations, and at various times in the day.

Two types of tests are performed when collecting the data: active background tests and passive measurements. Active background tests are those defined and initiated by Opensignal to measure the device's network performance irrespective of the user's activity. Passive measurements refer to the collection of information regarding the current network environment, such as signal strength, from the device operating system. We are only using the data from active background tests for our current assessment.

To ensure diversity and robustness of measurement data, active tests are conducted according to a set of triggers defined by Opensignal which are activated in a variety of situations, including changes to location and network configuration, app activity and when certain timers elapse. The combination of these triggers ensures measurements are performed throughout the day, providing a broad and representative set of network performance data.

From the measurements calculated during the active background tests, we focussed on 6 Key Performance Indicators (KPIs), namely; **Download speed, Upload speed, Latency, Jitter, Packet Discard and Time to First Byte.** These metrics are in line with activities performed in consumer mobile applications and are likely to closely reflect the perceived consumer quality of service. The way each of these metrics are calculated is further detailed below.[1]

[1] More detailed information on KPI metrics can be found in Opensignal's website.

Download speed

Download throughput tests are performed by downloading files of various sizes using the Hypertext Transfer Protocol (HTTP)[2] from content delivery networks (CDNs) set up on several edge test servers. File sizes of 2MB, 5MB and 10MB are used to perform download throughput tests and these simulate a typical user download activity i.e. downloading a webpage. Note that this measurement does not measure the maximum throughput of the network but is typical of customer interactions and can be used to measure relative performance between operators.

Upload speed

For upload throughput tests, Opensignal uses a 1MB file to perform an upload throughput test. The file is sent using HTTP to a CDN. This test simulates a user performing a typical upload activity i.e. uploading an image. As with download, the upload tests do not give the maximum upload throughput of the network but is representative of typical usage and can be used to compare relative performance between operators and devices.

Server response tests (UDP Latency, Jitter and Packet Discard)

Server response tests are performed over the user datagram protocol (UDP), a connectionless protocol. Unlike the Transmission Control Protocol (TCP), connectionless protocols such as UDP utilise minimum handshaking[3] and are used in latency sensitive applications such as video and VoIP calling. For these tests, the closest test server is chosen based on geographical distance between test server and SDK device. 100 packets are then sent from the SDK to the test server and the precise time that each packet is sent is recorded to millisecond precision. 20ms spacing is kept between each packet transmission. The test server receives the packets sent from the SDK and immediately echoes them back to the listening SDK device. The precise time that each packet is received is also recorded. Based on these gathered data, the server response test metrics are calculated as below:

- **Latency** is half of the round-trip travel time of each packet from the time the packet is sent from the SDK to when the same packet is received;
- **Jitter** is calculated by measuring the change in latency from packet to packet received by the SDK; and
- **Packet Discard Rate** is calculated by determining the ratio, in percentage terms, between the number of packets lost or out of sequence compared to the overall number of packets sent.

[2] HTTP is a set of rules that governs how communication is exchanged over the internet between a client, for example a user web browser, and a server such as a website. This set of rules enable the transfer of information such as text, images, sound, video, and other multimedia files over the web.
[3] The processes of setting up a communication link.

Time to first byte

This is a measure of TCP latency associated with the download throughput test. When a download throughput test is conducted, the SDK will record the time it takes for the first payload byte in the download packet to reach the device.

Data analysis methodology

We set out below our methodology for processing this data to enable us to meet our objective of assessing the quality of network performance available to a typical user. These are:

a) We processed a dataset which had been gathered over a nine-month period, recognising the need to maintain a good balance in terms of providing a recent view of the network performance (considering that networks evolving over time) but also ensuring that enough historical samples are included to be statistically significant for making an assessment.

b) Splitting the view of the data by geographical areas. In the case of this Connected Nations report, UK postcode districts have been used to provide detailed local insight while ensuring sufficient samples in most areas to provide statistical significance.[4]

c) Grading the consumer quality of service in each area into 4 main performance classes, namely, **Basic, Good, High and Very high performance** based on whether all the Key Performance Indicators (KPIs) mentioned above meet or exceed all the thresholds defined for a given performance class as shown in Table 1.

Table 1: Key Performance Indicator thresholds for each performance class

	Download throughput	Upload throughput	Latency	Jitter	Packet loss	Time to First Byte
Basic	*<2 Mbps*	*<0.5 Mbps*	*>100ms*	*>20ms*	*>4%*	*>1.2s*
Good	*>=2 Mbps*	*>=0.5 Mbps*	*<=100ms*	*<=20ms*	*<=4%*	*<=1.2s*
High	*>=5 Mbps*	*>=1 Mbps*	*<=50ms*	*<=15ms*	*<=2%*	*<=0.8s*
Very high	*>=10 Mbps*	*>=2 Mbps*	*<=30ms*	*<=10ms*	*<=1%*	*<=0.5s*

[4] Postcode district means a geographical area indicated by the (alphabetical) letters and numbers in a postcode preceding the space in the code e.g. SE1.

d) Assigning a postcode district's overall performance class is done by determining the highest performance class achieved by at least 80% of the crowdsourced data recordings that took place within it. For example, suppose 90% of all individual data recordings taken within a specific postcode district meet the criteria for 'good performance', 80% meet the criteria for 'high performance', and 70% meet the criteria for 'very high' performance. In that case, the postcode district will be classed as 'high performance'. If less than 80% of measurements meet the 'good performance' criteria, the postcode district will be classified as 'basic performance'. If there's not enough data recording within a postcode district to make a statistically reliable assessment, then that postcode district will be classed as having 'unknown performance'.

e) Increasing the reliability of our performance assessment by setting out a minimum unique device type and sample count that must be met before a postcode district can be classified as having a particular performance class (see statistical confidence section). This will limit the possible margin of error around these percentage estimates.

We will continue to evolve our approach to ensure we provide the most accurate picture of network performance, even as networks and consumer devices continue to evolve and new capabilities are enabled by crowd source platforms.

Statistical confidence

To ensure that the sample data from each postcode district is statistically robust, we have only included a performance view for postcode districts that meet both the following criteria:

- Minimum of 100 crowdsourced data recordings taken within the postcode district; and
- Minimum of 15 unique device types contributing crowdsourced data recordings in this postcode district.

These criteria ensure that the data recordings used to summarise the mobile performance of a postcode district come from a variety of individuals (rather than only one or two phones contributing all the measurements) and are from a sufficiently large sample of measures, reducing the margin of error around the percentage of data recordings meeting a performance class.

Given these criteria, if 80% of the data recordings taken within a postcode district are classed as 'high performance', the maximum margin of error around this percentage would be around 8.7% at the 97% level.[5] The true margin of error around the percentage figures for each postcode district will

[5] Margin of error quantifies the degree of uncertainty we have in the results due to them coming from a sample of data recordings taken from a sample of phones within a postcode district rather than from all possible phones in this area. It is half the size of the confidence interval. A 97% confidence interval around a % means: if we randomly took samples from this same area/during the same period, 97% of the time, the 97% confidence interval around this figure will contain the true population estimate. Suppose a postcode district has 80% of its data recordings meet the criteria for 'high performance', and when we calculate the 97% confidence interval, we get a lower bound of 75% and an upper bound of 85%. In that case, we can say that there is a 97% chance that this confidence interval for this sample percentage contains the true population percentage. Given a confidence interval between 75% and 85%, the margin of error for the 80% figure will be 5% at the 97% level.

depend on the percentage itself and the number of data recordings taken within each postcode district.[6]

Our use of statistical parameters is informed by analysis of the overall dataset and the distribution within, and our view on best practice approach to statistical significance, as well as consultation with Opensignal. As with our measurement methodology, we will keep this statistical approach under review as we develop our further assessment of mobile network performance.

Limitations

As noted in our UK report, we recognise that no single tool is likely to provide a complete view of mobile network performance that distinguishes all influencing factors. For example, this approach does not currently enable us to provide insights across the entirety of UK, given the limited samples available in the many rural areas. It also aggregates samples from different device locations at the time of the test (e.g. whether a device was indoors or outdoors) and whether the test was conducted at a very busy period or not, such that additional, more granular insight may be possible over time. We will consider whether and how our approach can be refined as we make greater use of crowdsourced data in the future.

[6] We have set a limit of 10% for the margin of error that must be achieved for a set of data recording from a postcode district to be classed as having a particular performance class. If the margin error exceeds this limit, then the performance class is downgraded to a lower performance class where the calculated margin of error is under the 10%. For example, suppose more than 80% of all individual data recordings taken within a specific postcode district meet the criteria for 'high performance', but has margin of error of 12%, and the same postcode district achieves a margin of error of 9% with a grading of ‘good performance’, then that postcode district will be classed as having ‘good performance’.