ETSI TR 103 559 V1.2.1 (2023-10)

TECHNICAL REPORT

Speech and multimedia Transmission Quality (STQ);
Best practices for robust network QoS
benchmark testing and scoring

Reference

RTR/STQ-00229m

Keywords

3G, benchmarking, data, GSM, LTE, network, QoE, QoS, scoring, service, speech, video, ViLTE, VoLTE

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 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Important notice

The present document can be downloaded from:
https://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 prevailing version of an ETSI deliverable is the one made publicly available in PDF format at www.etsi.org/deliver.

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

If you find a security vulnerability in the present document, please report it through our Coordinated Vulnerability Disclosure Program:
https://www.etsi.org/standards/coordinated-vulnerability-disclosure

Notice of disclaimer & limitation of liability

The information provided in the present deliverable is directed solely to professionals who have the appropriate degree of experience to understand and interpret its content in accordance with generally accepted engineering or other professional standard and applicable regulations.
No recommendation as to products and services or vendors is made or should be implied.
No representation or warranty is made that this deliverable is technically accurate or sufficient or conforms to any law and/or governmental rule and/or regulation and further, no representation or warranty is made of merchantability or fitness for any particular purpose or against infringement of intellectual property rights.
In no event shall ETSI be held liable for loss of profits or any other incidental or consequential damages.

Any software contained in this deliverable is provided "AS IS" with no warranties, express or implied, including but not limited to, the warranties of merchantability, fitness for a particular purpose and non-infringement of intellectual property rights and ETSI shall not be held liable in any event for any damages whatsoever (including, without limitation, damages for loss of profits, business interruption, loss of information, or any other pecuniary loss) arising out of or related to the use of or inability to use the software.

Copyright Notification

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are 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 Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, 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.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

Foreword

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

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

Countrywide mobile network benchmarking and scoring campaigns published in the press enjoy great public interest and are of high importance for the operators of mobile networks. A first place score in press releases associated with such measurements is often used in the advertisements of the winning operator to boost their corporate identity. Though published results are often well documented, they are not always completely transparent about how the actual scoring has been achieved. Methods and underlying assumptions are mostly not described in detail.

The present document discusses the construction and methods of such a countrywide measurement campaign, with respect to the area and population to be covered, the collection and aggregation of the test results and the weighting of the various aspects tested. The applicability of the results of such a campaign, for inter country comparison purposes, is not covered in the present document.

Based on established methods and quality metrics, such as success ratio and setup times, the results of the data collected in the benchmarking are aggregated individually. The individual aggregated values are weighted and further aggregated for each application like telephony, video and data services. The application fields are then in turn weighted and aggregated over the different areas where the data is collected. Finally, calculation of an overall score or a joint score is performed.

The experienced quality of service varies over time so that the individual score of a particular throughput cannot be fixed once and for all. As well as the test metrics changing over time, so does the importance of the various services. The present document describes a typical set of tests that could be performed and related evaluation criteria. In the annexes, actual real-world examples of weightings and score mapping parameters are given.

1 Scope

The present document describes the best practices for benchmarking of mobile networks. The goal of the benchmarking is to determine the best provider or operator for a designated area with respect of the services accessed with a mobile phone. The tests conducted are telephony, video streaming, data throughput and more interactive applications such as browsing, social media and messaging. This goal is achieved by executing benchmarking tests in designated test areas that represent or actually cover a major part of the users of mobile services. The results collected in the various areas are individually and collectively weighted and summarized into an overall score.

Due to the rapid development of the mobile technology and consumption habits of the users, the quality of experience of the users changes over time even when the objective to measure the quality of service does not change. The present document needs to keep up with those changes and does so by parameterizing the individual factors that contribute to the score.

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 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".

[i.2] Void.

[i.3] Void.

[i.4] ETSI TR 101 578: "Speech and multimedia Transmission Quality (STQ); QoS aspects of TCP-based video services like YouTube™".

[i.5] ETSI TR 102 678: "Speech and multimedia Transmission Quality (STQ); QoS Parameter Measurements based on fixed Data Transfer Times".

[i.6] ETSI TR 103 138: "Speech and multimedia Transmission Quality (STQ); Speech samples and their use for QoS testing".

[i.7] Recommendation ITU-T E.840: "Statistical framework for end-to-end network-performance benchmark scoring and ranking".

[i.8] Recommendation ITU-T P.1401: "Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models".

[i.9] Recommendation ITU-T P.863: "Perceptual objective listening quality prediction".

[i.10] Recommendation ITU-T P.863.1: "Application guide for Recommendation ITU-T P.863".

[i.11] IETF RFC 9000: "QUIC: A UDP-Based Multiplexed and Secure Transport".

[i.12] ETSI TR 103 733: "Speech and multimedia Transmission Quality (STQ); Best practices of testing the performance of web content delivery".

[i.13] Recommendation ITU-T Y.1540: "Internet protocol data communication service - IP packet transfer and availability performance parameters".

[i.14] Recommendation ITU-T G.1051: "Latency measurement and interactivity scoring under real application data traffic patterns".

[i.15] ETSI TR 103 702: "Speech and multimedia Transmission Quality (STQ); QoS parameters and test scenarios for assessing network capabilities in 5G performance measurements".

[i.16] Recommendation ITU-T G.1035: "Influencing factors on quality of experience for virtual reality services".

[i.17] Recommendation ITU-T P.565.1: "Machine learning model for the assessment of transmission network impact on speech quality for mobile packet-switched voice services".

[i.18] iperf3, 26.08.2023.

3 Definition of terms, symbols and abbreviations

3.1 Terms

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

live web page: web pages considered as dynamic content, content changes over time and some content might be different caused by the hosting server or the access network

static web page: web pages considered as static content, content stays constant over time and access network

3.2 Symbols

Void.

3.3 Abbreviations

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

AMR	Adaptive Multi-Rate
API	Application Programming Interface
CDN	Content Delivery Network
CST	Call Setup Time
DL	DownLink
EVS	Enhanced Voice Services
FB	FullBand
FTP	File Transfer Protocol
HD	High Definition
HTTP	HyperText Transfer Protocol
IMS	IP Multimedia Subsystem
IP	Internet Protocol
ITU	International Telecommunication Union
ITU-T	International Telecommunication Union Telecommunication
KPI	Key Performance Indicator
LTE	Long Term Evolution (mobile networks)
MB	MegaByte
MOS	Mean Opinion Score
OTT	Ove The Top (services)
PDV	Packet Delay Variation

QoE	Quality of Experience
RTT	Round Trip Time
SMS	Short Messaging Service
SMTP	Simple Mail Transfer Protocol
TCP	Transmission Control Protocol
TP	Throughput
TS	Technical Specification
UDP	User Datagram Protocol
UDPST	UDP Speed Test
UL	UpLink
VoD	Video on Demand
VoIP	Voice over IP
VoLTE	Voice over LTE
VoNR	Voive over New Radio
VSSSR	Video Streaming Service Success Ratio
WB	WideBand

4 Governing Principles for Mobile Benchmarking

4.1 General

The accurate benchmarking and scoring of networks which cover large geographic areas requires careful consideration of a number of factors. These include the technology used, the extent of coverage offered, mobile device evolution, customer population distribution, network usage and tariff offerings. The following principles should be adhered to where possible to ensure that benchmarking scoring outcomes are always meaningful.

4.2 Fair Play

Benchmarking outcomes can be significantly influenced by specific targeting of test devices for superior performance. In such cases the results obtained no longer reflect the experience of a customer using that network. Steps should be taken to ensure that the measured results are truly representative of the real customer experience.

EXAMPE 1: If Operator A implements a special QoS construct specifically for the devices used to collect Benchmarking data, and Operator B does not, the results should not be compared for the purpose of drawing conclusions about the relative experience of customers on each network. The networks should not be compared for benchmarking purposes.

EXAMPLE 2: If Vendor A implements a special functionality in their equipment/device software or firmware to recognize benchmark testing and boost performance, and Vendor B does not, the results may show one vendor to be superior to another for test cases no longer relevant to usual network usage. Vendor performance, from a customer perspective, can no longer be reliably compared.

4.3 Comparing networks with different coverage extents

Often networks are built with differing coverage objectives. Network rollout often varies between operators. This is often an important differentiator for customers making decisions about which network is best for them. Benchmarking should be performed in such a way that it highlights coverage differences in the results. From a scoring perspective, operators should never be penalized for providing coverage where other operators do not. In fact they should instead be rewarded in the scoring system. It should be the intention of any comprehensive mobile benchmark to include coverage comparison as a differentiating factor in the scoring.

EXAMPLE: If Operator A offers significantly more geographic coverage than Operator B, Benchmarking data collection methodology and scoring should be such that this difference is always reflected in the scoring as a 'bonus' rather than a 'penalty' and the Benchmarking methodology should be such that this difference is measured. Failures occurring due to lack of coverage should always be included in scoring calculations and weighted appropriately to reflect the true customer experience.

4.4 Comparing networks with differing technology use

Network evolution and the adoption rate of new technologies often varies between operators. Benchmarking should be performed in such a way that it incorporates the use of the latest technology available. This is to reflect the network capability and customer experience available with the latest devices. Benchmark scoring should account for Operators who offer performance differentiation through early adoption of new technologies by way of a 'bonus' for such deployment.

EXAMPLE: If operator A deploys 5G technology whilst operator B continues to deploy 4G technology, the benefits 5G technology offer to the customer experience should be captured in the Benchmarking data collection and scoring.

4.5 Test device selection

Mobile network benchmarking is performed mainly using drive testing. This relies heavily on the choice of test device(s). Care should be taken in the selection of such devices to ensure they do not favour one Operator's network over another in the results. The same devices may perform differently on two different networks depending on factors such as the antenna placement in the device for varying frequency bands, variations due to manufacturing tolerances, firmware version differences, modifications made to devices for metric data collection and device placement and mounting in the test vehicle.

4.6 Test server selection

Data tests are commonly performed to a test server or selected web page (or pages). The selection of such servers/sites can influence the benchmarking result. Test servers should be selected so they do not favour one network compared to another. Web pages should be selected such that they represent a cross section of pages commonly used by customers.

EXAMPLE: If Operator A hosts the sever selected for 'ping' testing and the same server is also used to test Operator B, it is likely that performance levels for Operator B will be worse than those for Operator A due to the difference in latency to the selected server. This miss-represents the performance difference for this metric. Such situations should be avoided.

4.7 Test method transparency

Given the importance of the clear interpretation of benchmark results, all results should be accompanied by a declaration containing information about the following:

1) The scoring model/methodology used including all coefficients, targets and weightings.

2) The underlying KPI values as measured in the test.

3) The number of samples collected or number of tests performed for each KPI measured for each sub category.

4) The test methodology used including details of equipment setup, call sequences, test servers and web pages.

5) The areas/routes used for the data collection.

6) The device model and firmware version used for the data collection.

7) The tariff/data plan used for the data collection.

The intention of this is to provide the transparency required so that parties receiving the results are able to understand them fully. All factors required for this understanding should be provided.

4.8 Advice and best practice for web-page selection

Web page selection can impact on webpage load test results. To ensure a representative performance comparison can be made the following information and advice should be considered:

- For sufficient diversity and robustness of results, a minimum of 6 different pages is recommended to be considered for the scoring. It is good practice to measure more pages (e.g. 10), to retain enough diversity in case the dynamic behaviour requires to eliminate certain pages from the overall result.
- It is recommended to select pages according to their relevance to end customers. Preferably, public information of popular ranking per country is used and referenced. If possible, pages should be selected from Top 50 list, where an extension of that range is justifiable if not enough suitable pages exist within the Top 50.
- In case the download of a predefined data amount is used as success criteria as described in ETSI TR 103 702 [i.15], all pages should exceed a minimum size to cover the minimum amount of data The page size needs to be observed on a daily basis throughout the measurements. In case of the severe size changes, a reaction may be needed.
- Internationally popular live pages and country dependent pages may be used in reasonable proportion (e.g. 10 live pages - 4 are common, 6 are country dependent).
- Ad blockers should not be used.
- A web-page selection that is hosted pre-dominantly by one CDN should be avoided.
- Websites of services that are predominantly accessed via a dedicated app on a smartphone should not be selected. For example, Facebook™, YouTube™ and similar websites/services are typically not accessed via a mobile browser and should therefore not be used as websites for HTTP Browsing tests in mobile benchmarking campaigns.
- No website should be selected that is a sub-page/site of another already selected website.
- No website should be selected where the content is legally suspicious or contains harming, racism or sexist content.

5 General Description

In the present document the benchmarking and scoring of networks over a large geographical area, e.g. entire countries in various modes and for diverse services provided by mobile networks is described. A comprehensive manner to compare the tested networks is to calculate an overall score per network based on the individual measurement results collected during a test campaign. The individual measurement results are aggregated using a weighted accumulation into an overall network score. This overall score finally allows the scoring of the tested networks. To arrive there, the weighted aggregation is done over several layers.

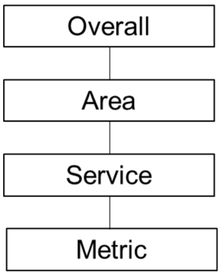


Figure 1: Aggregation layers

Weights are used for the aggregation of the different metrics, mobile services and areas to obtain the final score.

The accumulation of the measurements is done over several levels. The first or lowest layer consists of the measurement metrics for the services delivered over the mobile network. The services or applications considered are telephony, video, data transfer and services including browsing, social media and messaging. The metrics collected for one mobile service and a certain area are aggregated into an individual score for each metric; the scores of the metrics are then aggregated into an overall score of the mobile service.

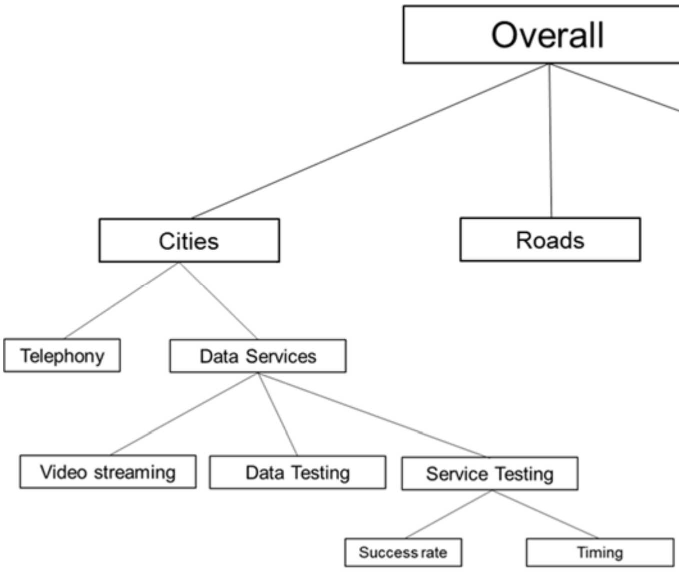


Figure 2: Aggregation over services and layers for a mobile network

In this aggregation, the metrics have a score weight according to the weight they were given for that particular mobile service. The scores for the individual mobile services are then in turn aggregated into a score for telephony and data services, and then together for the area they were collected in.

Finally, the various areas are weighted and accumulated over the various areas covered in the measurement. The different areas can have further geographical subdivisions. The weighted aggregation of the areas results in an overall score that characterizes the network.

6 Test Areas

6.1 General

The choice of the areas to be tested are an important part of the test setup. In order to be representative, the areas have to cover a majority of the population and main areas of mobile use; in case of limited countrywide coverage a representative proportion of the covered population. Drive testing is the method of choice but can be supplemented by walk testing in designated areas.

In the choice of areas and the distribution of time between individual subdivisions such as big cities and roads the geographical and topological properties of the respective country need to be considered. This may impair, to some extent, the comparability between countries. The aim should be that the chosen sites are appropriate for the respective country under test.

In order to be representative or to paint a more detailed picture, the areas of test such as cities and roads can be supplemented by measurements in trains and hot spot locations.

To maintain comparability, test areas that are not covered by all the networks under test need to be considered appropriately. In general, limiting the tests only to areas that are served by all networks is certainly the first choice, but in case important parts of the country and population would not be tested, the respective operator that does not cover these areas can be excluded from the countrywide testing or the limitations need to be included in the overall scoring.

The various areas need to be tested in an appropriate manner. Since some areas might not be accessible by drive testing, walk testing can be considered.

6.2 Geographical divisions

6.2.1 Cities

Cities are varying in size and density and the categorization of big, medium and small cities varies by country. The city size and importance are sometimes reflected in requirements set by the spectrum licensing authorities. The cities can be, but do not necessarily need to be, divided up into three categories, namely big cities, medium cities and small cities.

The big cities are defined as the major cities of a country from the population and commercial point of view, e.g. high rise buildings and high density of population are found in the big cities. Most of the hot spot areas are found in the big cities. Testing big cities means driving the main roads including tunnels and bridges.

Medium cities are smaller cities than the big cities with less inhabitants and less commercial importance. Occasionally they have high rise buildings and in general the density of the population is lower than in big cities.

Small cities or towns have fewer inhabitants than medium cities and have an even lower commercial importance.

The choice of the possible subdivision and distribution in defining city types is to reflect their relevance on the countrywide scale.

6.2.2 Roads

The highways are multi lane roads that can carry high traffic and connect big and medium cities of the test area. They are going across the country and have no intersections or traffic lights. Tests performed on city highways that are within a big or a medium city are counted in the results for cities rather than roads.

Main roads are roads that carry high traffic and connect cities of the test area. These roads may have traffic lights and intersections. The main roads that are driven within cities are counted for the cities.

Rural roads are roads that do not carry high traffic and connect medium and small cities. They can run through open landscape and can also cover dispersed settlements.

6.2.3 Complementary areas

Complementary tests, if appropriate, vary from country to country. E.g. trains and railways are established locations for tests in countries with strong commuting or highly frequented intercity connections whilst in other countries trains can be disregarded.

Other hot spots of use such as train stations, airports, pedestrian zones, parks, stadiums or tourist attractions are locations frequented by users of mobile phones. Those areas are to be considered appropriately.

7 User Profiles

Different users have different requirements and expectations with regards to mobile services. These expectations are the basis of what is perceived as excellent, good or poor. In addition, the type of service that is requested might differ between different user groups whom each put a different emphasis on the various service aspects like telephony, video, data or other social media. These groups can be assigned different subscription profiles, however, for the purpose of a network comparison, the best or highest commercially available profile yields the results that represent the performance of the network in best fashion. Using standard or budget profiles may produce interesting insights in the services received by the respective subscriber but are not in the position to assess what the network is capable of.

8 Test Metrics

8.1 Introduction

The test metrics are, with a few exceptions, generally defined in ETSI TS 102 250-2 [i.1] therefore the tests are whenever possible referenced to that document. In ETSI TS 102 250-2 [i.1] a success ratio is in most cases a two-step metric divided into successful access and successful conclusion.

These can be by weighted calculation aggregated into a single value, possibly even incorporating additional metrics or criteria that are decisive for the user's perception of a working service. This applies to all occurrences of Success Ratio in clauses 8.2 to 8.5.

The video, data throughput and service testing are often summarized as data tests as opposed to telephony tests, this separation is not excluded.

8.2 Telephony

8.2.1 General

Telephony tests are tests with a fixed call length where two terminals, either both mobile or one landline and one mobile call each other. Landline connections usually do not support new higher codecs such as AMR WB or EVS. In order to measure these codecs mobile to mobile circuit switched calls are necessary and at times even VoLTE calls over packet switched are needed. To consider unsustainable quality in a call, for a low speech quality score (e.g. MOS < 1,6) or silent periods for consecutive measurement samples (e.g. > 20 s), the call can be counted as unsustainable, and as an unsuccessful call or treated by a separate indicator. The proposed QoS and QoE parameters are not tied to any technology, they are obtained on user plane and can be obtained e.g. for VoLTE, VoNR, OTT VoIP telephony or circuit switched voice calls.

8.2.2 Telephony Success Ratio

The success ratio of the voice service independent of access or relay technology is the Telephony non Accessibility and the Telephony Cut-Off Call Ratio in ETSI TS 102 250-2 [i.1], clauses 6.6.1 and 6.6.5. For the purpose of the present document, the Voice Over LTE (VoLTE) service is treated as a telephony service.

8.2.3 Setup Time

The setup time for voice calls is defined in ETSI TS 102 250-2 [i.1], clause 6.6.2. It starts with the initiation of the call and ends when the alerting of the called side is indicated. Alternatively, the time when the acceptance or successful setup of the call is signalled to the user can be used as the end trigger. Triggerpoints based on initiation of the call, the indication of alerting or the successful setup taken at user interface or application layer may differ in time from triggerpoints on signalling layer due to the processing time of the device's software and operating system.

8.2.4 Listening Quality

The value is calculated on a per sample basis as described in ETSI TS 102 250-2 [i.1], clause 6.6.4 where Recommendation ITU-T P.863 [i.9] in FB mode is recommended to be used. The measurement is set up according to ETSI TR 103 138 [i.6] and Recommendation ITU-T P.863.1 [i.10]. In case it can be guaranteed that there is pure IMS calling (e.g. VoLTE) or OTT VoIP connections are used without transcoding or re-packaging, listening quality can be obtained by Recommendation ITU-T P.565.1 [i.17].

8.3 Video Testing

8.3.1 General

Video testing is in the standard case IP based video streaming. Video streaming quality of service aspects can be found in ETSI TS 102 250-2 [i.1] and in ETSI TR 101 578 [i.4]. For the purposes of the present document the Smartphone app based testing as in Figure 1 of ETSI TR 101 578 [i.4] is used. In order to collect details of the transport and reproduction, the length of the observation period of the video should reflect the relevant delivery mechanisms and the typical usage profile of a mobile user.

8.3.2 Video Streaming Service Success Ratio

The video streaming success ratio is the end-to-end success ratio of the requested video stream. It starts with the request of the video and ends with the end of the playout. This is derived from the metrics in ETSI TR 101 578 [i.4] as a combination of Video Access Failure Ratio and Video Playout Cut-off Ratio.

8.3.3 Setup Time

The setup time is the time from stream request to the display of the first picture and start of playout. This is Video Access Time from ETSI TR 101 578 [i.4].

8.3.4 Video Quality

The quality of a video reproduction is determined by freezing, frame-rate, resolution and compression depth and scheme by the codec. Freezing is most common and annoying impairment experienced by the user. The handling of freezes is described in clause 4.5.4 in ETSI TR 101 578 [i.4].

A comprehensive measure for the perceived quality that combines the impact of the above mentioned parameters is the Mean Opinion Score (MOS) scale and is done according to clause 6.5.8 in ETSI TS 102 250-2 [i.1]. In the case of video streaming with a respective app on the smartphone an encrypted stream and a range of different resolutions (up to HD) is expected. The Video Quality parameter in ETSI TR 101 578 [i.4] reflects such measure. In addition to this, Video Freezing Time proportion in ETSI TR 101 578 [i.4] provides an insight about the proportion of the accumulated video freezing duration in relation to the actual video playout duration.

8.4 Data Testing

8.4.1 General

For data testing the throughput bandwidth for the user is tested. This is done by downloading and uploading incompressible files over HTTP. In clause 6.8 of ETSI TS 102 250-2 [i.1] the up and download of entire files is described. The description of an upload and download using fixed duration is in clause 5.2 of ETSI TR 102 678 [i.5]. Both approaches can be used, either alone or combined, for the purpose of evaluating throughput bandwidth. Both described test setups are named HTTP Browsing in the references; however for down - and uploading HTTP files, a legacy HTTP server and client is nowadays used instead of a web browser for achieving the highest possible throughput to the user.

Testing data rate or data throughput refers typically to transport layer as TCP or to IP. To initiate TCP connections, a higher layer test scenario as e.g. a HTTP session is used. There are other test scenarios using FTP or SMTP utilizing TCP transport too. HTTP/TCP is used as most common test scenario, while others like FTP/TCP become less relevant for retrieving information on TCP level.

Note that HTTP is typically associated with TCP as transport protocol. However, HTTP/3 additionally supports the UDP-based QUIC [i.11] protocol. The use of UDP/QUIC depends on server and device capabilities.

8.4.2 Success Ratio

The determination of the success ratio for HTTP uploads and downloads of entire files is included in clause 6.8 of ETSI TS 102 250-2 [i.1] and in clause 5.2 of ETSI TR 102 678 [i.5].

8.4.3 Throughput

8.4.3.1 File-based and fixed duration throughput

The determination of the mean data rate or throughput for HTTP uploads and downloads is included in clause 6.8 of ETSI TS 102 250-2 [i.1] and in clause 5.2 of ETSI TR 102 678 [i.5]. Throughput can be obtained in up- and downloads of entire files. The achievable throughput depends on the number of active, parallel file transfers and it can result in different results depending on the transfer time of individual files. File based throughput measurements are preferably based on up- and download of single files to achieve comparable test situations and to emulate the transfer of single files as typical in many use-cases. The best view of the actual achievable transfer data rate to a single user is provided by a multi-threaded, fixed duration HTTP up- and download test from and to an HTTP server. The measurement should be stopped before any of the files transferred in parallel is entirely transmitted. This enables a constant number of active transfer threads during the test. The methodology introduced in the following clauses applies the same for download and upload measurements.

Measuring data rate on TCP level provides a throughput value as delivered to the user based on the TCP payload. Hence, it does not consider TCP and lower layer headers neither SYN/ACK packets. This throughput value is usually lower than the amount of data transported over the physical radio channel.

8.4.3.2 Throughput in fixed duration tests

To provide testing-time predictability, fixed-duration testing - as opposed to fixed-size file transfer - is often used to obtain the maximal achievable data throughput delivered to a single user. Also, to get a better estimation of the overall data rate available to a single mobile device under given network conditions, multi-threaded data transfer is used in those tests to overcome possible limitations for a single-threaded transfer.

Data throughput obtained in time based, fixed duration tests is often considered as single user throughput under given network conditions. However, the throughput calculation as defined in ETSI TR 102 678 [i.5] also covers initial ramp-up of the data channel, where the finally achievable throughput is not reached yet. The influence of this initial lower data rate on the throughput calculated over the entire fixed duration depends on the chosen test duration. Especially, for short test durations the throughput is influenced by the duration and data rate of the channel ramp-up. Note, this single user throughput should not be mistaken for network capacity or single user data capacity.

8.4.3.3 Sustainable throughput in fixed duration tests

To overcome the influence of data channel ramp-up, this initial phase can be excluded from the throughput calculation. The initial trigger point for the throughput calculation is chosen after ramp-up is finished or after a short waiting period. The data throughput is calculated only in a transfer phase where the transport is actually in a stable state and a saturated throughput is assumed. These sustainable throughput measurements are independent from test duration and provide results closer to typical speed test applications.

8.4.3.4 HTTP/TCP throughput and UDP IP-capacity in fixed duration tests

Data testing and throughput measurements are historically derived from browsing tests based on HTTP/TCP. To obtain the actually achievable throughput for up- and download HTTP/TCP or other reliable protocols are typically used.

As UDP is becoming more dominant as a transport protocol, especially for media applications, data delivery capacity based on UDP is complementing legacy TCP throughput measurements.

However, the amount of data transported and available to the user via UDP can differ from using TCP due to network settings as well as the dependency of TCP from the performance in the return channel.

In case of using UDP for testing data delivery performance on the device, differences to TCP based measurements should be considered. UDP is an unreliable protocol, there is no implicit re-transmission and there is no additional traffic by SYN/ACK packets. If applying UDP data transfer measurements, usually the UDP header counts as transmitted data. Consequently, the measured data rate can be considered as gross IP capacity that is closer to the transmitted data on physical layer compared to TCP throughput.

Those IP capacity measurements using UDP should be considered as describing network performance too. A method for UDP IP capacity measurements is described as UDPST in Recommendation ITU-T Y.1540 [i.13], Annex B.

Another, commonly used but not standardized tool for IP capacity measurements is named iPerf. The open source iPerf library is described and available under [i.18].

8.4.3.5 Data latency

Data latency and latency variation plays an important role for perceived quality of real-time and interactive applications [i.15]. There are several approaches to obtain latency information like e.g. Ping RTT or SYN/ACK in TCP connections as defined in clause 6.3 of ETSI TS 102 250-2 [i.1]. Latency values can also be obtained directly accessing the user interface while running a real interactive application by motion-to-photon evaluation [i.16].

However, latency is not constant over time and varies. Therefore, a series of individual latency values should be measured to receive a sufficient amount of values for valid statistics of latency and latency variation. Furthermore, latency values should be obtained under realistic load situations in the data channel to be tested.

The most recent approach is defined in Recommendation ITU-T G.1051 [i.14], where the latency is integrated over time and combined with delay jitter and packet loss and forms an Interactivity Score.

8.5 Services Testing

8.5.1 General

Besides the browsing of web pages as in clause 6.8 of ETSI TS 102 250-2 [i.1], services like social media and messaging systems (SMS is not considered) are not described or standardized for mobile testing. Some overall interesting aspects of all of these services are the success ratio and the duration or timing of the interaction.

8.5.2 Services

8.5.2.1 Browsing and web-content delivery

For web browsing tests web pages are accessed and downloaded. These pages are preferably popular dynamic pages. The respective browsing metrics are in clause 6.8 of ETSI TS102 250-2 [i.1] and ETSI TR 103 733 [i.12]. For dynamic web pages a success criterion can be defined by the time it takes until a predefined data volume of the overall page session is received as in ETSI TR 103 733 [i.12].

8.5.2.2 Social Media

For social media like Facebook™ and Instagram™, an action such as posting of pictures, text and video is the typical activity that is tested.

These are the usual activities where the user interacts with the media application. Popular social media vary in their popularity over time and across countries, therefore the list of services and their weight in the calculation can change. Since at the time of publication of the present document there are no standardized metrics for the use of these services over mobile networks, the metrics cannot be referenced to any document. In cases where interfaces (API) exist to those applications, these can be used to test the respective service.

It should be highlighted that the perceived user experience depends on the tested service platform in addition to the mobile network performance, because the observed timings inevitably include processing time on the service platform. As such, it depends on the focus of the testing activity as to whether inclusion of such services is useful or not.

8.5.2.3 Messaging

Sending a text message, line and measuring the delivery time and success ratio is a convenient way to characterize the perceived quality of service. For these services, delivered over a mobile network, no standard for quality of service metrics exists.

Although there are no standardized metrics for social media and messaging services delivered over mobile network, metrics based on legacy messaging services can be established.

8.5.3 Success Ratio

In general, the successful conclusion of an activity is to be measured in social media and messaging. The number of successful trials versus the number of trials is the success ratio.

$$\text{Service Success Ratio [\%]} = \frac{\text{number of successful activities}}{\text{number of trials}} \times 100$$

An activity starts with triggering an action on the device by e.g. pushing a button to send a text message, to open a Facebook™ profile, posting a picture on Instagram™ or opening a web page. The activity is successful when the application indicates a confirmation that the triggered process is successfully concluded. This can be done, for example, by a graphic indicator like a check or by other means.

8.5.4 Timings

The duration of a social media or messaging activity is the time between triggering the activity and the indication of the successful conclusion of it. In the case of browsing, social media and for messaging, it is the time until confirmation of successful reception is indicated.

$$\text{Service activity Duration [s]} = \mathrm{t_{end}} - \mathrm{t_{start}}$$

The timing, the duration from the initiation to the successful conclusion of a test depends to a significant extent on the performance of the underlying web service. However, these factors are the same for all networks under test.

9 Weighting

9.1 General

In order to achieve an overall score, the individual test results for the various areas have to be given a weight. This weight is the importance with which the result enters into the overall valuation of the testing. The weighting of the results is done on each level of aggregation (Figure 1).

In the following clauses the general method of weighting and the individual measures are described.

Example values that are used in practice for the weighting of the areas and tests as well as actual values for the upper and lower limits of the target ranges are presented in the annexes to the present document.

9.2 Areas

For an area, all regional, daytime, geographical or morphologic categories are considered, where the scoring method is applied before further aggregating to an overall score. These different categories that are measured have a combined weight of 100 %, in case e.g. there are no complementary areas, cities and roads have alone a combined weight of 100 %. In case the areas have further subdivision, these areas are individually weighted and then make up 100 % of the next level e.g. if city category is subdivided into big cities and small cities. These two subareas add up to 100 % representing the whole weight of cities.

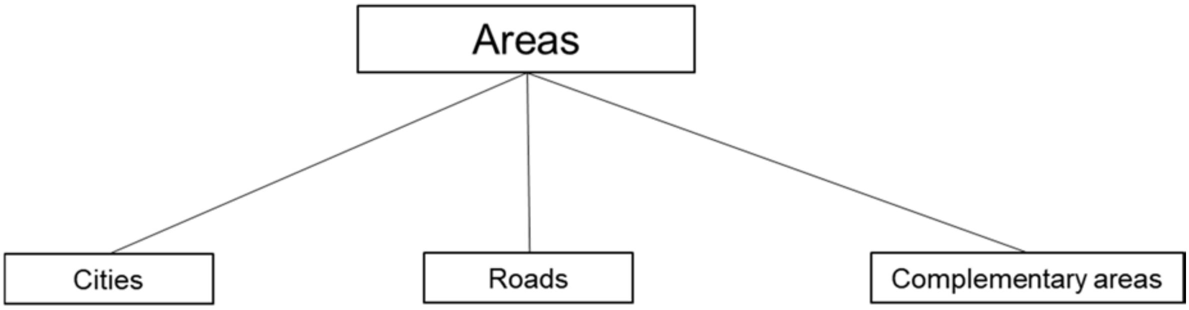


Figure 3: Examples of areas

The timing of tests can have an impact on the perceived performance on the network. Despite this, the weighing of different times is not considered part of best practice. It is advisable, however, that measurements are reasonably spread over the different times of day, e.g. to not deliberately exclude busy hours.

9.3 Tests

9.3.1 General

Each test is multi layered in nature. The upper layer provides the overall score of the mobile service tests, which is calculated from the weighted scores of the test scenarios for telephony and data services. The two scenarios have the combined weight of 100 %. The data services in turn consist of video streaming, data testing and service testing. The three types have also the combined weight of 100 %. The weight of the individual test types can be determined according to the intended user profile.

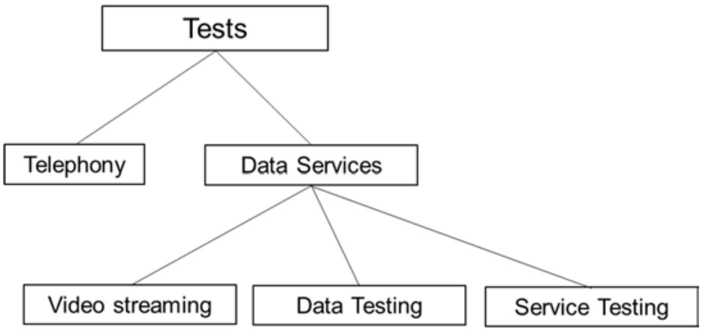


Figure 4: Service types for testing

The test metrics are evaluated as aggregated values. While the success ratio is aggregated already, for most of the other values such as listening quality, throughput, setup time, and duration etc. the average is taken into account. These individual metrics have a minimum and a maximum value. However, the metrics do also have a bad limit to saturation area in which the experience of the customer does not deteriorate significantly and a good limit to saturation area above which the customers experience does not improve further. The average values are expected to be between the good and the bad limits.

The scoring of the individual aggregates can be increasing or decreasing. If the score rises with the value, then it is an increasing value score. Starting from the minimum below the bad limit the value score is at 0 %. Between the bad and the good limit, the score is increasing to 100 %. In the saturation area between the good limit and the maximum, the value score stays at 100 %.

If the value score rises with the decreasing value, then it is a decreasing value score. In the saturation area between the good limit and the minimum value, the score is stable at 100 %. Between the good and the bad limit, the score decreases to 0 % and stays there above the bad limit.

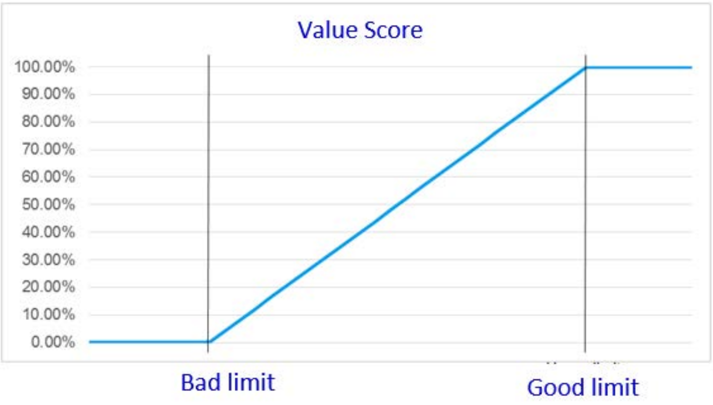


Figure 5: Weighting function

The general formula is:

$$\text{Score} = \frac{\text{value} - \text{Bad limit}}{\text{Good limit} - \text{Bad limit}} \times \text{weight}$$

NOTE: In the case of a decreasing value score, two things are expected:

1) the bad limit will be a higher numerical value than the good limit;

2) the resulting negative/negative calculation is expected and produces a positive result.

The scores for the average values are calculated between two limits. The negative impact of poor performance or the positive impact of excellent performance can be underrepresented by taking only scored averages into account. The aspects of the distribution of the results need to be taken into account. It is therefore useful to introduce limits for poor and excellent performance and calculate the percentage or percentile of results within these limits. In order to boost superior performance, an extra bonus can be applied, similarly the achievement of minimum performances can be awarded too.

In the given graph in Figure 5, a linear function is given to illustrate the general method, although not all service measurements are perceived in a linear manner. A non-linear relationship may occur where an increase or decrease in metric value at one end of the scoring interval is perceived to have a much larger or smaller effect than a similar increase or decrease at the other end and hence should have a higher or lower percentage of the available score. In this case non-linear functions may be applied to determine the score value. There are a number of functions which could be utilized to measure the non-linear score with square root, logarithmic or logistic sigmoid function being typically and widely used at this time. A working example for the application of a non-linear weighting is given in Annex B of the present document.

9.3.2 Telephony

9.3.2.1 General

The telephony service has three major aspects:

- overall success ratio;
- setup time; and
- listening quality (MOS).

These three values enter into the calculation of the overall score of the telephony service. The individual aspects can then in turn be weighted individually for the calculation of the overall telephony score. These factors have a combined weight of 100 %.

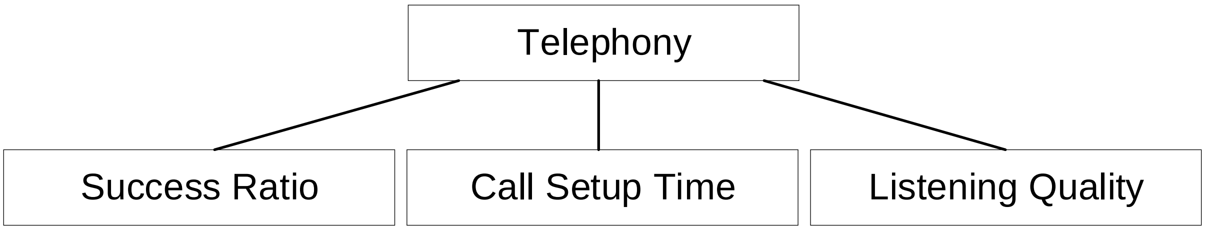


Figure 6: Contributing dimensions to telephony

9.3.2.2 Scoring

The higher the call success ratio and the MOS value the better is the experience of the user; the two values have an increasing value score. While the longer the call setup time is, the worse the experience of the user is; the call setup time has a decreasing score value. In addition to the consideration of the average values, extra bonus for excellent listening quality or very short call setup-times can be given, the same as extra bonus for the reduction of very bad experiences as very low MOS scores or very long call setup times. As examples, the 10th percentile of CST can be used for awarding very short CST and the 90th percentile for awarding excellent listening quality. To award the absence of negative experiences the 90th percentile of CST can be considered and a ratio of MOS < threshold (e.g. MOS < 1,6 or MOS < 2,2). A lower parameter value would lead to a higher score in these cases.

For telephony the following factors with example thresholds and percentiles can be taken into account:

- Call Setup Success Ratio.
- Call Drop Ratio.
- MOS.
- MOS < low MOS threshold.
- 90th percentile of MOS.
- Call Setup Time.
- Call Setup Time > long setup time threshold.
- 90th percentile of Call Setup Time.

9.3.3 Video streaming

9.3.3.1 General

The main aspects of the video streaming are the video streaming service success ratio, setup time (video access time) and the visual quality. These three factors are combined to the give the video streaming score together with extra bonus for superior performance values for selected metrics. All factors have the combined weight of 100 %.

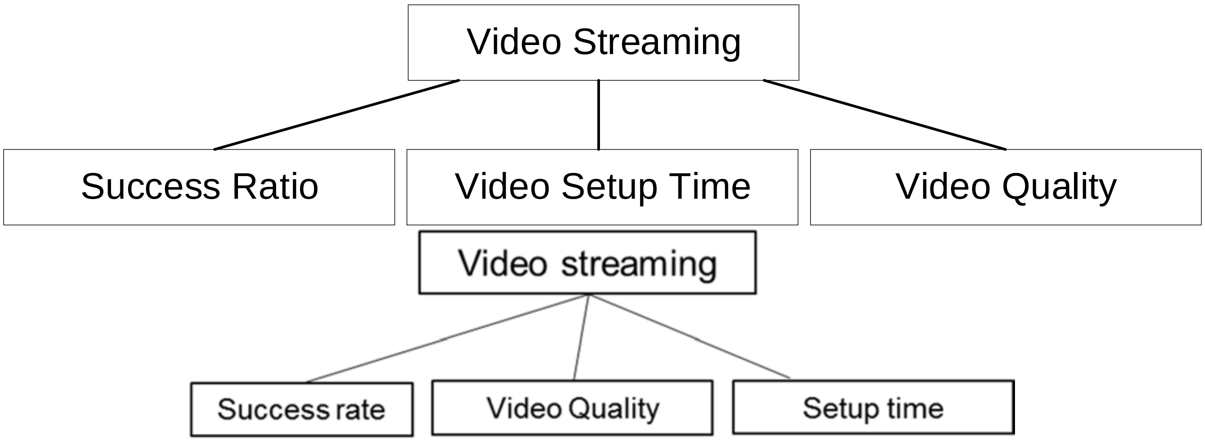


Figure 7: Contributing dimensions to video streaming

9.3.3.2 Scoring

The higher the video streaming service success ratio and the video quality MOS value the better the experience of the user is. The two values have an increasing value score, while the longer the time to first picture display and video start picture (that is video access time) is the worse is the experience of the user is; the setup time has a decreasing score value. The negative impact of bad video quality is represented by e.g. 10 percentile of video quality MOS between two limits and the negative impact of long setup times is represented by the percentage of video access time is above e.g. 10 s. Any impact of particularly good performance is not taken into account for video streaming since there is a very high proportion of HD expected that does not leave much headroom for technical improvement that can be rewarded with a bonus. For video streaming, the following factors with the proposed thresholds and percentiles can be taken into account:

- Video Success Ratio.
- Video Quality.
- Video Freezing Time proportion.
- Resolution.
- 10th percentile of Video Quality.
- Video Access Time > long setup time threshold.

The streaming success ratio together with the quality measures such as MOS, freezing, resolution and video access time can be combined to define a composite success criterion, with minimum requirements for the quality metrics. In this case, only video sessions are scored for quality aspects, which succeed in the composite success criterion.

9.3.4 Data Testing

9.3.4.1 General

The main aspects of data testing are the success ratio and the data rate or throughput. These two factors are combined to produce the data testing score. These have a combined weight of 100 %.

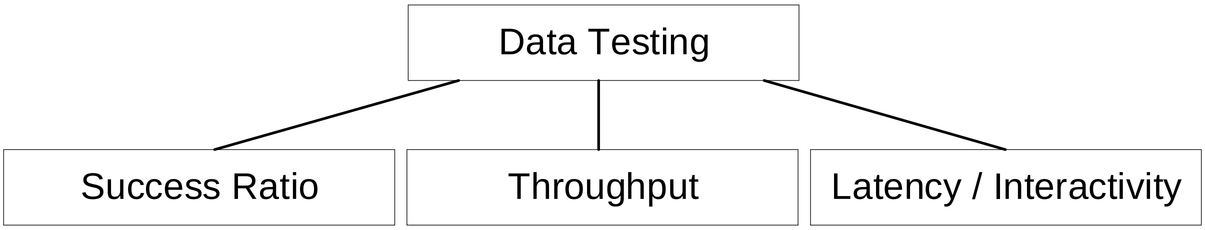


Figure 8: Contributing dimensions to data transfer testing

9.3.4.2 Scoring

The higher the success ratio and the throughput value the better the experience of the user is; the two values have an increasing value score. The negative impact of low throughput values is represented by the e.g. 10th percentile and the positive impact of high throughput is represented by e.g. 90th percentile. The Average Session Duration has a decreasing value score.

For data testing, the following factors with the proposed thresholds and percentiles can be taken into account. These apply for both uplink and downlink:

- Transfer Success Ratio.
- Average Session Duration.
- Average throughput.
- 10th percentile of (low) throughput.

- 90th percentile of (high) throughput.

To obtain Transfer Success Ratio and Average Session Duration preferably results from tests up- and downloading entire single files are used. Average and percentile throughput values are preferably obtained from time based, fixed-duration tests. A split into HTTP/TCP throughput and UDP based IP capacity values is possible.

For scoring data latency main QoS parameter to be considered is a rating score as in Annex A of Recommendation ITU-T G.1051 [i.14]:

- Interactivity Score.

Alternatively, can be used:

- Median of the individual two-way latency values as mean latency.
- Latency variation as e.g. derived from Packet Delay Variation (PDV).
- Ratio of lost packets and packets disqualified due to long delay.

9.3.5 Service Testing

9.3.5.1 General

The main aspects of the services are the success ratio and the timing or duration. These two factors are combined to determine the service score for browsing, social media and messaging. These aspects have a combined weight of 100 %.

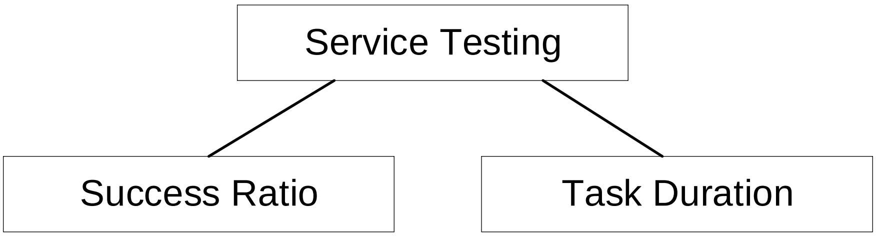


Figure 9: Contributing dimensions to data service testing

9.3.5.2 Scoring

The higher the success ratio the better the experience of the user. The success ratio value has an increasing score value. The timing or duration of an activity such as browsing or posting follows a decreasing function. The longer it takes the worse is the experience is. The negative impact of long activity duration is represented by the percentage of times above a long duration threshold.

For service testing, the following factors are taken into account:

- Activity Success Ratio.
- Average Duration.
- Activity Duration > long duration threshold.

For social media and messaging services only the activity duration samples of sending text message and sharing single picture can be taken into the factor calculation.

For Social Media and Messaging service testing, the following factors can be taken into account:

- Activity Success Ratio.
- Average Duration.
- Activity Duration > long duration threshold.

10 Statistical confidence and robustness

10.1 General

When comparing mobile networks, through benchmark metrics and scoring means, it is important that the statistical significance of the outcome is considered. This is especially true when using the results to conclude if one network is more preferable than another. The following clauses outline important considerations and provide one method to evaluate the validity of conclusions from a statistical viewpoint.

10.2 Influence of the derived scores on statistical confidence

The performance of a given network is estimated based on a set of measurements. These measurements are collected in certain geographical locations, at certain times on the dates selected to represent the general experience of customers under real-field load conditions. The measurements from any measurement campaign only represent a subset of measurements from the overall population.

The measured attributes and the derived metrics/scoring are subject to uncertainty. Two questions arise. How closely do the results derived from the measurements represent the performance experienced by the entire population? How repeatable are the results, it means how sensitive are the results to changes in the measurement points within the same basic population?

In general, the larger the sample set the less the uncertainty of the results, and the better representation of the population distribution of measurement results. Therefore, it is necessary that the sample set is selected to include samples from the various different environments that exist for the overall population. The uncertainty of the indicators and the derived score requires statistical analysis and is usually described by confidence intervals.

It should be mentioned that a particular measured attribute will have an individual confidence interval which may result in a stable average with more or less samples than another attribute. Contributors with low confidence typically drive and decrease the confidence of a final aggregated score. This should be considered when defining and scaling measurement campaigns. Low confidence stands for larger statistical confidence levels.

Success or failure ratios, in particular, require a sufficient amount of measurements to support a confident conclusion. For example, in a collection of 100 calls, one dropped call more or less will lead to a change in the Call Drop Ratio of 1 %, in case of 1 000 calls the Call Drop Ratio will change by 0,1 % by a single dropped call. The resolution of the Call Drop Ratio is defined by the number of measurements from which it is derived. If the actual Call Drop Ratio is within the range limited by the measurement resolution, deriving a reliable representation is difficult. In such cases measurement sample volumes should be increased. The minimum number and the targeted confidence interval for benchmarking campaigns highly depends on the purpose of the campaign.

10.3 Statistical confidence level estimation

10.3.1 General

This clause presents a pragmatic, empirical approach to assessing basic statistical properties of network benchmarking score results and deriving confidence levels. It will answer these main questions:

- How close can one expect the unknown results of the basic population of the observation area to be to the results and score of the obtained measurement result?
- How repeatable are the obtained measurement results and score when using different measurement points of the same basic population?

The following method describes an established method for deriving a confidence interval and other statistical metrics of the scoring result. The derived confidence interval is based on statistical evaluation of the measurement results and is a pure statistical representation of the measurement samples. It does not reflect the significance in human perception.

10.3.2 Statistical analysis using a bootstrap resampling method

The presented method describes a probability density (and distribution function) and the related statistical properties (average, standard deviation, confidence intervals) of the achieved score of each competing network. This helps to interpret results and establish a means of determining the confidence of an achieved score from the collected measurement data.

The probability density function of a scoring result for a network or an area can be derived following a bootstrapping approach according to this pseudo-algorithm:

1) For each measurement contributing to the score, use re-sampling from empirical distribution (i.e. set of measured values) to generate a 'bootstrap sample' of equal sample size as the set of measured values.

2) Calculate relevant statistics (i.e. aggregated QoS Parameters, score) from the bootstrap sample and map the result to the score domain.

3) Repeat steps 1) and 2) for a sufficiently large number of times (N) to estimate sufficiently the distribution of the QoS parameters or the aggregated score.

4) Assess the statistical properties of the result, e.g. confidence intervals, standard deviation, etc.

The respective algorithm is performed for each aggregated QoS Parameter. It is advisable that the re-sampling, though random in nature, be executed in a way that preserves (within the scope of one bootstrap subsample) the correlation between QoS parameters originating from the same service session.

Assuming bootstraps of size N = 1 000 are being obtained for each aggregated QoS Parameter, means having collected 1000 individual measurement results like e.g. Call Setup Times in a campaign, and there are M = 10 000 of those bootstrap re-samplings, the resulting distribution of scores can be interpreted as the results of 10 000 hypothetical testing campaigns carried out under similar conditions, each resulting in 1000 measurement results in a different mixture.

Having obtained these hypothetical results (i.e. distribution of scores) for a network or an area the confidence interval of the score or even an individual contributor can be estimated. Further information can be found in Recommendation ITU-T E.840 [i.7] and Recommendation ITU-T P.1401 [i.8]. However, for many obtained values or QoS Parameters a Gaussian distribution is not given (e.g. data throughput). Here, other appropriate methods to derive statistical confidence values need to be applied.

Furthermore, there are dependencies in between measurement results or QoS Parameters belonging to one test. Preferably, re-sampling is applied under consideration of those dependencies.

10.3.3 Interpretation of results

The method described in the previous clause allows assessing empirically the statistical variation of the score and its statistical confidence interval. The statistical confidence interval also allows analysis of whether the difference between two scores (e.g. two areas, two networks, two different time ranges) is significant by certain probability (e.g. 95 %). The significance of such a difference is purely based on the statistical evaluation and will not give any indication of the significance in human perception of the networks' performance.

Furthermore, applied significance levels and comparisons depend on the purpose of the measurement and are not generalized or recommended in the present document.

Annex A: Example set of weighting factors, limits and thresholds

A.1 General

This annex provides a first example which represents a best practise at the time of release of the present document. The information here is intended to provide an illustration of how to practically apply network benchmarking and scoring as described in the body of the present document. In this regard it identifies example weights, limits and thresholds that could be applied to areas and mobile services as well as providing example worked network scoring calculations.

A.2 Area

A.2.1 Geographical divisions

A.2.1.1 General

The three areas can be weighted in the following manner.

Area	Weight
Cities	50 %
Roads	40 %
Complementary areas	10 %

A.2.1.2 City type

In case of three subdivisions of the cities a possible weighting is as follows.

City Type	Weight
Big cities	60 %
Medium cities	30 %
Small cities	10 %

A.2.1.3 Road type

The three types of roads can be weighted in the following manner.

Road Type	Weight
Highways	60 %
Main Roads	30 %
Rural Roads	10 %

A.2.1.4 Complementary areas

The two general walk tests can be weighted in the following manner.

Type	Weight
Trains	40 %
Hotspots (train stations, airports, pedestrian zones, parks, stadiums or tourist attractions)	60 %

A.3 Mobile services

Service Type	Weight
Telephony	40 %
Data Services	60 %

A.4 Test metrics of mobile services

A.4.1 General

For all the considered service types, the basic idea is to rate three aspects of the service, where possible:

- Service availability/retainability (e.g. success ratio, drop ratio).
- Service access time (e.g. video access time or call setup time).
- Quality of the media transfer (e.g. listening quality, time to present the webpage content).

Each of the aspects is described and scored by one or more quality indicators. Usually, there is one indicator for rating the average performance (e.g. average call setup time), other indicators rate superior or low performance (e.g. 10th and 90th percentile or the ratio of tests exceeding a certain threshold).

In this Annex A the limits and thresholds are adjusted to high performance networks and towards physical or perceptual limits. The weights of the service categories and individual QoS parameters haven been aligned e.g. by weighting the success ratio in the services always with 50 % regardless of the service itself, while access time and quality a weighted with 25 % each for voice telephony and video streaming. For the other data services (browsing and messaging), transfer time get 50 % as the only QoE indicator.

A.4.2 Telephony

Factor	Bad limit	Good limit	Weight
Call Setup Success Ratio	90,00 %	100,00 %	25 %
Call Drop Ratio	10,00 %	0,00 %	25 %
MOS	2,50	5,00	15 %
MOS < 1,6	10,00 %	0,00 %	10 %
Call Setup Time [s]	10,00	3,00	15 %
Call Setup Time > 15 s	3,00 %	0,00 %	10 %

A.4.3 Data Services

A.4.3.1 General

Data Service Type	Weight
Data Testing	30 %
Browsing	25 %
Video Streaming	15 %
Social Media and Messaging	15 %
Data Latency and Interactivity	15 %

A.4.3.2 Video Streaming

Factor	Bad limit	Good limit	Weight
Video Streaming Service Success Ratio (VSSSR)	80,0 %	100,0 %	50 %
Video Quality MOS	3,8	5,0	15 %
Video MOS < 3,8	10,0 %	0,0 %	10 %
Video access time [s]	5,0	0,0	15 %
Video access time > 5 s	10,0 %	0,0 %	10 %

A.4.3.3 Data Testing

Factor	Bad limit	Good limit	Weight
Transfer Success Ratio DL (single-file, 10 MB)	80 %	100 %	10 %
Average throughput DL (fix-duration) [Mbit/s] (see note)	(0,5)	(2 000)	14 %
10th percentile of (low) throughput DL [Mbit/s]	(0,5)	(200)	18 %
90th percentile of (high) throughput DL Mbit/s]	(400)	(2 000)	8 %
Transfer Success Ratio UL (single-file, 5 MB)	80 %	100 %	10 %
Average throughput UL fix-duration) [Mbit/s] (see note)	(0,05)	(500)	14 %
10th percentile of (low) throughput UL [Mbit/s]	(0,05)	(50)	18 %
90th percentile of (high) throughput UL [Mbit/s]	(10)	(500)	8 %
NOTE: Each individual data throughput measurement is transferred into a dimensionless 0-1 000 scale by a logarithmic weighting function. The average on this scale is used for weighted consideration in the score. The values in parenthesis provide the data throughput in Mbit/s as at 0 and 1 000 of the scale after transformation. Downlink: TP0-1 000 = 600 × log10(TPMbit/s + 46) - 1 000 Uplink: TP0-1 000 = 400 × log10(TPMbit/s + 5,6) - 300			

A.4.3.4 Browsing

Factor	Bad limit	Good limit	Weight
Activity Success Ratio (1 MB download < 5 s) (see note)	80,0 %	100,0 %	50 %
Average Duration [s]	3,0	0,0	50 %
NOTE: The web page completion criterion is set at the successful download of 1 MB of content in accordance with ETSI TR 103 733 [i.12].			

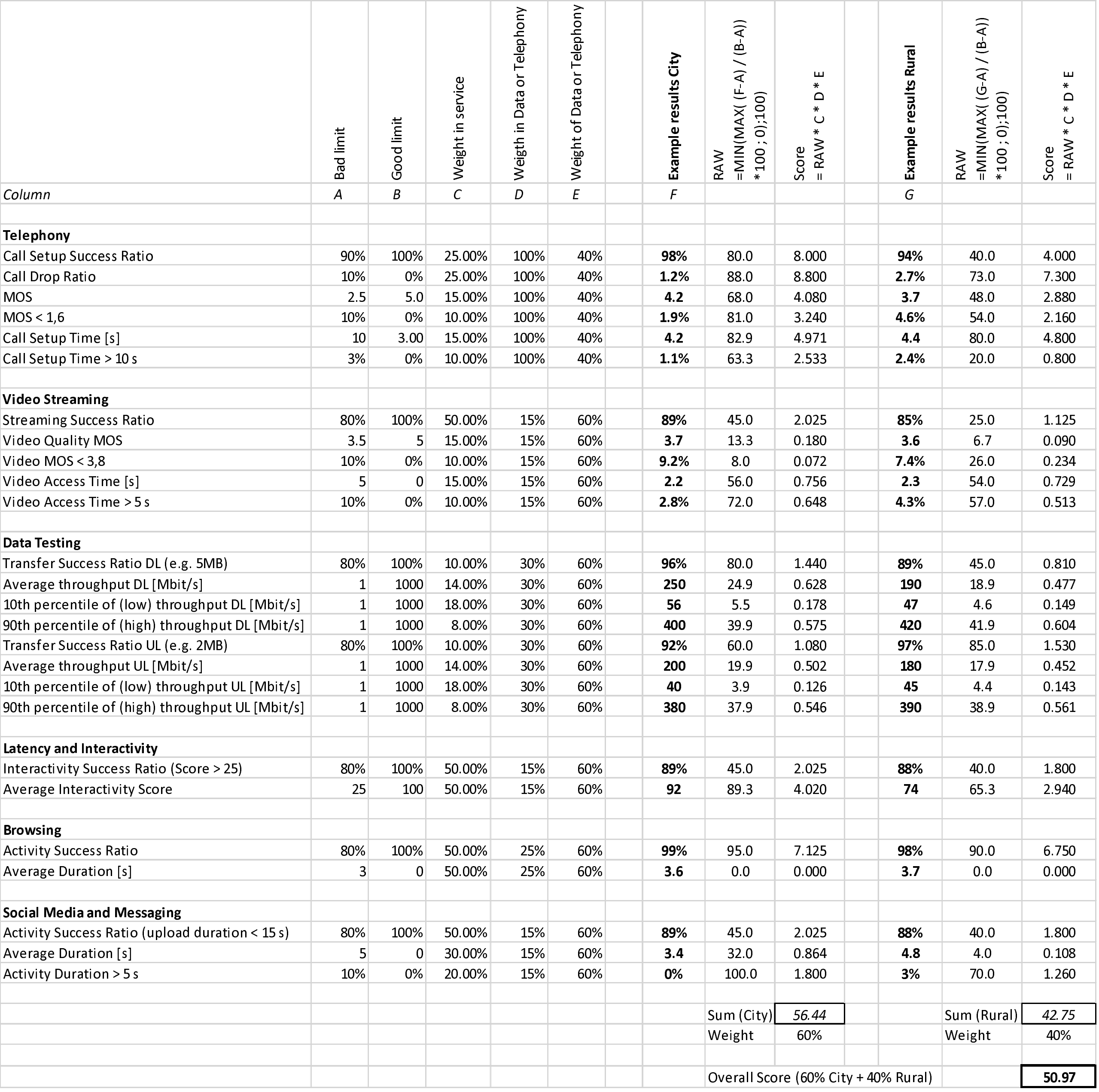
A.4.3.5 Social Media and Messaging

Factor	Bad limit	Good limit	Weight
Activity Success Ratio (upload duration < 15 s)	80,0 %	100,0 %	50 %
Average Duration [s]	5,0	0,0	30 %
Activity Duration > 5 s	10,00 %	0,00 %	20 %

A.4.3.6 Data Latency and Interactivity

Factor	Bad limit	Good limit	Weight
Interactivity Success Ratio (Score > 25) (see note)	80,0 %	100,0 %	50 %
Average Interactivity Score (see note)	25,00	100,0	50 %
NOTE: Score calculation according to Recommendation ITU-T G.1051 [i.14].			

A.5 Example Calculation



	Bad limit	Good limit	Weight in service	Weigth in Data or Telephony	Weight of Data or Telephony		**Example results City**	RAW =MIN(MAX((F-A) / (B-A)) *100 ; 0);100)	Score = RAW * C * D * E		**Example results Rural**	RAW =MIN(MAX((G-A) / (B-A)) *100 ; 0);100)	Score = RAW * C * D * E
Column	*A*	*B*	*C*	*D*	*E*		*F*				*G*		
Telephony													
Call Setup Success Ratio	90%	100%	25.00%	100%	40%		**98%**	80.0	8.000		**94%**	40.0	4.000
Call Drop Ratio	10%	0%	25.00%	100%	40%		**1.2%**	88.0	8.800		**2.7%**	73.0	7.300
MOS	2.5	5.0	15.00%	100%	40%		**4.2**	68.0	4.080		**3.7**	48.0	2.880
MOS < 1,6	10%	0%	10.00%	100%	40%		**1.9%**	81.0	3.240		**4.6%**	54.0	2.160
Call Setup Time [s]	10	3.00	15.00%	100%	40%		**4.2**	82.9	4.971		**4.4**	80.0	4.800
Call Setup Time > 10 s	3%	0%	10.00%	100%	40%		**1.1%**	63.3	2.533		**2.4%**	20.0	0.800
Video Streaming													
Streaming Success Ratio	80%	100%	50.00%	15%	60%		**89%**	45.0	2.025		**85%**	25.0	1.125
Video Quality MOS	3.5	5	15.00%	15%	60%		**3.7**	13.3	0.180		**3.6**	6.7	0.090
Video MOS < 3,8	10%	0%	10.00%	15%	60%		**9.2%**	8.0	0.072		**7.4%**	26.0	0.234
Video Access Time [s]	5	0	15.00%	15%	60%		**2.2**	56.0	0.756		**2.3**	54.0	0.729
Video Access Time > 5 s	10%	0%	10.00%	15%	60%		**2.8%**	72.0	0.648		**4.3%**	57.0	0.513
Data Testing													
Transfer Success Ratio DL (e.g. 5MB)	80%	100%	10.00%	30%	60%		**96%**	80.0	1.440		**89%**	45.0	0.810
Average throughput DL [Mbit/s]	1	1000	14.00%	30%	60%		**250**	24.9	0.628		**190**	18.9	0.477
10th percentile of (low) throughput DL [Mbit/s]	1	1000	18.00%	30%	60%		**56**	5.5	0.178		**47**	4.6	0.149
90th percentile of (high) throughput DL [Mbit/s]	1	1000	8.00%	30%	60%		**400**	39.9	0.575		**420**	41.9	0.604
Transfer Success Ratio UL (e.g. 2MB)	80%	100%	10.00%	30%	60%		**92%**	60.0	1.080		**97%**	85.0	1.530
Average throughput UL [Mbit/s]	1	1000	14.00%	30%	60%		**200**	19.9	0.502		**180**	17.9	0.452
10th percentile of (low) throughput UL [Mbit/s]	1	1000	18.00%	30%	60%		**40**	3.9	0.126		**45**	4.4	0.143
90th percentile of (high) throughput UL [Mbit/s]	1	1000	8.00%	30%	60%		**380**	37.9	0.546		**390**	38.9	0.561
Latency and Interactivity													
Interactivity Success Ratio (Score > 25)	80%	100%	50.00%	15%	60%		**89%**	45.0	2.025		**88%**	40.0	1.800
Average Interactivity Score	25	100	50.00%	15%	60%		**92**	89.3	4.020		**74**	65.3	2.940
Browsing													
Activity Success Ratio	80%	100%	50.00%	25%	60%		**99%**	95.0	7.125		**98%**	90.0	6.750
Average Duration [s]	3	0	50.00%	25%	60%		**3.6**	0.0	0.000		**3.7**	0.0	0.000
Social Media and Messaging													
Activity Success Ratio (upload duration < 15 s)	80%	100%	50.00%	15%	60%		**89%**	45.0	2.025		**88%**	40.0	1.800
Average Duration [s]	5	0	30.00%	15%	60%		**3.4**	32.0	0.864		**4.8**	4.0	0.108
Activity Duration > 5 s	10%	0%	20.00%	15%	60%		**0%**	100.0	1.800		**3%**	70.0	1.260
								Sum (City)	*56.44*			Sum (Rural)	*42.75*
								Weight	60%			Weight	40%
								Overall Score (60% City + 40% Rural)					**50.97**

Annex B: Example set of weighting factors, limits and thresholds

B.1 General

This annex provides a second example which represents a best practise at the time of release of the present document. The information here is intended to provide an illustration of how to practically apply network benchmarking and scoring as described in the body of the present document. In this regard it identifies example weights, limits and thresholds that could be applied to areas and mobile services as well as providing example worked network scoring calculations.

B.2 Area

B.2.1 Geographical divisions

The areas could be weighted and subdivided in the following manner.

EXAMPLE 1:

Area Type	Cumulative Area Type Weight	Area	Subdivision	Weight
Cities	80 %	Cities	Big and Medium	45 %
		Cities	Towns	20 %
		Complementary Areas	Hotspots	15 %
Outside Cities	20 %	Roads	n/a	12,5 %
		Complementary Areas	Railways	7,5 %

The Area Type is not a dimension introduced in the main part of the present document and serves merely to illustrate how the introduced Areas and their subdivisions are logically grouped.

This allows for alternatives which keep the high-level distribution between Area Types if, depending on the scope of the exercise, other combinations are possible where complementary areas are not in scope.

Two further examples are shown below.

EXAMPLE 2:

Area Type	Cumulative Area Type Weight	Area	Subdivision	Weight
Cities	80 %	Cities	Big and Medium	45 %
		Cities	Towns	20 %
		Complementary Areas	Hotspots	15 %
Outside Cities	20 %	Roads	n/a	20 %

EXAMPLE 3:

Area Type	Cumulative Area Type Weight	Area	Subdivision	Weight
Cities	80 %	Cities	Big and Medium	60 %
		Cities	Towns	20 %
Outside Cities	20 %	Roads	n/a	20 %

B.3 Mobile services

Service Type	Weight
Telephony	36 %
Data Services	64 %

B.4 Test metrics of mobile services

B.4.1 Telephony

Factor	Lower limit		Upper limit	Weight
	Cities	Outside Cities		
Composite Call Success Criterion combining: • Call Setup Success Ratio • Call Setup Time < 15 s • Inverse of Call Drop Ratio • Ratio of calls with no 2 consecutive speech samples < 1,3 MOS	90 %	85 %	100 %	55 %
10th percentile of MOS across all Samples	2,3 MOS		4,5 MOS	22,5 %
Percentage of calls supporting Data Connectivity	95 %	90 %	100	4,5%
90th percentile of Call Setup Time [s]	1,5 s		6 s	18 %

B.4.2 Data Services

B.4.2.1 General

Data Service Type	Weight	Subdivision	Subdivision Weight
Video Streaming	22,5 %	VoD	11,25 %
		Livestream	11,25 %
Data Testing	45 %	Fixed Size File DL	10 %
		Fixed Size File UL	10 %
		Fixed Duration File DL	12,5 %
		Fixed Duration File UL	12,5 %
Browsing	22,5 %	Web Pages	22,5 %
OTT Conversational App, see Recommendation ITU-T P.565.1 [i.17]	5 %		
Interactivity "eGaming Pattern", see Recommendation ITU-T G.1051 [i.14] and clause 8.4.3.5	5 %		

B.4.2.2 Video Streaming

Factor	Lower limit		Upper limit		
	Cities	Outside cities	Cities	Outside cities	Weight
Composite Session Success Criterion combining: • Success to access and stream the video (Video Streaming Service Success Ratio, VSSSR) • Absence of freezes above 1 s duration	90,0 %	85,0 %	100,0 %		55,0 %
Average Video Resolution [p]	960 p		1 080 p		27,0 %
Video access time [s] (for VoD)	1,2 s	1,5 s	5,5 s		18,0 %
Video access time [s] (for livestream)	1,6 s	1,9 a			

B.4.2.3 Data Testing

B.4.2.3.1 File Download (based on 10 MB File Size)

Factor	Lower limit		Upper limit		
	Cities	Outside cities	Cities	Outside cities	Weight
Composite Session Success Criterion combining: • Success to access and download the file • Achievement of min. throughput of 1 000 kbit/s	90,0 %	85,0 %	100,0 %		55,0 %
Average Download Session Duration [s]	1,0 s	1,5 s	6,0 s		13,5 %
10th percentile of Download throughput [kbit/s]	1 000 kbit/s		40 000 kbit/s	25 000 kbit/s	22,5 %
90th percentile of Download throughput [kbit/s]	50 000 kbit/s	20 000 kbit/s	300 000 kbit/s	200 000 kbit/s	9,0 %

B.4.2.3.2 File Upload (based on 5 MB File Size)

Factor	Lower limit		Upper limit		
	Cities	Outside cities	Cities	Outside cities	Weight
Composite Session Success Criterion combining: • Success to access and download the file • Achievement of min. throughput of 500 kbit/s	90,0 %	85,0 %	100,0 %		55,0 %
Average Upload Session Duration [s]	1,0 s	1,5 s	6,0s	8,0s	13,5 %
10th percentile of Upload throughput [kbit/s]	500 kbit/s		20 000 kbit/s	15 000 kbit/s	22,5 %
90th percentile of Upload throughput [kbit/s]	20 000,0 kbit/s	10 000 kbit/s	75 000 kbit/s	50 000 kbit/s	9,0 %

B.4.2.3.3 File Download (based on 7 s Fixed Download Time)

Factor	Lower limit		Upper limit		
	Cities	Outside cities	Cities	Outside cities	Weight
Composite Session Success Criterion combining: • Success to access and download the file for the full duration • Achievement of min. throughput of 1 000 kbit/s	90,0 %	85,0 %	100,0 %		55,0 %
Share of samples faster than 20 Mbps [%]	60 %	40 %	100 %		22,5 %
Share of samples faster than 100 Mbs [%]	25 %	0 %	80 %	60 %	13,5 %
90th percentile of Download throughput [kbit/s]	50 000 kbit/s	30 000 kbit/s	600 000 kbit/s	250 000 kbit/s	9,0 %

B.4.2.3.4 File Upload (based on 7 s Fixed Upload Time)

Factor	Lower limit		Upper limit		
	Cities	Outside cities	Cities	Outside cities	Weight
Composite Session Success Criterion combining: • Success to access and upload the file • Achievement of min. throughput of 500 kbit/s	90,0 %	85,0 %	100,0 %		55,0 %
Share of samples faster than 2 Mbps [%]	90 %	80 %	100 %		22,5 %
Share of samples faster than 5 Mbs [%]	70 %	60 %			13,5 %
90th percentile of Upload throughput [kbit/s]	20 000 kbit/s	15 000 kbit/s	120 000 kbit/s	100 000 kbit/s	9,0 %

B.4.2.4 Browsing

B.4.2.4.1 Void

B.4.2.4.2 Web Pages

Factor	Lower limit		Upper limit		
	Cities	Outside cities	Cities	Outside cities	Weight
Dynamic Web Pages Composite Session Success: • Success to access and download min 1 MB of web page content see ETSI TR 103 733 [i.12] • Achievement of min. throughput of 1 000 kbit/s on first 1 MB of content	90,0 %	85,0 %	100,0 %		55,0 %
Average Time to Download first 1 MB of content	1,5 s		4 s		45,0 %

B.4.2.4.3 OTT Conversational App

Factor	Lower limit		Upper limit	Weight
	Cities	Outside Cities		
Composite Call Success Criterion combining: • Call Setup Success Ratio • Call Setup Time < 15 s • Inverse of Call Drop Ratio • Ratio of calls with no 3 consecutive speech samples < 1,3 MOS	90 %	85 %	100 %	55 %
10th percentile of MOS across all Samples	2,0 MOS		4,0 MOS	45 %

B.4.2.4.4 Interactivity "eGaming Pattern"

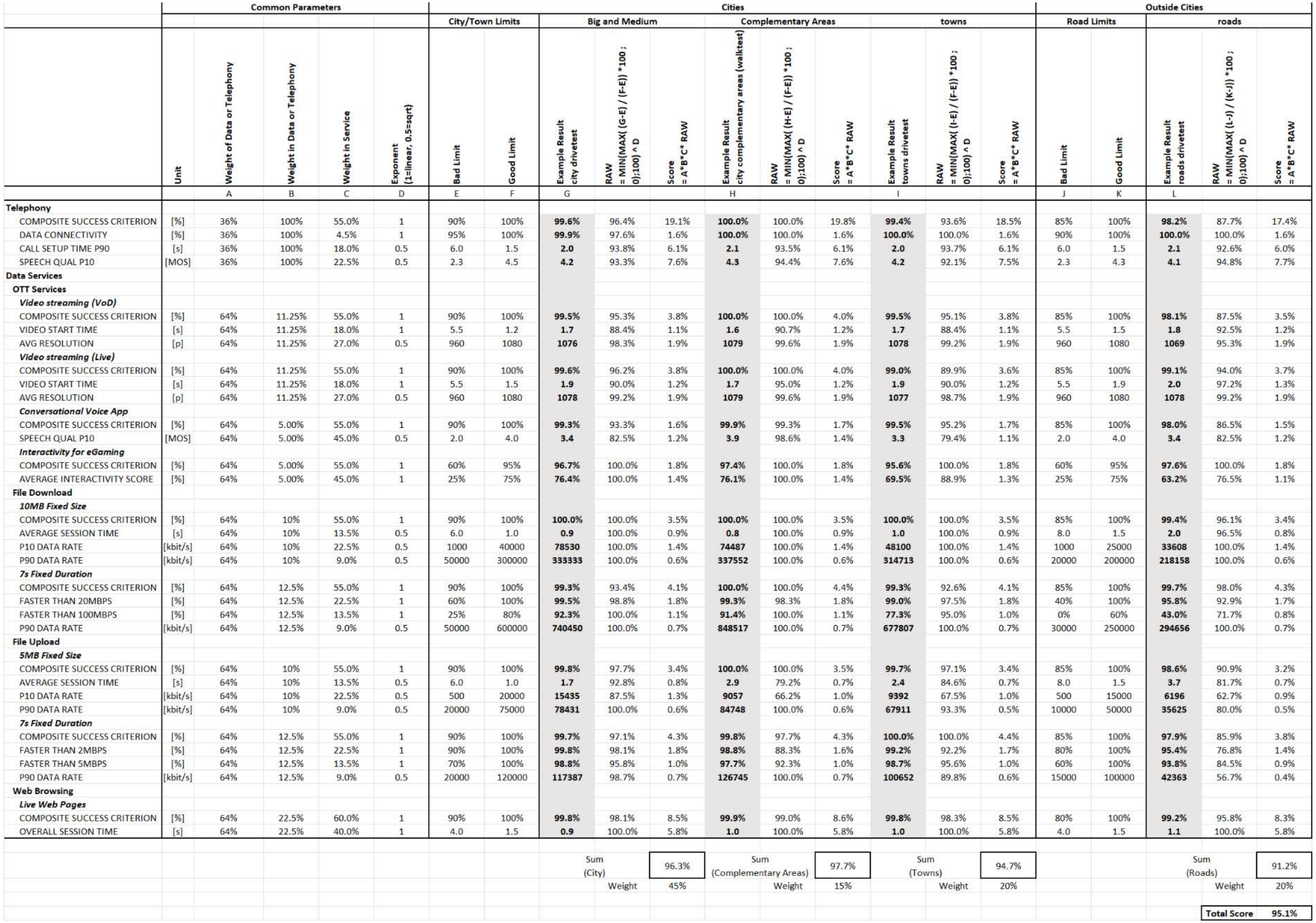
Factor	Lower limit	Upper limit	Weight
Composite Success Criterion: • Percentage of tests with Interactivity Score > 0	60 %	95 %	55 %
Avg Interactivity Score of all samples	25 %	75 %	45 %

B.5 Remarks on mapping functions

The example presented in the above clauses assumes that different types of mapping functions are used for the different factors/QoS parameters.

Linear functions are used for all QoS parameters with the following exceptions, where square root-shaped mapping functions are used: Speech Quality, Data Throughput, Avg. Video Resolution and File DL and UL Average Session Duration.

B.6 Example Calculation



	Common Parameters					Cities											Outside Cities				
						City/Town Limits		Big and Medium			Complementary Areas			towns			Road Limits		roads		
	Unit	Weight of Data or Telephony	Weight in Data or Telephony	Weight in Service	Exponent (1=linear, 0.5=sqrt)	Bad Limit	Good Limit	Example Result city drivetest	RAW = MIN(MAX((G-E) / (F-E)) *100 ; 0);100) ^ D	Score = A*B*C* RAW	Example Result city complementary areas (walktest)	RAW = MIN(MAX((H-E) / (F-E)) *100 ; 0);100) ^ D	Score = A*B*C* RAW	Example Result towns drivetest	RAW = MIN(MAX((I-E) / (F-E)) *100 ; 0);100) ^ D	Score = A*B*C* RAW	Bad Limit	Good Limit	Example Result roads drivetest	RAW = MIN(MAX((L-J) / (K-J)) *100 ; 0);100) ^ D	Score = A*B*C* RAW
		A	B	C	D	E	F	G			H			I			J	K	L		
Telephony																					
COMPOSITE SUCCESS CRITERION	[%]	36%	100%	55.0%	1	90%	100%	**99.6%**	96.4%	19.1%	**100.0%**	100.0%	19.8%	**99.4%**	93.6%	18.5%	85%	100%	**98.2%**	87.7%	17.4%
DATA CONNECTIVITY	[%]	36%	100%	4.5%	1	95%	100%	**99.9%**	97.6%	1.6%	**100.0%**	100.0%	1.6%	**100.0%**	100.0%	1.6%	90%	100%	**100.0%**	100.0%	1.6%
CALL SETUP TIME P90	[s]	36%	100%	18.0%	0.5	6.0	1.5	**2.0**	93.8%	6.1%	**2.1**	93.5%	6.1%	**2.0**	93.7%	6.1%	6.0	1.5	**2.1**	92.6%	6.0%
SPEECH QUAL P10	[MOS]	36%	100%	22.5%	0.5	2.3	4.5	**4.2**	93.3%	7.6%	**4.3**	94.4%	7.6%	**4.2**	92.1%	7.5%	2.3	4.3	**4.1**	94.8%	7.7%
Data Services																					
OTT Services																					
Video streaming (VoD)																					
COMPOSITE SUCCESS CRITERION	[%]	64%	11.25%	55.0%	1	90%	100%	**99.5%**	95.3%	3.8%	**100.0%**	100.0%	4.0%	**99.5%**	95.1%	3.8%	85%	100%	**98.1%**	87.5%	3.5%
VIDEO START TIME	[s]	64%	11.25%	18.0%	1	5.5	1.2	**1.7**	88.4%	1.1%	**1.6**	90.7%	1.2%	**1.7**	88.4%	1.1%	5.5	1.5	**1.8**	92.5%	1.2%
AVG RESOLUTION	[p]	64%	11.25%	27.0%	0.5	960	1080	**1076**	98.3%	1.9%	**1079**	99.6%	1.9%	**1078**	99.2%	1.9%	960	1080	**1069**	95.3%	1.9%
Video streaming (Live)																					
COMPOSITE SUCCESS CRITERION	[%]	64%	11.25%	55.0%	1	90%	100%	**99.6%**	96.2%	3.8%	**100.0%**	100.0%	4.0%	**99.0%**	89.9%	3.6%	85%	100%	**99.1%**	94.0%	3.7%
VIDEO START TIME	[s]	64%	11.25%	18.0%	1	5.5	1.5	**1.9**	90.0%	1.2%	**1.7**	95.0%	1.2%	**1.9**	90.0%	1.2%	5.5	1.9	**2.0**	97.2%	1.3%
AVG RESOLUTION	[p]	64%	11.25%	27.0%	0.5	960	1080	**1078**	99.2%	1.9%	**1079**	99.6%	1.9%	**1077**	98.7%	1.9%	960	1080	**1078**	99.2%	1.9%
Conversational Voice App																					
COMPOSITE SUCCESS CRITERION	[%]	64%	5.00%	55.0%	1	90%	100%	**99.3%**	93.3%	1.6%	**99.9%**	99.3%	1.7%	**99.5%**	95.2%	1.7%	85%	100%	**98.0%**	86.5%	1.5%
SPEECH QUAL P10	[MOS]	64%	5.00%	45.0%	0.5	2.0	4.0	**3.4**	82.5%	1.2%	**3.9**	98.6%	1.4%	**3.3**	79.4%	1.1%	2.0	4.0	**3.4**	82.5%	1.2%
Interactivity for eGaming																					
COMPOSITE SUCCESS CRITERION	[%]	64%	5.00%	55.0%	1	60%	95%	**96.7%**	100.0%	1.8%	**97.4%**	100.0%	1.8%	**95.6%**	100.0%	1.8%	60%	95%	**97.6%**	100.0%	1.8%
AVERAGE INTERACTIVITY SCORE	[%]	64%	5.00%	45.0%	1	25%	75%	**76.4%**	100.0%	1.4%	**76.1%**	100.0%	1.4%	**69.5%**	88.9%	1.3%	25%	75%	**63.2%**	76.5%	1.1%
File Download																					
10MB Fixed Size																					
COMPOSITE SUCCESS CRITERION	[%]	64%	10%	55.0%	1	90%	100%	**100.0%**	100.0%	3.5%	**100.0%**	100.0%	3.5%	**100.0%**	100.0%	3.5%	85%	100%	**99.4%**	96.1%	3.4%
AVERAGE SESSION TIME	[s]	64%	10%	13.5%	0.5	6.0	1.0	**0.9**	100.0%	0.9%	**0.8**	100.0%	0.9%	**1.0**	100.0%	0.9%	8.0	1.5	**2.0**	96.5%	0.8%
P10 DATA RATE	[kbit/s]	64%	10%	22.5%	0.5	1000	40000	**78530**	100.0%	1.4%	**74487**	100.0%	1.4%	**48100**	100.0%	1.4%	1000	25000	**33608**	100.0%	1.4%
P90 DATA RATE	[kbit/s]	64%	10%	9.0%	0.5	50000	300000	**333333**	100.0%	0.6%	**337552**	100.0%	0.6%	**314713**	100.0%	0.6%	20000	200000	**218158**	100.0%	0.6%
7s Fixed Duration																					
COMPOSITE SUCCESS CRITERION	[%]	64%	12.5%	55.0%	1	90%	100%	**99.3%**	93.4%	4.1%	**100.0%**	100.0%	4.4%	**99.3%**	92.6%	4.1%	85%	100%	**99.7%**	98.0%	4.3%
FASTER THAN 20MBPS	[%]	64%	12.5%	22.5%	1	60%	100%	**99.5%**	98.8%	1.8%	**99.3%**	98.3%	1.8%	**99.0%**	97.5%	1.8%	40%	100%	**95.8%**	92.9%	1.7%
FASTER THAN 100MBPS	[%]	64%	12.5%	13.5%	1	25%	80%	**92.3%**	100.0%	1.1%	**91.4%**	100.0%	1.1%	**77.3%**	95.0%	1.0%	0%	60%	**43.0%**	71.7%	0.8%
P90 DATA RATE	[kbit/s]	64%	12.5%	9.0%	0.5	50000	600000	**740450**	100.0%	0.7%	**848517**	100.0%	0.7%	**677807**	100.0%	0.7%	30000	250000	**294656**	100.0%	0.7%
File Upload																					
5MB Fixed Size																					
COMPOSITE SUCCESS CRITERION	[%]	64%	10%	55.0%	1	90%	100%	**99.8%**	97.7%	3.4%	**100.0%**	100.0%	3.5%	**99.7%**	97.1%	3.4%	85%	100%	**98.6%**	90.9%	3.2%
AVERAGE SESSION TIME	[s]	64%	10%	13.5%	0.5	6.0	1.0	**1.7**	92.8%	0.8%	**2.9**	79.2%	0.7%	**2.4**	84.6%	0.7%	8.0	1.5	**3.7**	81.7%	0.7%
P10 DATA RATE	[kbit/s]	64%	10%	22.5%	0.5	500	20000	**15435**	87.5%	1.3%	**9057**	66.2%	1.0%	**9392**	67.5%	1.0%	500	15000	**6196**	62.7%	0.9%
P90 DATA RATE	[kbit/s]	64%	10%	9.0%	0.5	20000	75000	**78431**	100.0%	0.6%	**84748**	100.0%	0.6%	**67911**	93.3%	0.5%	10000	50000	**35625**	80.0%	0.5%
7s Fixed Duration																					
COMPOSITE SUCCESS CRITERION	[%]	64%	12.5%	55.0%	1	90%	100%	**99.7%**	97.1%	4.3%	**99.8%**	97.7%	4.3%	**100.0%**	100.0%	4.4%	85%	100%	**97.9%**	85.9%	3.8%
FASTER THAN 2MBPS	[%]	64%	12.5%	22.5%	1	90%	100%	**99.8%**	98.1%	1.8%	**98.8%**	88.3%	1.6%	**99.2%**	92.2%	1.7%	80%	100%	**95.4%**	76.8%	1.4%
FASTER THAN 5MBPS	[%]	64%	12.5%	13.5%	1	70%	100%	**98.8%**	95.8%	1.0%	**97.7%**	92.3%	1.0%	**98.7%**	95.6%	1.0%	60%	100%	**93.8%**	84.5%	0.9%
P90 DATA RATE	[kbit/s]	64%	12.5%	9.0%	0.5	20000	120000	**117387**	98.7%	0.7%	**126745**	100.0%	0.7%	**100652**	89.8%	0.6%	15000	100000	**42363**	56.7%	0.4%
Web Browsing																					
Live Web Pages																					
COMPOSITE SUCCESS CRITERION	[%]	64%	22.5%	60.0%	1	90%	100%	**99.8%**	98.1%	8.5%	**99.9%**	99.0%	8.6%	**99.8%**	98.3%	8.5%	80%	100%	**99.2%**	95.8%	8.3%
OVERALL SESSION TIME	[s]	64%	22.5%	40.0%	1	4.0	1.5	**0.9**	100.0%	5.8%	**1.0**	100.0%	5.8%	**1.0**	100.0%	5.8%	4.0	1.5	**1.1**	100.0%	5.8%
								Sum (City)		96.3%	Sum (Complementary Areas)		97.7%	Sum (Towns)		94.7%			Sum (Roads)		91.2%
									Weight	45%		Weight	15%		Weight	20%				Weight	20%
																				Total Score	**95.1%**

History

Document history		
V1.1.1	August 2019	Publication
V1.2.1	October 2023	Publication