

Recommendation

ITU-T G.1052 (05/2024)

SERIES G: Transmission systems and media, digital systems and networks

Multimedia Quality of Service and performance – Generic and user-related aspects

Testbed framework for mobile application quality of service and quality of experience evaluation

ITU-T G-SERIES RECOMMENDATIONS
Transmission systems and media, digital systems and networks

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100-G.199
GENERAL CHARACTERISTICS COMMON TO ALL ANALOGUE CARRIER-TRANSMISSION SYSTEMS	G.200-G.299
INDIVIDUAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON METALLIC LINES	G.300-G.399
GENERAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON RADIO-RELAY OR SATELLITE LINKS AND INTERCONNECTION WITH METALLIC LINES	G.400-G.449
COORDINATION OF RADIOTELEPHONY AND LINE TELEPHONY	G.450-G.499
TRANSMISSION MEDIA AND OPTICAL SYSTEMS CHARACTERISTICS	G.600-G.699
DIGITAL TERMINAL EQUIPMENTS	G.700-G.799
DIGITAL NETWORKS	G.800-G.899
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900-G.999
MULTIMEDIA QUALITY OF SERVICE AND PERFORMANCE – GENERIC AND USER-RELATED ASPECTS	G.1000-G.1999
TRANSMISSION MEDIA CHARACTERISTICS	G.6000-G.6999
DATA OVER TRANSPORT – GENERIC ASPECTS	G.7000-G.7999
PACKET OVER TRANSPORT ASPECTS	G.8000-G.8999
ACCESS NETWORKS	G.9000-G.9999

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T G.1052

Testbed framework for mobile application quality of service and quality of experience evaluation

Summary

Recommendation ITU-T G.1052 proposes a testbed set-up and a methodology to evaluate the quality of experience of multiple 5G and beyond mobile use cases. A testbed, its calibration and a testing methodology are described, able to provide consistent and repeatable results in an automated fashion. A key focus of the set-up is the use of commercial devices and applications, aiming at making results more valuable by mimicking the set-up of a real user.

Recommendation ITU-T G.1052 focuses on cellular Third Generation Partnership Project networks. A similar set-up could be thought of using other technologies, e.g., wireless fidelity, also interoperating to emulate inter-radio access technology handovers, although this lies outside the scope of Recommendation ITU-T G.1052.

History*

Edition	Recommendation	Approval	Study Group	Unique ID
1.0	ITU-T G.1052	2024-05-29	12	11.1002/1000/15898

Keywords

5G and beyond, methodology, MOS, QoE, testbed.

* To access the Recommendation, type the URL <https://handle.itu.int/> in the address field of your web browser, followed by the Recommendation's unique ID.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the appropriate ITU-T databases available via the ITU-T website at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2024

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Table of Contents

		Page
1	Scope	1
2	References.....	1
3	Definitions	2
	3.1 Terms defined elsewhere	2
	3.2 Terms defined in this Recommendation.....	3
4	Abbreviations and acronyms	3
5	Conventions	6
6	Scoring methodology.....	6
	6.1 Measurements.....	6
	6.2 Key performance indicators	6
	6.3 Session quality score	7
	6.4 Aggregated scores	8
7	Use cases and network scenarios.....	8
	7.1 Use cases	9
	7.2 Definition of network scenarios	11
	7.3 Network scenario parameters	14
8	Testing methodology	15
	8.1 Test system configuration.....	15
	8.2 Test case initial conditions	16
	8.3 Test case post-conditions.....	17
	8.4 Application test case.....	17
9	Testbed calibration.....	17
	9.1 Calibration of individual measurement instruments.....	18
	9.2 Characterization of the testbed	18
	9.3 Radio frequency connection compensation.....	19
	9.4 Measurement probes.....	19
	9.5 Baseline calibration standard reference	20
	9.6 Testbed health tracking.....	20
10	Testbed description.....	20
	10.1 Devices, operating systems, and applications	21
	10.2 Communication system	21
	10.3 Service provider.....	31
	10.4 Measurements.....	31
	10.5 Test coordination and automation	32
	Appendix I – Relevant ITU-T Recommendations	34

	Page
Appendix II – Examples of key performance indicator normalization mapping functions.....	35
II.1 Linear mapping.....	36
II.2 Logarithmic mapping	37
Appendix III – Scenario parameter settings.....	39
III.1 Parameters explained.....	39
III.2 Scenarios.....	42
Appendix IV – Opentap overview	50
IV.1 Test plans.....	50
IV.2 Test steps	51
IV.3 Verdicts.....	51
IV.4 Resources.....	53
IV.5 Result listeners.....	53
IV.6 Session logs	54
IV.7 Packages and plugins.....	54
Bibliography.....	56

Introduction

5G is expected to bring faster transmission, massive connectivity, lower latency and higher reliability, enabling better and new services and applications. Understanding end user expectations and experience is of paramount importance to service providers and application developers, as it helps to improve the service quality perception in a highly competitive service provision market.

Traditionally, quality of service (QoS) is the most widely used metric, which can be used to assess the performance of telecommunication services by performing a series of technical measurements such as throughput, delay, jitter, error probability and packet loss. Although QoS has been used to indicate the effectiveness of a service, it does not take subjective factors into account, which directly reflect the user-perceived service. Quality of experience (QoE), defined by the ITU-T as "the degree of delight or annoyance of the user of an application or service" (clause 6.209 of [ITU-T P.10]), whose influencing factors

"include the type and characteristics of the application or service, context of use, the user's expectations with respect to the application or service and their fulfilment, the user's cultural background, socio-economic issues, psychological profiles, emotional state of the user, and other factors whose number will likely expand with further research" (clause 6.210 of [ITU-T P.10]), making it highly subjective.

QoE assessment in an end-to-end (E2E) communication system is a complex task, as it depends on many factors, such as the network conditions, including both the radio access network and core network, the terminal capabilities, the service providers, and human factors. Furthermore, the combination of various use cases and network configurations further complicates the situation and makes the theoretical evaluation of QoE for a given 5G service impossible.

Typically, QoE is evaluated via subjective methods or alternatively, if available, using objective method such as with parametric models. In a subjective testing approach, a group of end users rates the perceived quality by giving the experimenter an opinion score, later aggregating into a mean opinion score [b-Kozamernik] called human panels. Being a laborious and time-consuming method, this subjective rating cannot be conducted in real time. Furthermore, since people participating in the experiment are typically paid for their contribution, and each additional sequence needs to be re-evaluated, the approach is hardly scalable. On the other hand, objective methods applied in field testing utilize QoE models with many measured parameters to predict user-perceived QoE. The lack of control over the testing environment, e.g., changing radio propagation and network conditions, makes results unpredictable. Therefore, it is of great interest to design a fully laboratory-controlled end-to-end testbed with a homogeneous testing framework that allows QoE evaluations for new mobile applications and services in an automated and repeatable manner, which can then be used to calibrate such tools.

Since testbed approaches can be criticized for being artificial and far from reality, the aim of this Recommendation is to keep experiments as close to reality as possible. This can be done by using the following:

- commercial devices: a selection of the most common consumer devices should be used to test the desired use case;
- real applications: applications have become so complex that it is hard to emulate their traffic artificially using the same adaptation algorithms.

Calibration is of the utmost importance for a laboratory-based testbed, and it comes in different forms. The calibration related to the measurement block includes the calibration of an individual measurement instrument, the characterization of testbed performance, e.g., latency and throughput, compensation of radio frequency connection losses and the impacts of various measurement agents. Furthermore, calibration is needed to convert measured key performance indicators into session quality score values, accurately predicting human panel results.

Finally, automation is another key feature of a laboratory-based testbed, allowing the creation of repeatable experiments and thus reliably comparing different settings, e.g., different devices, radio or network settings. With an automated system, measurements can be taken with great confidence, always following the recommended guidelines, including periodic calibrations and compensations needed to obtain stable results over time (e.g., over multiple measurement campaigns) and over space (e.g., over different testbeds).

Recommendation ITU-T G.1052

Testbed framework for mobile application quality of service and quality of experience evaluation

1 Scope

This Recommendation specifies an automated laboratory-based (testbed) scoring framework for various types of human-machine or machine-machine applications over a series of varying use cases, network conditions, terminals and measurement domains. All parts of the testbed are described in this Recommendation, including its components, the calibration process, the testing methodology, the most relevant use cases and network scenarios, and the scoring methodology.

The framework consists of a single laboratory-controlled cellular network testbed and a set of test cases to automatically evaluate application (app) performance under various network conditions or user equipment (UE) and provide a single quality score based on the measured key performance indicators (KPIs). The score is computed by weighting the results obtained while testing the different uses cases for each app, for the different aspects relevant to the user, and under the network scenarios pertinent to the app. All KPI measurements are available to the user. The framework incorporates specific quality of service (QoS) to quality of experience (QoE) estimations based primarily on standardized models. In the absence of standardized models, subjective experiments may be used to determine fixed coefficients for the calibration of the testbed.

This Recommendations describes components and methodology to obtain stable and reproducible results, so that different parties implementing such a testbed would obtain comparable results. This can be useful to device vendors, mobile network operators, app developers, test and measurements companies, among others.

Existing and in-progress Recommendations (sufficiently mature) are referenced for QoE estimation purposes, with accommodation for the inputs needed provided. The proposed use cases are for further study.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

- [ITU-T F.743.17] Recommendation ITU-T F.743.17 (2022), *Requirements for cloud gaming systems*.
- [ITU-T K.127] Recommendation ITU-T K.127 (2017), *Immunity requirements for telecommunication equipment in close proximity use of wireless devices*.
- [ITU-T P.10] Recommendation ITU-T P.10/G.100 (2017), *Vocabulary for performance and quality of service*.
- [ITU-T Q.1743] Recommendation ITU-T Q.1743 (2016), *IMT-advanced references to release 11 of LTE-advanced evolved packet core network*.
- [ITU-T Q.4068] Recommendation ITU-T Q.4068 (2021), *Open application program interfaces (APIs) for interoperable testbed federations*.



- [ITU-T Y.1540] Recommendation ITU-T Y.1540 (2019), *Internet protocol data communication service – IP packet transfer and availability performance parameters*.
- [ITU-T Y.4500.15] Recommendation ITU-T Y.4500.15/Q.3955 (2018), *oneM2M – Testing framework*.
- [IEC 62264-3] International Standard IEC 62264-3:2016, *Enterprise-control system integration – Part 3: Activity models of manufacturing operations management*.
- [ISO 10605] International Standard ISO 10605:2023, *Road vehicles – Test methods for electrical disturbances from electrostatic discharge*.
- [3GPP TS 36.331] Technical Specification 3GPP TS 36.331 V18.2.0 (2024), *Evolved universal terrestrial radio access (E-UTRA); Radio resource control (RRC); Protocol specification (Release 18)*.
- [3GPP TS 36.508] Technical Specification 3GPP TS 36.508 V18.5.0 (2024), *Evolved universal terrestrial radio access (E-UTRA) and evolved packet core (EPC); Common test environments for user equipment (UE) conformance testing (Release 18)*.
- [3GPP TS 38.508-1] 3GPP TS 38.508-1 V18.3.0 (2023), *5GS; User equipment (UE) conformance specification; Part 1: Common test environment (Release 18)*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

- 3.1.1 **cell** [ITU-T Q.1743]
- 3.1.2 **cloud gaming** [ITU-T F.743.17]
- 3.1.3 **device under test (DUT)** [ISO 10605]
- 3.1.4 **end-to-end quality** [ITU-T P.10]
- 3.1.5 **heterogeneous network** [ITU-T Q.1743]
- 3.1.6 **key performance indicator (KPI)** [IEC 62264-3]
- 3.1.7 **macro cells** [ITU-T Q.1743]
- 3.1.8 **micro cells** [ITU-T Q.1743]
- 3.1.9 **pico cells** [ITU-T Q.1743]
- 3.1.10 **primary cell** [3GPP TS 36.331]
- 3.1.11 **quality of experience (QoE)** [ITU-T P.10]
- 3.1.12 **QoE influencing factors** [ITU-T P.10]
- 3.1.13 **QoE assessment** [ITU-T P.10]
- 3.1.14 **quality of service (QoS)** [ITU-T P.10]
- 3.1.15 **radio frequency (RF)** [ITU-T K.127]
- 3.1.16 **testbed** [ITU-T Q.4068]
- 3.1.17 **test case** [ITU-T Y.4500.15]
- 3.1.18 **user equipment (UE)** [ITU-T Q.1743]

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 application under test (AUT): Application whose performance is under investigation.

3.2.2 application under test backend service: Remote endpoint of the application under test service layer, providing connectivity for the necessary automation and data gathering.

3.2.3 conducted radio frequency connection: Connection via capacitive and inductive coupling on to different lines and cabling, including power and data.

3.2.4 measurement probe: A software or hardware tool used to collect data on one or more aspects of the performance, behaviour and metrics of a system, for the purpose of analysis, optimization and debugging.

3.2.5 network scenario: Usage scenario determined by a parameterization of the air interface and the core network.

3.2.6 session quality score (SQS): A value representing the quality score for a session of the application under test.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ACK	Acknowledgement
ACR	Absolute Category Rating
AMF	Access and Mobility Management Function
API	Application Program Interface
app	application
AR	Augmented Reality
AUT	Application Under Test
AWGN	Additive White Gaussian Noise
BLER	Block Error Rate
BWP	Bandwidth Part
CG	Cloud Gaming
CLI	Command Line Interface
C-plane	Control plane
CPU	Central Processing Unit
CS	Configured Scheduling
DC	Dual Connectivity
DCI	Downlink Control Information
DL	Downlink
DPS	Dynamic Point Selection
DS	Delay Spread
DUT	Device Under Test
E2E	End-to-End



EN	Evolved universal mobile telecommunication system terrestrial radio access network New radio
eNB	Evolved Node B
EPA	Extended Pedestrian A
EPC	Evolved Packet Core
ETU	Extended Typical Urban
EARFCN	E-UTRA Absolute Radio-Frequency Channel Number
E-UTRA	Evolved Universal Terrestrial Radio Access
E-UTRAN	Evolved Universal mobile telecommunication system Terrestrial Radio Access Network
EVA	Extended Vehicular A
FDD	Frequency Division Duplex
FR	Frequency Range
gNB	Grouped Node B
GPS	Global Positioning System
GPU	Graphics Processing Unit
HARQ	Hybrid Automatic Repeat Request
HSI	High-Speed Internet
HSS	Home Subscriber Server
HST	High-Speed Train
IP	Internet Protocol
IPA	iPhone application Archive
KPI	Key Performance Indicator
LAN	Local Area Network
LSS	Live Streaming Service
LTE	Long-Term Evolution
MCS	Modulation and Coding Scheme
MIMO	Multiple-Input Multiple-Output
MME	Mobility Management Entity
MOS	Mean Opinion Score
NR	New Radio
NSA/SA	Non-Standalone/Standalone
NSSF	Network Slice Selection Function
OFDM	Orthogonal Frequency Division Modulation
OS	Operating System
OTA	Over The Air
PCF	Policy Control Function
PCRF	Policy and Charging Rules Function

PDCCH	Physical Downlink Control Channel
PDSCH	Physical Downlink Shared Channel
PGW	Packet data network Gateway
PRB	Physical Resource Block
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
QoE	Quality of Experience
QoS	Quality of Service
RAM	Random-Access Memory
RAN	Radio Access Network
RAT	Radio Access Technology
RB	Resource Block
RF	Radio Frequency
RMC	Reference Measurement Channel
RRC	Radio Resource Control
RSRP	Reference Signal Received Power
RTT	Round-Trip Time
Rx	Receiver
SCS	Subcarrier Spacing
SDU	Service Data Unit
SFN	Single Frequency Network
SGW	Serving Gateway
SMF	Session Management Function
SN	Social Networking
SNR	Signal-to-Noise Ratio
SPS	Semi-Persistent Scheduling
SQS	Session Quality Score
SS-RSRP	Synchronization Signal-Reference Signal Received Power
TCP	Transmission Control Protocol
TDD	Time Division Duplex
TDL	Tapped Delay Line
TM	Telemeeting
Tx	Transmitter
UDM	Unified Data Management
UE	User Equipment
UL	Uplink
U-plane	User Plane

UPF	User Plane Function
VR	Virtual Reality
Wi-Fi	Wireless Fidelity

5 Conventions

None.

6 Scoring methodology

This clause clearly describes a scoring methodology with the stages shown in Figure 6-1, their terminology and notation. Specification of numeric models for use cases lies outside the scope of this Recommendation. In this Recommendation, use case shall be interpreted as an app or service tested on a device (see clause 7.1). A list of existing model standards (i.e., existing ITU-T Recommendations) for specific use cases according to the scoring approach proposed here is contained in Appendix I. Note that the list mentioned therein may not include all relevant Recommendations.

Measurements are first explained in clause 6.1, describing what they are and how they can be taken. Next, clause 6.2 describes how measurements can be combined to compute high-level KPIs. Finally, KPIs are integrated into a session quality score (SQS), described in clause 6.3, and results can be aggregated with methods described in clause 6.4.

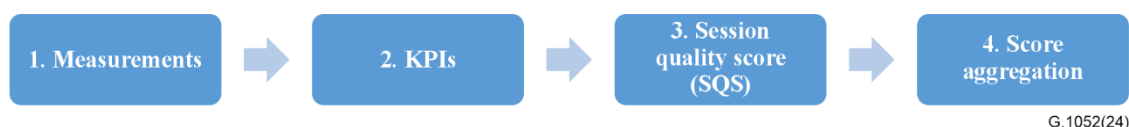


Figure 6-1 – Scoring methodology

6.1 Measurements

Measurements are the basic information needed to compute scores about any system. In this case, measurements can be obtained from either software or hardware probes, which help extract the desired quantity from the set-up (see clause 10.4).

Measurement collection is an integral and relevant part of automated test cases (see clause 10.5). These measurements are then combined to compute one or multiple KPIs.

6.2 Key performance indicators

The performance of a use case is measured by a set of KPIs, i.e., the quantifiable critical indicators of the performance of the test. Evaluation criteria can target devices under test (DUTs) or applications under test (AUTs), and various sets of KPIs can be considered for different objectives.

KPIs are derived from measurements. There can be a one-to-one mapping, e.g., power consumption, or, more typically, many-to-one mapping, e.g., time to load video stream is the difference between the video streaming play timestamp and the video streaming demand request timestamp. KPIs may be specified at different layers, e.g., in terms of network information such as packet loss, round-trip times (RTTs) or throughput, or at the application layer in terms of service specific parameters. KPIs can also be of different data types, e.g., integer value (number of user actions within a time window), real value (time to load a reference picture on a social media), boolean (message sent successfully) or vector values (content stall durations over a telemeeting session).

In the proposed methodology, a result from another test can become a KPI for a new test, intrinsically creating a hierarchy of KPIs. A practical example is the video streaming use case, where KPIs for the SQS are typically video quality, audio quality, audio-video synchronization and loading-related

information, such as stalling. For this particular use case, objective models in terms of SQSs already exist, as listed in Appendix I. Each one of these KPIs can also be studied individually, comprising themselves more basic KPIs, e.g., video resolution for video quality. This hierarchical structure is not considered in the proposed methodology, leaving to the experimenter the freedom to determine the best KPIs for its use case, and eventually determining suitable sub-problems to best model the hierarchy of KPIs – also depending on what input parameters or input information is required by existing objective models (see Appendix I).

Although KPIs can be described in many different forms, they should be precisely specified, avoiding ambiguity.

KPIs may be normalized if deemed useful, as explained in Appendix II.

6.3 Session quality score

The SQS is a metric designed to evaluate the quality of a test session. It is crucial to compare different set-ups, DUTs, AUTs, or experimental parameters. For instance, when testing to assess the influence of a network parameter on the performance of an AUT, the SQS should provide clear insights into the optimal values for the parameters being tested. This metric is derived by meaningfully aggregating all the identified KPIs. Measuring or ranking the performance of test sessions should be done on the E2E behaviour of the AUT, based on user preferences, thus meaning that SQS should strive to mimic the user QoE.

If an ITU-T Recommendation of an objective model (see Appendix I) exists for the target use case, the model intrinsically addresses this mapping. That model should be used if the AUT can provide the necessary input KPIs for the recommended objective model. If not, the experimenter should evaluate the possibility of using a different AUT within the same use case from which the necessary KPIs can be extracted.

The experimenter may use an algorithm not coming from an ITU-T Recommendation if no such objective model is available or cannot be applied. The algorithm might come from state-of-the-art scientific publications.

If an ITU-T Recommendation of a subjective test methodology exists for the target use case (see Appendix I), the experimenter should produce data to build an objective model with acceptable performance, taking inspiration from techniques explained, e.g., in [b-ITU-T P.1401] and [b-ITU-T P.1402] and other existing objective models. If this is the case, the experimenter is encouraged to submit this model to standards development organizations, such as ITU, for further discussion and validation by a group of experts and possibly inclusion in a Recommendation.

If no ITU-T Recommendation exists for subjective tests of the target use case, the experimenter is encouraged to propose a methodology to standards development organizations, such as ITU, for further discussion, validation by a group of experts, and possibly inclusion in a Recommendation.

Disclaimer: If no ITU-T Recommendation of an objective model or a subjective test methodology exists for the target use case, the experimenter may use a different model or methodology to evaluate its experiments. However, it shall be clearly stated that the model or methodology used does not come from an ITU-T Recommendation.

The SQS shall be defined on a rating scale. If the score is calibrated on a ground truth, the same rating scale should be used (e.g., a five-point absolute category rating (ACR) scale for data from most subjective tests). If calibration data comes from something other than subjective tests, the five-point ACR scale should not be used to avoid confusing the reader into thinking otherwise. Since specification of a scale lies outside the scope of this Recommendation, more details can be found in subjective quality assessment Recommendations, such as those listed in Appendix I.

6.4 Aggregated scores

Clause 6.3 describes a methodology to get an SQS. In general, though, multiple objective scores among multiple dimensions might be of interest, such as, but not limited to the following.

- Different realizations: repeating the same experiment with the same set-up multiple times allows the experimenter to get various samples of the same conditions. If the testbed, DUT and AUT are well controlled, the differences between realizations should be minor. In general, it is a good idea to obtain multiple realizations to both check the stability of the overall experiment set-up and determine whether relevant random processes are present within the set-up.
- Different DUTs: different smartphones might have different performances, especially if they have different price tags, have been launched in different years, and have different network capabilities (e.g., support to 5G or different modems), to mention a few. It is necessary to specify which DUTs have been used for an experiment, so that results can (or cannot) be compared.
- Different use cases: an experiment can target a single or multiple use cases.
- Different AUTs: use cases will likely be tested using a specific AUT. It is possible, in general, to use different AUTs for the same use case. For example, multiple video streaming platforms and multiple messaging and teleconferencing platforms exist.
- Different testbeds: different testbeds or testbed parameters might be of interest. For example, different traffic shaping algorithms, network slicings or networks from different operators.
- Different scenarios: network performance, and thus SQS, heavily depend on the network and radio interface conditions. Scenarios described in clause 7.2 represent a set of common conditions. Observation of how the scenario affects the experiment's performance could be of interest.

In general, it can be helpful to aggregate scores over one, multiple or all dimensions, making it easier to understand the performance obtained during the experiment.

Aggregating SQS can take several forms, depending on what interests the user and which dimension is aggregated. Likewise, aggregating functions can take many forms, the most typical being arithmetic average, weighted average, standard deviation, minimum and maximum, and median (or, in general, n th percentile).

For example, scores will likely be averaged over different realizations, although computing the standard deviation can help the user understand whether any configuration is unstable.

Furthermore, standards already exist for condensation of multiple dimensions into a single comprehensive score. For example, in [b-ETSI TR 103 559] weighted averages are used to consider experiments subdivided into several layers, i.e., what they call normalized KPIs, mobile services, and areas loosely corresponding to SQS, use cases and network scenarios described in this Recommendation.

Score aggregation is not further discussed here, since it lies outside the scope of this Recommendation; some standards already exist.

7 Use cases and network scenarios

This clause specifies scenarios and the primary use cases of a framework to test and benchmark the E2E QoE performance of applications, services and devices on various domains under realistic deployment conditions. Although QoE is perceived as a subjective metric, it is crucial to user satisfaction with a specific service, as defined in [ITU-T P.10]. The QoE evaluation is closely associated with the type of use case. For example, indicators such as time to log in, post or download content are relevant KPIs for the social network use case. In contrast, response time, video resolution

and framerate are appropriate KPIs for the virtual reality (VR) use case. Therefore, one of the first steps in determining a test and certification framework is the identification of appropriate use cases and relevant network scenarios. This approach was initially adopted by the Next Generation Mobile Network [b-NGMN] and later by Fifth Generation Infrastructure Public-Private Partnership [b-5G PPP], where the use cases and the network scenarios were tightly coupled. In this Recommendation, the use cases are decoupled from the network scenarios to broaden the scope of the investigation.

7.1 Use cases

In this Recommendation, use case shall be interpreted as an app or service tested on a device. The most relevant use cases in the context of 5G services are described in the remainder of this clause and reported in Table 7-1.

Table 7-1 – Use cases

Use case	Description
Augmented reality (AR)	For mobile phones, or AR glasses or headgear.
Cloud gaming (CG)	Used in mobile phones or connected consoles.
Content distribution streaming services	Typical streaming service in downlink (DL). This includes content on demand as well as live streaming.
High-speed Internet (HSI)	Traditional browsing or files up/download.
Live streaming services (LSSs)	Users can live stream their experiences, while other users can consume the live video both uplink (UL) and DL are put under pressure, and the service can use both reliable and non-reliable transport protocols.
Social networking (SN)	Content posting on online platforms.
Telemeetings (TMs)	Video conferencing applications typically allow the stream of both video and audio between two or more interacting users.
Virtual reality (VR)	For mobile phones or dedicated VR gear.

7.1.1 Augmented reality

From clause 6.3 of [b-ITU-T P.1320],

"Augmented reality overlays digital objects on the real-world environment with spatial registration that enables geometric persistence concerning placement and orientation within the real world. Augmented reality superimposes a layer of digital content over the real physical world such that users see virtual models at fixed positions and orientations mixed in with the actual world around them."

Thus, the user becomes part of this semi-virtual world and can manipulate objects or perform a series of actions there.

Multiple sub-systems in the device are involved, such as the global positioning system (GPS), the camera, the accelerometers, the speakers and the microphone, requiring a significant amount of processing either on-board or offloaded to a remote processing unit.

7.1.2 Cloud gaming

CG is a particularly challenging case where each game presents its own requirements. Moreover, games no longer purely rely on user-screen interaction but encompass multiple components such as GPS, camera or accelerometers. Therefore, the network demand is also quite different from one game to another. However, two main aspects are to be highlighted: latency and throughput. Latency in gaming is a crucial aspect, and it is typically assumed that a delay of about 50 ms is tolerable for non-VR games. An upper limit in terms of throughput is 4K resolution, which means an average of 50 Mbit/s.

7.1.3 Content distribution streaming services

With the increasing popularity of on-demand film viewing, streaming services are crucial apps to consider in the context of 5G services.

Customers use such services everywhere: at home, on public transportation, in cars, on trains, etc. These services are DL-centric, with extremely high data throughput requirements in the DL, while the UL is only used to control the stream.

Today's popular streaming services require 5 Mbit/s for 1 080 p videos and 20 Mbit/s for 4K resolution. As a result, video quality can be downgraded if the user network is not good enough, with 0.7 Mbit/s being the minimum the service requires for operation.

For this use case, several recommended objective models for network and service monitoring already exist, listed in Appendix I.

7.1.4 High-speed Internet

One of the most common applications for 5G will remain Internet broadband access. While current Internet speeds are acceptable for existing services and websites, there is a continuous trend in making websites fancier, containing many multimedia sources that increase their download size.

To maintain a good quality of experience for users and a low loading time, speed increases are required for Internet access. However, service availability with such rates can be challenging in specific scenarios.

The number of applications and services that use HSI is pervasive. However, it is commonly assumed that broadband everywhere means provision of a minimum of 20 Mbit/s everywhere. Therefore, this is the minimum throughput that services in this category are expected to cope with.

7.1.5 Live streaming services

Users keep increasingly demanding to broadcast and share their experiences through live video. LSSs also puts high demand on UL bandwidth for the person streaming, which is a key difference from content distribution services, where only DL resources are required.

Streamers have little visual indication if the network is impacting the UL streaming of its video. It is therefore recommended to assess the success of this service from the point of view of the receiving client. The set-up and UL network for the streamer act as a limiting factor for the highest video quality that can be achieved for the system. In addition, viewer conditions can further degrade the QoE.

For this use case, several recommended objective models for network and service monitoring already exist, listed in Appendix I.

Some care has to be taken depending on the type of data loss handling that is used in the service.

For reliable services where no late loss is allowed, services attempt to adapt the streaming compression to make it possible for the video to pass through the network. If the conditions degrade enough and the client jitter buffer gets too low or empty, a stalling event will occur to allow the client to buffer some extra video before restarting the playout. Models such as [b-ITU-T P.1203] and [b-ITU-T P.1204] cover this use case.

In contrast, for unreliable services where losses can occur, content is skipped in various amounts when degradations that are too large happen. Services using unreliable transport protocols can induce errors on a sub-frame level, trying to conceal losses on a macroblock level or with full-frame-skipping where the client simply throws away the video frames until something without errors can be recreated. Bitrate adaptation can also be used here to limit degradation occurrences, but possibly errors will be large enough or the data arrive late enough to compel the decoder to deal with an incomplete frame. In this case, models from [b-ITU-T P.1201] and [b-ITU-T P.1202] can be used, in tandem with the integration module contained in [b-ITU-T P.1203].

7.1.6 Social networking

SN is one of the most throughput-hungry use cases today. People post media files to SN platforms and consume other people's posts, often using a lot of DL throughput while also requiring large UL throughput.

Requirements are lower in terms of latency, while the throughput is the main focus. With the increased quality of media files and expectations from the user of good interactivity with the app, throughput demand can easily reach hundreds of megabits per second.

7.1.7 Telemeeting

Since the COVID-19 pandemic, the necessity to telemeet has increased exponentially, both in personal and business settings. TMs can include voice, video and media sharing, and one of the main objectives is smooth interactivity among participating users.

7.1.8 Virtual reality

From clause 6.2 of [b-ITU-T P.1320],

"Virtual reality immerses users in a fully artificial (computer-generated) digital environment. The term virtual reality refers to a computer-generated environment in which the user can perceive, feel and interact in a manner that is similar to a physical place. This is achieved by combining stimulation over multiple sensory channels – such as sight, sound and touch – with force-feedback, motion tracking and control devices. In an ideal virtual reality ... system, the user would not be able to distinguish an artificial environment from its physical counterpart. While none of the current VR systems would be able to pass this criterion, significant advances in the perceptual fidelity of virtual environments have been achieved over the last few years. The virtual environment itself is a computer-generated three-dimensional model of a physical environment, in which the user can navigate and interact with the objects that it contains."

7.2 Definition of network scenarios

In contrast to what is available in the literature [b-5G PPP], this Recommendation aims at the perspective of the end user rather than an aggregated network-wise vision, presenting some of the main 5G scenarios in this clause. 5G usage scenarios have been divided into macro-categories according to their commonalities in network deployment and channel conditions. Each usage scenario is then subdivided into user and traffic present in the surroundings, as described in Table 7-2.

Table 7-2 – Network scenarios

Scenario description	Sub-scenario description
Urban	Office
	Pedestrian
	Driving, normal
	Driving, traffic jam
	Driving, emergency driving
	Internet café, busy hours
	Internet café, off-peak
Suburban	Festival
	Stadium
	Shopping mall, busy hours
	Shopping mall, off-peak
High-speed train (HST)	Relay
	Direct passenger connections

7.2.4 Urban

Dense network deployments generally characterize urban scenarios. For example, consider locations such as London city centre, Manhattan or Tokyo. Network deployments in urban scenarios are usually characterized by the presence of macro sites with a reduced inter-site distance, along with the presence of small cells. Furthermore, given the need to carry a high amount of traffic, a dense heterogeneous network is required, allowing the terminal to be connected to small cells as primary cells.

Under this scenario, several 5G sub-scenarios characterized by the location and mobility of the users have been identified.

7.2.4.1 Office

This is the typical indoor office environment where users are static. It is characterized by indoor wireless network access points (wireless fidelity (Wi-Fi) and small cells). The channel and interference conditions are also dictated by the penetration of signals from outdoor cells, mostly macro sites. The expected traffic scenario is related to typical broadband access with various applications.

7.2.4.2 Pedestrian

This is the usual scenario of a user walking down an urban street at around 1.4 m/s (5 km/h). It is expected that the channel conditions can change due to moving obstacles, e.g., trucks and buses. Even in this case, typical broadband access is a baseline for traffic-generating applications.

7.2.4.3 Driving

It is becoming increasingly common to have wireless modems inside cars that allow direct connectivity to the onboard computer and entertainment system. Given the urban environment, the expected speed of the vehicle is up to 60 km/h (16 m/s). Conditions depend on the app, since vehicle to infrastructure traffic for vehicular control could also be foreseen.

7.2.4.3.1 Normal

In normal vehicular traffic conditions, the speed is expected to be at the high end of urban limits, possibly 40 to 60 km/h (11 to 16 m/s). User density on the road here is reasonably low, given the regulatory distance between vehicles.

7.2.4.3.2 Traffic jam

In a traffic jam, it is expected that vehicles will be more packed, resulting in a higher density, while the speed tends to be very low to almost static. Users are then expected to make heavier use of, for example, their entertainment system creating a heavier load on the network.

7.2.4.3.3 Emergency driving

Emergency driving can be performed by public entities acting on emergency calls for safety and security, e.g., ambulances, police or firefighters. Despite the urban settings, driving speeds higher than typical urban limits must be considered, up to 120 km/h (33 m/s).

7.2.4.4 Internet cafe

The scenario is a typical multi-radio access technology (RAT) scenario with users mostly static, with some channel variations due to the street-level nature of the location (moving objects on the street, some indoor penetration). Different scenarios can be designed during busy hours, e.g., lunchtime, and during off-peak hours, e.g., early afternoon.

7.2.4.4.1 Busy hours

During busy hours, the cafe is expected to be full of customers trying to access the network for consumer services. This affects the network load and the resources that each customer can get from the network.

7.2.4.4.2 Off-peak

During off-peak hours, the number of customers in the cafe decreases, creating easier conditions for network access. At the same time, the types of imperfections that affect the channel (interference from other customers, amount of moving shielding objects) is also reduced.

7.2.5 Suburban

The suburban umbrella contains all those scenarios generally located outside the city centre, in less densely populated areas. For this reason, most sites of access to the network are macro only, mounted on higher-rise towers for coverage, with a more significant inter-site distance.

Few sporadic small cell hotspots are eventually encountered within the coverage area of the macros, either for localized capacity boosting or to reinforce coverage. All sub-cases of interest in these networks and propagation conditions are related to the concentrated and massive presence of users.

7.2.5.1 Festival

This usage scenario is typically outdoor, with additions of a few localized macro sites and small cells for increased capacity. The expected traffic type is mostly consistent with very dense broadband access, with possible broadcasting of concerts from the festival area.

7.2.5.2 Stadium

This is the scenario with the highest density of users. Hotspots cover the stadium area, but its main characteristic is balancing the number of access points and the interference generated.

7.2.5.3 Shopping mall

Large indoor spaces with the presence of certain materials reflecting electromagnetic radiation characterize the physical environment. Diverse applications are expected to run in the network for businesses and external customers. Similar to Internet cafes, vastly different behaviours can be observed during busy hours, e.g., on a Sunday afternoon, and during off-peak hours, e.g., on a weekday afternoon.

7.2.5.3.1 Busy hours

Considering busy hours, the shopping mall is expected to be flooded with customers, and business internal and customer traffic is expected to overload the network.

7.2.5.3.2 Off-peak

Fewer customers are present on the premises, partly changing not only traffic, but also propagation conditions.

7.2.6 High-speed train

High-speed collective transport is considered a valuable scenario for 5G, where both commercial traffic and critical vehicular control should coexist in the same radio access. The propagation condition is rural, with relatively sparse macro sites densified for the railway coverage purposes. Train speeds are very high, up to 500 km/h (140 m/s). Two different ways of dealing with user affiliations represent the main technical challenges for 5G and how the system should behave.