ETSI TR 103 501 V1.1.1 (2018-10)

TECHNICAL REPORT

Speech and multimedia Transmission Quality (STQ); Guidelines for the Measurement of Data Throughput on Devices connected to Mobile Networks

Reference
DTR/STQ-00217m

Keywords
3G, data, GSM, network, QoS, service

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the only prevailing document is the print of the Portable Document Format (PDF) version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Speech and multimedia Transmission Quality (STQ).

Throughput, or data rate, is the single most important property of a packet data network. The definition of throughput - transferred volume of data per unit of time - is essentially simple. However, there are many different methodologies available to measure it. To select the most appropriate one for a given purpose, and to assess comparability of results, requires thorough understanding of these methodologies. The present document addresses the measurement methodologies that can be used from an end-user perspective, i.e. embedded in devices or in dedicated test equipments connected to mobile networks.

While there is extensive coverage of IP layer centric methodology (such as ETSI EG 203 165 [i.1] ("Throughput measurement Guideline"), the content of such ETSI Guide does not actually cover methodologies and aspects such as application-level measurements. The present document takes also in consideration methods known as 'crowdsourcing', which have gained considerable audience (and potentially relevance) in the last years but until the time of publication of the present document have not been subject to extensive treatment in the framework of standardization work (however there are activities under way, e.g. the E.MTSM work item in Question 12 of the ITU-T Study Group 12).

Likewise, the present document integrates multi-threaded measurements into a common methodological frame.

The present document has been written to provide a holistic, organized view of the entire measurement process, which also includes elements such as definition of system under test and system boundaries, post processing of data and relation between basic methodologies and their relation to intended targets of measurements.

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Introduction

Throughput, or data rate, is the single most important characteristic of packet data networks. While its definition - transferred data volume per unit of time - is simple, there is a wide range of possibilities how actual measurements can be carried out. Consequently, it is hard to decide if results of different measurements are comparable of how results from one type of measurement can be used to predict the outcome of another type of usage.

Also, there are entirely different views on performance characteristics of packet data networks. From a low-level perspective (on the IP layer), a network transfers data packets and its performance is characterized by packet transfer time, latency or delay and packet loss rate; also, the variations in time of these metrics, as well as less frequent events such as packet re-ordering have an effect. On higher protocol layers such as TCP, there is no packet loss; lost packages on lower layers translate into lower overall data rates. From an end-user, QoS or QoE perspective, the performance of a network may again vary as the dynamics of a particular application interact with network characteristics and behaviour of other components in sometimes complex ways. In addition, networks typically use resource and performance optimization mechanisms which further increase the complexity of dynamic behaviour.

Mobile connectivity has become an important element of modern life. Both business and consumers have great interest in knowledge about the performance of mobile networks and useful information about this performance is in high demand. There are few actors which have the means to use professional measurement tools to obtain such information. This is one of the reasons why in recent years, a substantial number of companies have emerged which develop and distribute crowdsourcing tools - subsequently termed 'speed test apps' although this description is not entirely correct - to measure mobile network performance. Due to promised or expected cost reduction, even network operators use and rely on such tools today. In some countries, regulators operate crowdsourcing tools too.

From their mode of operation, these tools are considered to effectively measure network performance from an end customer perspective, as the tools run on end-user smartphones. Use cases typically are http or ftp upload and download and therefore results can be attributed to be QoS values and do not represent actual internet speed measurements as understood by e.g. laboratory measurements or assessments in the regulatory context.

Strictly speaking, the scenarios used represent only a small fraction of actual end-user behaviour as pure upload and download only plays a role in use cases such as app download or transfer of larger number of data as e.g. in transfer of photos or videos (typically in e-mail or cloud storage contexts). Nevertheless, such tests play a significant role in the public perception of mobile networks. Through the interplay between public media and PR of mobile network operators, they have a substantial economic impact.

The basic requirement for any meaningful, professional measurement is repeatability. As long as the properties of the network under test and of the test equipment or application stay the same, running the same test is expected to produce the same results. Repeatability allows then comparison. But comparison can also be understood between measurement tools or applications, and in such a case, it requires that the relevant procedures (parameters, set up) of the test are fully documented. This is usually not the case with current 'speed test' applications, and even less so regarding full interoperability, e.g. by open access to servers used as counterpart of throughput testing.

The present document provides a contribution to the evolution of network performance testing towards a professional degree of transparency. This begins with a consistent framework of definitions and technical terms. The elements of the testing process are then described within this context.

Apart from the obvious direct parameters of throughput testing, such as time windows or transferred data volumes, there are numerous other elements which can have an impact on data values obtained. In this sense, methodology and definition of metrics cannot be decoupled from each other. The process starts with selecting the boundaries to the system under test, i.e. insertion or demarcation points. Next comes the way the system under test is accessed. For instance, if the test is run over a radio access network using a mobile device such as a smartphone, the type and degree of influence needs to be assessed. The type of stimulus is likewise important, such as the protocol type, the structure of data traffic (e.g. TCP or UDP based), and the number of parallel connections. Depending on these selections, other choices also become parameters for testing. An example would be to use some kind of real application to create a particular type of traffic, versus using synthetically generated traffic.

The need for careful consideration is not limited to the generating side of measurement data. The way data is processed may also have an impact on results and is therefore subject to documentation and transparency. This applies e.g. to rules about which data to include in computation of results, and which to ignore or discard. For instance, a methodology may require discarding a certain number or range of extreme values to reduce the volatility of results.

Beyond the direct uses of throughput measurement, one of the driving forces is the prospect of using respective data to predict or infer QoS or QoE for a broader range of services. This is of course desirable from a commercial and practical point of view, in order to reduce the complexity of testing as compared to running actual service test use cases. There is no doubt that doing this is possible in principle, using different approaches such as a model-based ones or empirical methods, i.e. a data driven mapping between results from both domains, can be used. It is beyond the scope of the present document to discuss this topic in detail. It is however clear that a meaningful way to do so involves a large degree of caution and professional care - e.g. in calibration and validation of methods. In any case, it will be necessary to gauge the efficiency, data quality the overall effort of such approaches against direct service tests to obtain QoS or QoE parameters by running actual use cases.

1 Scope

The present document provides a systematic overview of methods to measure throughput in mobile networks, with special focus on measurements using a viewpoint at, or close to, application level. Also, it provides a holistic, integrated view of the measurement process, which also includes a selection of methodologies according to intended goals of measurement, and also covers post-processing and data aggregation aspects.

2 References

2.1 Normative references

Normative references are not applicable in the present document.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] ETSI EG 203 165 (V1.1.1): "Speech and multimedia Transmission Quality (STQ); Throughput Measurement Guidelines".

[i.2] ETSI TS 102 250-2: "Speech and multimedia Transmission Quality (STQ); QoS aspects for popular services in mobile networks; Part 2: Definition of Quality of Service parameters and their computation".

NOTE: The content of this document series has also been used (copied) in Recommendation ITU-T E.804.

[i.3] Recommendation ITU-T Q.3960: "Framework of Internet related performance measurements".

NOTE: Available at http://www.itu.int/rec/T-REC-Q.3960-201607-I/en.

[i.4] ETSI TR 102 678: "Speech and multimedia Transmission Quality (STQ); QoS Parameter Measurements based on fixed Data Transfer Times".

[i.5] ETSI TS 138 521-3: "5G; NR; User Equipment (UE) conformance specification; Radio transmission and reception; Part 3: Range 1 and Range 2 Interworking operation with other radios (3GPP TS 38.521-3)".

[i.6] ETSI TS 138 101-3: "5G; NR; User Equipment (UE) radio transmission and reception; Part 3: Range 1 and Range 2 Interworking operation with other radios (3GPP TS 38.101-3)".

[i.7] Recommendations ITU-T Y.154x series: "Quality of service and network performance".

[i.8] Recommendation ITU-T Y.1545.1: "Framework for monitoring the quality of service of IP network services".

[i.9] IETF RFC 7398: "A Reference Path and Measurement Points for Large-Scale Measurement of Broadband Performance".

3 Definitions of terms and abbreviations

3.1 Terms

For the purposes of the present document, the following terms apply:

data path: sequence of entities or elements (physical/virtual) which a data packet transferred between endpoints traverses

endpoint A: local entity in an end to end test scenario (typically, the actual test system)

endpoint B: remote party in an end to end test scenario (in packet data tests, often a server connected to the public internet; can also be a CDN in case of live public content

fixed-size method: throughput measurement method where a fixed amount of data is transferred and the time for transfer is recorded to calculate a throughput value

fixed-time method: throughput measurement method where data transfer is performed for a fixed period of time, and the amount of data is recorded to calculate a throughput value

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

API	Application Programming Interface
CDN	Content Delivery Network
DL	DownLoad
FTP	File Transfer Protocol
HTTP	Hyper Text Transfer Protocol
HTTPS	HTTP Secure
HW	HardWare
ICMP	Internet Control Message Protocol
IP	Internet Protocol
IT	Information Technology
ITU-T	International Telecommunication Union - Telecommunication
MDR	Mean Data Rate
NuT	Network under Test
OTT	Over The Top
PoO	Point of Observation
PR	Public Relations
QoE	Quality of Experience
QoS	Quality of Service
QUIC	Quick UDP Internet Connections
RAT	Radio Access Technology
RTT	Round Trip Time
SI	International System of units (Système international d'unités)
TBKPI	Time Based Key Performance Indicator
TCP	Transport Control Protocol
TP	ThroughPut
UDP	User Datagram Protocol
UL	UpLoad

4 Background

There are standards documents that deal, in part quite extensively, with pure throughput measurement at IP level, or with measurements where throughput or round trip time is part of a larger metric for a given use case or service. The main reason for considering round-trop time together with throughput is that TCP throughput is dependent on both the minimum capacity of all path segments, and the round-trip time (because of the feedback loop in TCP flow-control, this would apply to QUIC or any retransmission protocol in general).

The common factor in these standards is that they focus strongly or exclusively on methods based on events of lower protocol levels (typically, the IP or TCP plane). However, the reality at the time of publication of the present document is that there are various tools available - which sometimes appear to be quasi-standards making use events from higher layers - for which there are good reasons, explored further in the course of the present document.

NOTE 1: The term *application plane* is used here as a synonym for Points of Observation (PoO) above the IP layer. Actually, events used in throughput measurement may come from any layer between the API for basic data transfer (e.g. the operating system's socket API) and user-interface indicators in case real smartphone apps are used. According to the principle laid out in ETSI TS 102 250 [i.2], mixing events from different PoO should be avoided whenever reasonably possible.

NOTE 2: Measurement on radio carrier are also possible, but outside of the scope of the present document. For details on this topic, one can refer to ETSI TS 138 521-3 [i.5] and ETSI TS 138 101-3 [i.6].

At the time of writing, there appears to be no document which provides a comprehensive, practically oriented overview of all aspects of the entirety of packet-data network performance measurements, which goes far beyond core measurement methodology and integrates aspects of system boundaries and data aggregation.

While IP-level methods are also mentioned in the present document for completeness, its main focus is on the application plane, and reference is made to documents which treat IP-level measurements extensively, such as ETSI EG 203 165 [i.1] or Recommendation ITU-Ts of the Y.154x series [i.7].

Application-level measurements can be run on practically all devices while low-level data is typically only available if devices are modified, granting apps full system access. However, such modifications, usually called *rooting*, render a device potentially unsafe (typically, this process removes some security features). Even if users would accept that, which is considered very unlikely, the process of rooting is not easy to perform, can lead to permanent damage of devices of something goes wrong, and voids device warranty. Therefore, it is safe to assume that any larger distribution of measurement apps, and in particular crowdsourcing, will have to be based on application-level measurement methods.

This does not mean that application-level methods provide an entirely different point of view. It is assumed that there are clear, deterministic relations between the layers of packet data transfer, so in principle a direct relation to QoS metrics based on low-level events can be established. In this respect, application-level methods provide access to a much broader range of information sources.

Remark: Relations between layers are testable so respective validation can be done when needed.

One aspect has to be taken into account, namely the potential higher degree of device dependency, which requires additional professional care at the level of test design and conduction. Also, application-level measurements may produce less diagnostic depth than measurements on IP level. On the bottom line, the method of choice will be selected based on the type and depth of required information and the relative cost and effort of obtaining it.

5 Basics of throughput measurements

Throughput is defined as transferred data volume per time, or:

$$\text{TP} = \frac{\mathit{Amount\ of\ Transferred\ Data}}{\mathit{Duration}}$$

Where the unit of TP is typically kbit/s.

NOTE: As the amount of transferred data (data volume) is often given in kByte or Mbyte, some caution is indicated. The SI defines prefix k stands as 10^3 and prefix M as 10^6. In the IT world and even in parts of related standardization literature, the prefixes k and M are used for factors based on powers of 2. The prefix k equals 2^{10} (1 024) and M equals 1 024 × 1 024. Obviously, even a single misuse of the k prefix already has a potential error of 2,4 %. This error may multiply for consecutive wrong usages. For technical purposes, the safest way is to not use prefixes at all, but byte values. When this cannot be applied, great care is advisable throughout the documentation chain.

Values for data volume and duration are easy to obtain. However, there are multiple sources for each of them (taken at different points of observation, PoO) and values can only be compared if the PoO are the same.

Practically, events can be closely linked i.e. by a (demonstrable) systematic deterministic relationship (e.g. a function-call chain) so even if the formal PoO is not the same, it can be treated as equal. This applies even to a case where this chain involves a fixed time differential which then would have to be taken into account.

The PoO aspect is not just a formal one, it is also related to the QoS level. If, for instance, a given data transmission path has a non-negligible packet loss rate, the low-level data transmission activity or rate may be high but with no practical use from a user's perspective. Also, transmission protocols may have different amounts of overhead that lead to different user-perception throughput values for the same low-level data rate.

This aspect has a special dimension when using API functions of certain device classes, e.g. the traffic counters of smartphones. When using such data sources, it is necessary to know the level at which these traffic counters are actually counting. At least, the methodology in the context of repeatability of measurement will need to include an element of continuity assurance.

Likewise, the duration needs a proper definition. It is given by two points in time. The throughput of a network under test is not constant over time, in particular at the beginning of a transmission. Therefore, the choice of specific events or otherwise defined points in time will have an impact on the throughput value obtained from a measurement.

In the case of upload data rates, it is important to understand if the data regarded as transmitted is just successfully handed over to the next stage in the chain, or if it is actually transferred across the network under test.

In case of benchmarking measurements, as all channels are using the same architecture and methodology, all aspects mentioned above are less critical. Usually, they do not affect the ranking order of candidates. In case of single-channel measurements, results are usually compared to those measured with other systems or taken at an earlier point in time. Here, it is important to record the used parameters, and to properly understand eventual effects.

Calibration or connection measurements, as used in other fields of engineering, can also be a useful method to ensure consistency. Such measurements would involve measurements on the same system under test using the previous and the new tool or methodology, and producing a statistically relevant number of samples in order to understand eventual differences in output values.

As a summary, throughput measurements are not intrinsically complex. In order to make sure that the results are correct and consistent, there are, however, some principles of proper craftsmanship to be observed.

6 Treating measurement and evaluation methodology as a unit

Practical experience shows that measurement and evaluation methods cannot be treated separately. For a full understanding of applicable evaluation methods and interpretation of results, the method used needs to be known. The single most important example is the selection of file size (or time window) in throughput measurements.

A certain minimum size is necessary to avoid ramp-up and quantization effects. With typical TCP window sizes in the range of 1 Mbit and beyond, and data rates in the order of 100 Mbit/s, a transfer of less than 3 Mbytes to 6 Mbytes will not produce any meaningful results.

Typically, throughput figures obtained from measurements with a particular set of parameters (protocol type e.g. ftp/http; file size) need to be extrapolated to obtain predictions for other parameters, too. It is important to understand how reliable these extrapolations are, considering that a network under test may exhibit a different behaviour if parameters are different. The term *prediction horizon* will be used to elucidate this situation.