ITUPublications
Recommendations

International Telecommunication Union
Standardization Sector

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.

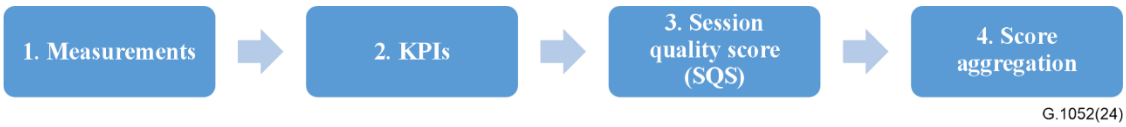



**Figure 6-1 – Scoring methodology**

### 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.

#### 7.2.6.1 Relay

One way of dealing with user affiliations to the network is by installing a relay (or self-backhaul as sometimes referred to in 5G literature) or an access gateway for Wi-Fi on the train. Since only one device is affiliated with the fixed wireless network, the control plane (C-plane) is not the bottleneck. Still, the aggregated traffic of all the users creates a massive "super-user," which requires a tremendous amount of the user-plane (U-plane) resources.

#### 7.2.6.2 Direct passenger connections

In contrast to the previous scenario, each user is dealt with individually by the fixed network. While each user has limited U-plane resources to be scheduled, the C-plane and the so-called massive handover can create difficulties in networks and impairments to data flow for users, impacting their QoE.

### 7.3 Network scenario parameters

Wireless cellular networks have a vast number of parameters to be set to ensure the correct functioning of the system. Besides the parameters of the technology itself, while specifying scenarios, more fine-grained details about the environmental conditions need to be set in the testbed.

Some parameters to be set are dependent on the DUT. Various devices support features from different technology releases. This factor partially limits the number of typical parameters specified in the scenario description for testing purposes. Furthermore, different technologies need different parameters to describe scenarios.

Parameters such as antenna configuration, transmission mode, and maximum achievable modulation and coding scheme (MCS) should be considered device dependent and should be set to achieve maximum stable performance for the DUT.

Other parameters are set to the *de facto* default values. Examples of such parameters are listed in Table 7-3.

**Table 7-3 – De-facto default values for common parameters**

| Parameter | Value |
|---|---|
| Cyclic prefix length | Short |
| Max number of hybrid automatic repeat request (HARQ) re-transmissions | 4 |
| Block error rate (BLER) target for the first transmission | 30% |
| Standard bearer QoS | Default bearer |

Several sub-scenarios have been specified to guarantee the most realistic conditions in each scenario. Each sub-scenario depicts, for example, a different time of the day when a different number of users are present in the network, or the environment can suffer from more radical changes. For presets for these parameters, see Appendix III.

### 7.3.1 Network scenario applicability

Test cases should be executed for each network scenario applicable to each use case, as listed in Table 7-4.

**Table 7-4 – Use case/network scenario applicability**

| Network scenario | Use cases | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | AR | CG | CDSS | HSI | LSS | SN | TM | VR |
| Urban office | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ |
| Urban pedestrian | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |

**Table 7-4 – Use case/network scenario applicability**

| Network scenario | Use cases | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **AR** | **CG** | **CDSS** | **HSI** | **LSS** | **SN** | **TM** | **VR** |
| Driving normal | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| Traffic jam | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | |
| Emergency driving | ✓ | | | ✓ | ✓ | ✓ | ✓ | |
| Internet café – Busy hours | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ |
| Internet café – Off-peak | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ |
| Festival | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| Stadium | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| Shopping mall – Busy hours | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | |
| Shopping mall – Off-peak | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | |
| High-speed train relay | | ✓ | ✓ | ✓ | | ✓ | ✓ | |
| HST direct passenger connection | | ✓ | ✓ | ✓ | | ✓ | ✓ | |

# 8 Testing methodology

The use cases considered in this Recommendation are specified in clause 7.1. To make the evaluation of the user experience feasible, a systematic approach to decompose the problem into several actionable tests running on the testbed and generalize the QoE computation is needed. Tests are run considering an AUT running on a DUT, attached to the testbed. In this clause, the testing methodology is broken down and explained.

## 8.1 Test system configuration

The testbed architecture is described in clause 10.

The system configurations in clauses 8.1.1 to 8.1.3 shall serve as starting points for the experimenter. Depending on the experimental design, further configuration steps might be taken to ensure a correct set-up.

If human interaction or external recording is needed, the environment must be properly organized to avoid external factors influencing the test. The environment should be set up based on the target test. If objective models are standardized that provide SQSs, the respective configuration shall be used according to the associated standards. A list of recommended objective models can be found in Appendix I. If subjective tests are required because no SQS objective model has been standardized, respective subjective test methods shall be applied, such as those listed in Appendix I.

### 8.1.1 Testbed configuration

The testbed is configured to perform the test cases using one of the bands supported by the DUT. The selection of the frequency band depends on the test, but it shall be declared in the final test report. Unless required by the test, the same band shall be used across experiments to avoid uncontrolled variations of the results.

If needed by the specific test designed for the use cases, the testbed shall provide a way to rotate the DUT with sufficient degrees of freedom to perform the test, e.g., three translational axes, three rotational axes or six roto-translational axes.

The testbed is configured as follows.

- The testbed performs test cases with one of the frequency bands supported by the DUT.
- The testbed network is initially as specified by [3GPP TS 36.508] for long-term evolution (LTE), or [3GPP TS 38.508-1] for 5G new radio (NR). The experimenter must ensure that the same version of the specifications is applied across experiments; otherwise results are impossible to compare.
- The testbed has been successfully calibrated (see clause 9).

### 8.1.2 Device under test configuration

The DUT is grid powered if feasible. If the DUT is battery powered, the battery charge shall never fall below 50%. This can be done either by fully charging the phone before the test begins, if the test is not power-hungry, or by providing both connection and power when connecting the DUT to the testbed. If the test is power-hungry, the battery charge should be closely monitored.

The DUT shall be configured as follows.

- The DUT is ON and there are no applications running other than the required system applications.
- No antivirus or anti-malware app is running.
- Maximum brightness available.
- Maximum screen resolution available.
- No energy saving option is enabled.
- Do not disturb mode is enabled if available.
- Date and time configured to a target time zone. It can be either the local time zone or coordinated universal time.
- The DUT is allowed to use mobile data, thus being able to connect to the testbed.
- Mobile data usage is not restricted.
- The DUT Wi-Fi is turned off.
- The DUT audio volume is configured at the middle of the available range.
- The DUT has been turned on for at least 3 min to allow all boot processes to be completed.
- The DUT screen is configured with screen always on.
- The DUT has been successfully calibrated (see clause 9).

### 8.1.3 Application under test configuration

The AUT is configured as follows.

- The AUT is deployed in the DUT.
- Consent the necessary permissions to the AUT (e.g., microphone or camera)
- If needed, the AUT is logged in with correct test user.

## 8.2 Test case initial conditions

The following list of initial conditions shall serve as a starting point for the experimenter. Depending on the experimental design, more initial conditions might be needed to ensure a correct set-up.

Initial test conditions are specified as follows.

- The testbed, DUT, and AUT are configured as specified in clause 8.2.
- The testbed and the DUT are connected as described in clause 10.

- The testbed is configured according to the target network scenario (see clauses 7.2, 7.3 and Appendix III).
- The AUT is installed on the DUT.
- The testbed has established a data path with the DUT.
- The AUT backend service is accessible from the testbed.
- The testbed has cleared the AUT stored data and cache.
- The testbed has initiated the AUT.

### 8.3 Test case post-conditions

The following post-conditions shall serve as a starting point for the experimenter. Depending on the experiment design, further post-conditions might be taken to ensure a correct tear down. Post conditions should mirror the initial conditions, in order to return the AUT, DUT, and testbed to a state where the next test can be run with a cold start.

The following post-conditions shall be executed after the test case steps:

- The testbed terminates the AUT.
- The testbed shuts down the data path with the DUT.

### 8.4 Application test case

A test case comprises a set of actions that the experimenter shall take, and information that shall be recorded. Specify the following.

- A unique identifier for the test case that will be used to connect a database containing the relevant information about the test with the information (e.g., the measurements) recorded while running the test.
- The objective of the test.
- The configuration of the testbed, DUT, and AUT (see clause 8.1).
- The initial conditions (see clause 8.2).
- The post-conditions (see clause 8.3)
- Test steps that need to be taken during the test. Test steps are closely related to the target use case and AUT, as well as with the objectives of the test. Test steps should also ensure the correct configuration of the system, set the proper initial conditions, and ensure post-conditions on termination. Steps should be automated whenever possible, to ensure repeatability and scalability.
- Measurements that need to be obtained while running the test and their recorded values. Save them accordingly, using the unique test identifier.

## 9 Testbed calibration

The testbed shall be calibrated and proper compensations for specific measurements shall be made against the baseline before executing any test campaign. This clause describes the necessary calibration procedures needed for the testbed. This includes the calibration of the individual measurement instruments (clause 9.1), the characterization of the testbed (clause 9.2), radio frequency (RF) connection compensation (clause 9.3) and measurement compensation (clause 9.4). Finally, clause 9.5 describes the calibration and compensation process, and clause 9.6 shows how to track testbed health over time.

## 9.1 Calibration of individual measurement instruments

This clause describes the calibration requirements of the individual measurement instruments in the testbed.

Calibration is the relationship between a measurement standard of known uncertainty and test equipment of unknown uncertainty. To perform calibrations, a laboratory must be able to establish a complete chain of traceability from its standard to a nationally or internationally recognized standard held by an organization (such as the National Institute of Standards and Technology in the USA).

The calibration also determines the uncertainty of the measurement instrument, and it can have an expiration date. Electronic components, in fact, can degrade over time, both with and without usage.

The test instruments used in the testbed shall be individually calibrated and verified, either by the manufacturer or by an accredited third party laboratory.

## 9.2 Characterization of the testbed

This clause describes the characterization of the overall performance of the testbed in terms of the average RTT and the maximum achievable throughput in the testbed. These two have been selected as the most relevant parameters to establish the testbed's baseline and serve as reference.

To facilitate the measurement and make it repeatable, an automated characterization test case should be developed (see clause 10.5). Once measurements are made, they should be saved in an appropriate characterization file.

Average, minimum and maximum values shall be validated and, if the results differ significantly from reasonable values, the testbed should be inspected and fixed. Reasonable values are highly dependent on the testbed implementation, the specific Internet protocol (IP) addresses, the local area network (LAN) topology and connectivity, and the geographical location of the testbed when testing for external connectivity.

### 9.2.1 Round-trip-time

The RTT shall be measured towards multiple target IP addresses to characterize what a DUT will experience in the testbed. An application such as `ping` can be used to perform the test. It is important to characterize the minimum and maximum, and RTT variation to understand its intrinsic property and potential use case limitation.

Ping should be performed from the DUT. The relevant target IPs to ping are:

- a data server within the testbed;
- every other relevant component within the testbed LAN (e.g., components of the core network, orchestration VMs, data server).

Furthermore, if the test requires a connection to the Internet (outside the LAN):

- Internet gateway;
- geographically close external server (e.g., the Internet service provider of the testbed or a public domain name system IP address, 8.8.8.8);
- geographically far server (e.g., domain name system servers from a geographically distant country);
- if known, the server hosting the communication with the AUT.

Measuring the RTT of servers outside of the LAN is inherently unstable, and values might change over time. To appropriately take into account these variations, measurements should be performed throughout the duration of the experiment (see clause 9.4).

### 9.2.2 Throughput

The maximum supportable throughput of the testbed must be characterized. The configured data rate in each test case should not exceed the maximum throughput of the testbed. The testbed should support a minimum throughput based on the target use case, and this characterization shall ensure that the available throughput is indeed higher than the minimum required.

Evaluation of the maximum throughput of the testbed includes:

- throughput test within the testbed LAN;
- throughput test towards known servers outside the testbed LAN (if necessary and possible).

### 9.2.3 Time synchronization

Measurements might be taken from the DUT, as well as from different components making up the testbed. Typically, it is important to know when these measurements were taken, and thus all relevant components should either share a common clock or time synchronization that should be compensated to a sufficient precision. Such precision might vary depending on the KPI itself, e.g., if it is related to human perception (a few to 10s of milliseconds) or, e.g., network event (possibly sub-millisecond).

## 9.3 Radio frequency connection compensation

RF connection compensation differs based on the testbed architecture, specifically, if conducted or over the air (OTA) connection is used (see clause 10.2.1). In both cases, the path that the signal takes starting from the testbed and reaching the DUT must be taken into account and properly compensated to enable reproducibility and repeatability.

In conducted tests, cable and connection losses have to be compensated for. In OTA tests, the unknown relative position, distance, orientation and polarization of transmitter (Tx) and receiver (Rx) antennas have to be compensated for.

The loss is measured in DL per frequency band of interest on the DUT side. If possible, it should also be measured per DUT Rx antenna, e.g., for conducted connection to individual DUT Rx antennas. The procedure is achieved via a test plan that:

- generates RF DL signals with known power;
- measures the difference between the configured DL power and measured reference signal received power (RSRP) on the DUT side;
- stores the values in a compensation table for the specific DUT;
- compensates the output power according to the compensation table.

This approach guarantees identical received pilot power for all devices in the testbed for each test campaign.

## 9.4 Measurement probes

The DUT should run several measurement probes in parallel with the AUT, capturing the AUT resource usage and the DUT state during the test. Probes are typically assumed to be software probes, although hardware probes might be needed for specific tests (e.g., power consumption).

These probes track and log performance during test execution and might also impact central processing unit (CPU), graphics processing unit (GPU) and memory usage, possibly affecting the performance of the AUT. To quantify how many resources are used for each probe, their impact needs to be characterized for each DUT. A probe shall not be used if it monopolizes a key resource for the AUT.

Relevant resources to monitor on the DUT are, e.g., battery level, CPU usage, GPU usage, storage busy time, available or total storage, free or total random-access memory (RAM), temperature, network usage and network performance (like RTT). The monitor periodicity can vary from resource to resource, depending on its variability and on the extra resources used by the monitor itself. These

monitors can be used to warn the experimenter whether resources crossed critical thresholds (e.g., low battery level, high CPU usage or low available storage), and thus help to discard tests where resources are outside healthy conditions.

Other components from the testbed could also be monitored (e.g., control personal computer, data server, radio access network (RAN) emulator), although this is secondary and left to the discretion of the experimenter.

### 9.5 Baseline calibration standard reference

Once the testbed has been calibrated in accordance with clause 9.1, a calibration table should be generated, summarizing the characterization of the testbed as described in clause 9.2, the measurement of RF connection compensation from clause 9.3, and the impact of measurement probes on resource usage as described in clause 9.4.

The calibration table verifies set-up stability over time, best interprets the obtained results and early issue detection, and acts as a baseline reference for each new DUT connected to the testbed. The table should be reported together with the campaign results so that the reader is informed that calibration and compensation were performed and what the state of the system is, thus better understanding the campaign results.

### 9.6 Testbed health tracking

Before running a complete test campaign, a short test campaign, which consists of a subset of calibration test cases, shall be executed to check and validate the health of the testbed. The obtained testbed calibration results should be compared with the baseline results. If there are no differences, the testbed is ready for use. If there are noticeable differences between the calibration results and no environmental changes have been made in the testbed, necessary actions shall be taken to identify and solve the originating issue.

The testbed health tracking procedure should be repeated whenever an event that could potentially cause changes in the environmental settings of the testbed happen, or a new element has been introduced into theme testbed. Examples of such events include, but are not limited to:

- a new device is connected to the testbed;
- a calibrated device is disconnected and reconnected to the testbed;
- a calibrated device is moved within the shielding box;
- cabling or OTA antenna are changed, a new shielding box, etc;
- software updates in any part of the testbed or DUT;
- test parameter change, e.g., RF band change.

As specified in this Recommendation, the testbed calibration process shall be repeated, and new tables shall replace previous baseline results if necessary.

## 10 Testbed description

This clause describes the structure of the testbed and its main components.

Starting from the DUT and the AUT, clause 10.1 suggests how to choose the appropriate device(s) and app(s) to perform the desired tests. Next, clause 10.2 describes what the testbed should include, specifying every part of it and setting minimum requirements, clause 10.3 discusses how the AUT can obtain data from the service provider, and clause 10.4 shows how measurements can be taken from the testbed. Finally, clause 10.5 discusses how the testbed can be coordinated and automated.

The overall testbed is graphically represented in Figure 10-1.

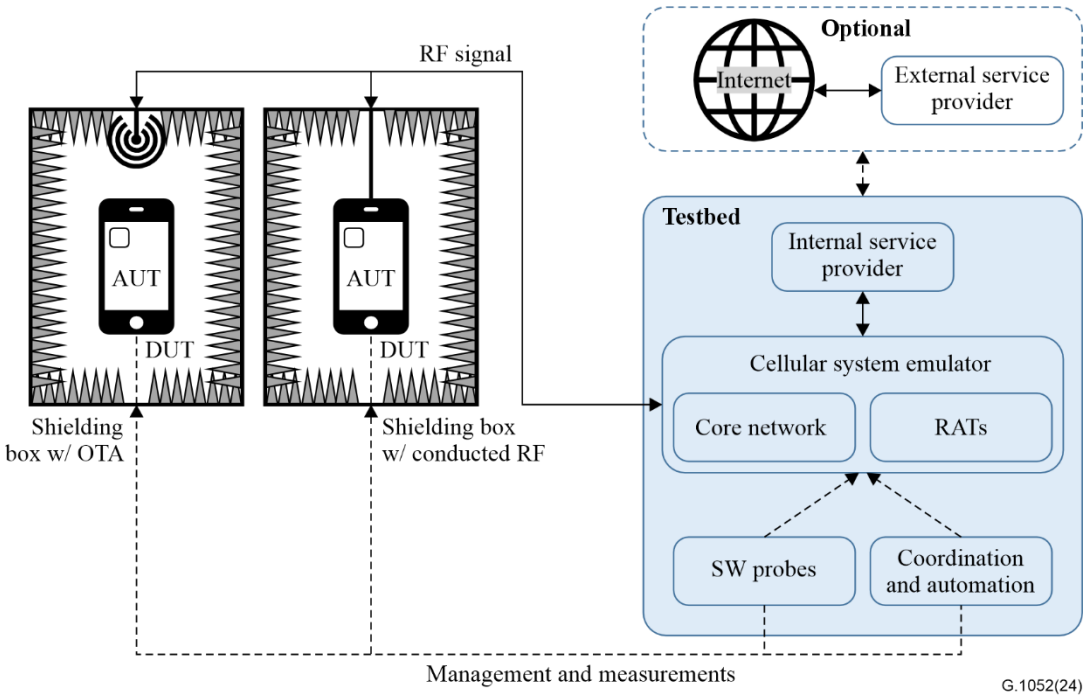



**Figure 10-1 – Test system architecture overview**

NOTE – The purpose of this figure is to give a simple idea of the main components of the testbed, without showing the details.

The DUT(s) within the shielding box can be connected to the cellular system emulator either via OTA connection or via conducted RF connection. Probes, coordination, and automation software communicate with the DUT via a secondary management link. This is done to avoid introducing extra undesired traffic in the data link and to have a secure, unimpaired link for management purposes.

## 10.1 Devices, operating systems, and applications

Results from a testbed are more useful if they are connected somehow with reality. For an experimenter, exploring a market of interest may make sense to understand the most popular devices (manufacturers and models), operating systems (OSs), and applications.

It should be taken into account that different markets (such as different regions, wealth, or demographics) can have other preferences, choosing, for example, high- vs. low-end brands and models or free vs. paid applications.

When communicating results, it is crucial to clarify all the details of the testbed, including DUT and testbed specifications. This is because different testing conditions can yield different results, even if the testbed is kept the same. A simple example is the type of modem that high- vs. low-end devices can include, the feature set of newer vs. older devices, the details of protocols, and, generally, networking software used by specific versions of an app, etc.

Technical reasons might also guide how to create the final set-up. For example, open-source apps are more manageable and flexible from an experimenter's point of view, although any modification to the app might reduce the accuracy of the obtained results. Furthermore, proprietary systems may require the use of resources owned by application developers, which might not be a convenient approach.

## 10.2 Communication system

### 10.2.1 Radio frequency connection

The testbed integrates channel emulation and digital generation of impairments such as additive white Gaussian noise (AWGN), a critical feature for achieving high accuracy when setting signal-to-noise ratio (SNR) conditions. In particular, standard multipath fading profiles defined by the Third

Generation Partnership Project (3GPP) are supported to emulate reference propagation conditions (clause 10.2.3). The signal from the testbed has to be brought to the DUT with minimal distortions and external factors affecting it. Different types of connections can be used, each with varying degrees of accuracy, stability over time, and complexity.

#### 10.2.1.1 Radio frequency shielding box and anechoic chamber

The device should be enclosed in a shielding box to shield the DUT from external connections, bringing it to ideal conditions, while shielding the outside from radiation that might be produced by the wireless components. The shielding box shall have a way to allow the DUT placed inside to connect to the testbed, both for RF signalling and control (often done over universal serial bus ports). If an OTA connection is used, the shielding box should also act as an anechoic chamber, creating ideal propagation condition which allow full control of the channel for the experimenter. See Figure 10-2.

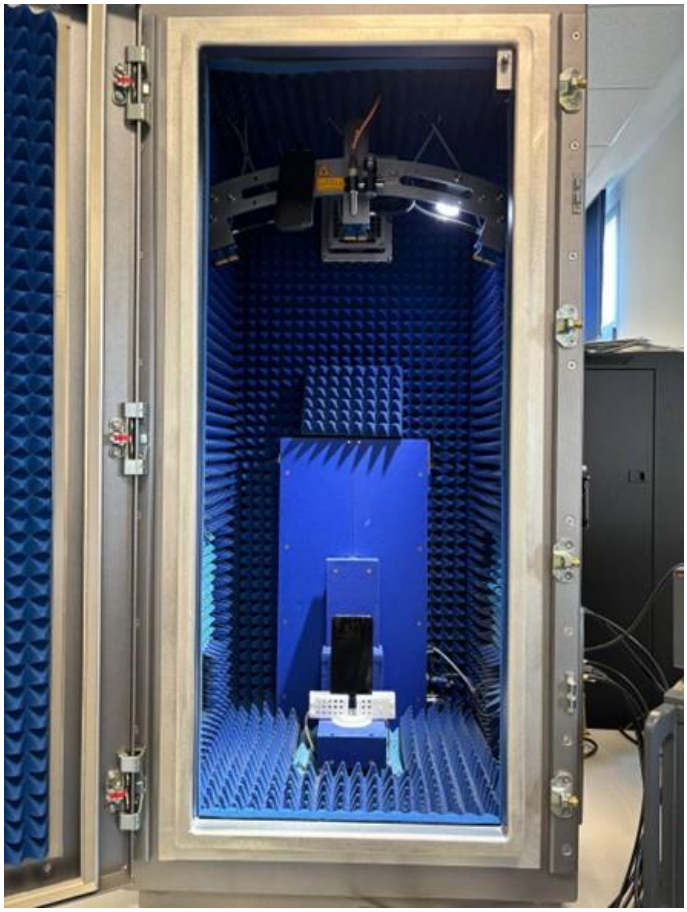


**Figure 10-2 – Example of an anechoic chamber**

#### 10.2.1.2 Conducted radio frequency connection

To preserve the radio conditions configured on the testbed at best, the radio connection can be conducted through cables attached directly to the antenna ports of the DUT. This is often a complex and costly operation, but it creates a very stable RF connection. Not all modern phones might support this configuration, the experimenter shall make sure to find out whether this is a possibility at all. See Figure 10-3.

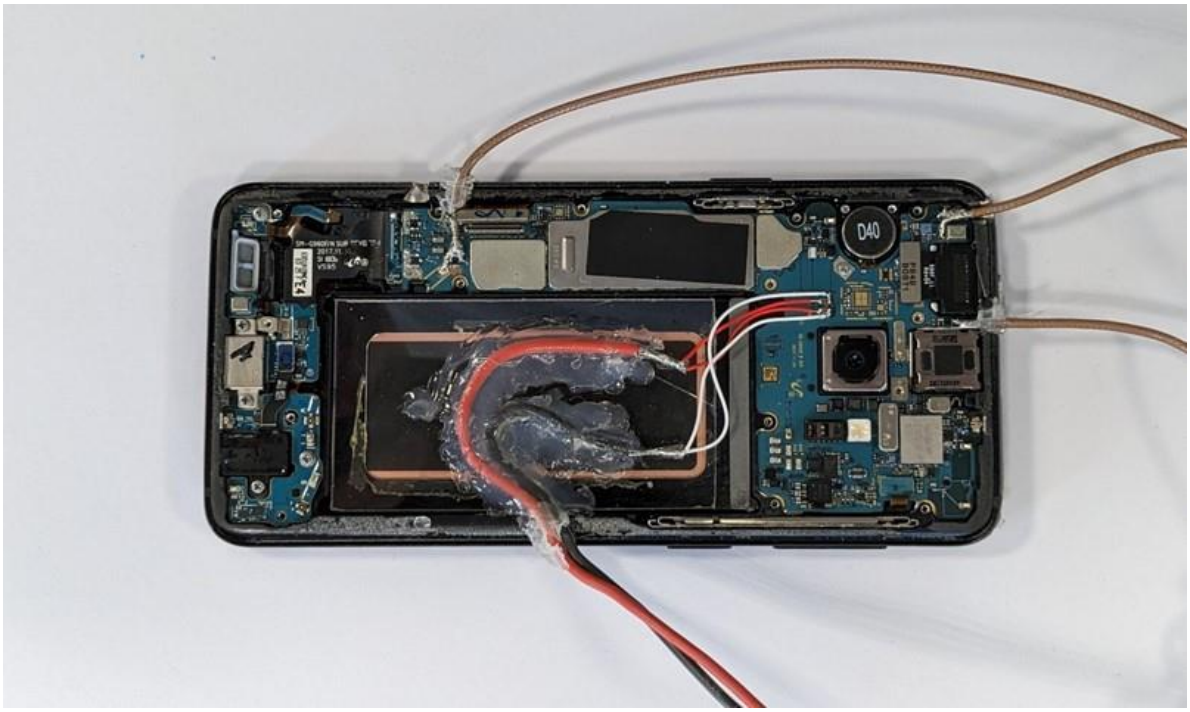


**Figure 10-3 – Example of device under test modified to support conducted radio frequency connection**

NOTE – In this case, the battery has also been removed to support power consumption analysis.

#### 10.2.1.3 Over-the-air connection

OTA connections use antennas to transmit the signal from the testbed to the DUT, similar to what happens in real deployments. Antennas can be of different types, and the relative placement between the device and the antenna can affect the received signal. This makes the configuration less stable, and calibration should thus be performed every time the DUT, the antennas, or the anechoic chambers are moved and interacted with.

#### 10.2.1.4 Comments on radio frequency connections

Both conducted and OTA connections can potentially be used within the scope of this Recommendation, although they have intrinsically different performance and use cases.

On one hand, conducted RF connection, once properly calibrated (see clause 9):

- is stable over time;
- is reliable once cables are properly attached ensuring a perfect reception of the desired signal to the DUT;
- allows for reproducible results over different systems.

Though, the approach also has some downsides, for example:

- unless the DUT supports conducted connection natively, equipping a DUT with proper cabling is an operation only possible for specialized companies, which can be extremely expensive;
- with most modern phones having four to eight antennas or more, each one of them is separately connected to the testbed's RFIO, making testing on multiple devices fiddly;
- connectorizing a device bypasses its antenna system, which might play a key role in its overall performance in the real world – this makes it impossible to fully characterize its performance, and does not allow a fair comparison among different devices.

On the other hand, OTA connection:

- does not require crucial hardware modification of the DUT;
- the set-up simply involves placing the DUT correctly within the shielding box.

However, note the following.

- The signal generated by the testbed is affected by the hardly controllable wireless channel, adding an unknown element to the test. Specifically, if the objective is to run a carefully emulated channel between the Tx and the Rx antennas with proper antenna correlation matrices, especially if multiple-input multiple-output (MIMO) is deemed as important, then the wireless channel will make this extremely hard in practice. Even in an isolation chamber, signals from all Tx antennas will mix at all Rx antennas, uncontrollably mixing the signals. Although wireless cable testing techniques exist, they are very hard to implement in practice, and hardly possible with commercial devices. OTA testing should thus rely on simple channel models, where different MIMO channel conditions do not play a key role.
- Stability over time is also harder to guarantee as the channel can greatly vary with minor changes within the radiated environment. To reduce this effect some actions might be taken, for example: (i) a high-quality isolation must be ensured; (ii) robust positioning systems must be in place to hold both the DUT and the radiating elements still throughout the duration of a test (e.g., bulky antennas should be secured in position), especially if the environment around the isolation chamber might produce vibrations or movement (e.g., powerful fans from a nearby server, moving equipment around the lab, heavy human presence in the lab); and (iii) vibration from the DUT should be disabled.
- Reproducibility is also not easy to ensure given the delicate relative positioning of Tx and Rx antenna elements. More frequent testbed health tracking should be performed (clause 9.6), especially regarding RF connection compensation (clause 9.3).

In general, conducted testing is the main method used in < 6 GHz 5G NR frequency range (FR1), while radiated or OTA testing is the main method used in the 5G NR mmWave spectrum (FR2) [b-3GPP TS 37.901-5]. Within the scope of this Recommendation, OTA testing can also be used in FR1, although great care must be taken when setting up the OTA system.

### 10.2.2 Radio access technologies

#### 10.2.2.1 Wi-Fi access

Although Wi-Fi does not lie within the scope of this Recommendation, the system may be able to connect to a Wi-Fi access point to conduct testing and coordinate Wi-Fi access points. The following parameters can be controlled to emulate appropriate conditions to match the desired end-user context use case (home, office, public environment, among others):

- operating channel;
- transmission power;
- channel bandwidth;
- radio status (on/off).

#### 10.2.2.2 3GPP access

The system shall have the ability to emulate a mobile data connection access with the following parameters:

- supported formats;
  - LTE;
  - LTE-advanced;
  - NB-Internet of things;

- 5G NR non-standalone/standalone (NSA/SA).

– Basic cell configuration (LTE).
  - Duplex Mode: frequency division duplex (FDD) and time division duplex (TDD).
  - DL bandwidth: 1.4 MHz, 3 MHz, 5 MHz, 10 MHz, 15 MHz and 20 MHz.
  - UL bandwidth: 1.4 MHz, 3 MHz, 5 MHz, 10 MHz, 15 MHz and 20 MHz.
  - Mandatory frequency setting method: The system shall support specification of the operating frequency by selecting the evolved universal terrestrial radio access (E-UTRA) absolute radio-frequency channel number (EARFCN). Other methods to select the operating frequency might be available.
  - LTE Bands supported in FDD: 1 to 32 (excluding 15 and 16), 252 and 255.
  - LTE bands supported in TDD: 33 to 44.
  - TDD frame configuration: 0 to 6 (Release 8). Additional frame configurations, as specified in later releases, are optional.
  - TDD special subframe configuration: 0 to 8 (release 8) and 9 (release 11).

– Basic cell configuration (5G NR).
  - Duplex Mode: FDD NSA/SA and TDD NSA/SA.
  - DL bandwidth: 10 MHz, 15 MHz, 20 MHz, 25 MHz, 30 MHz, 40 MHz, 50 MHz, 60 MHz, 70 MHz, 80 MHz, 90 MHz, 100 MHz.
  - UL bandwidth: 10 MHz, 15 MHz, 20 MHz, 25 MHz, 30 MHz, 40 MHz, 50 MHz, 60 MHz, 70 MHz, 80 MHz, 90 MHz, 100 MHz.
  - Mandatory frequency setting method: The system shall support specification of the operating frequency by selecting the EARFCN. Other methods to select the operating frequency might be available.
  - NR bands supported in FDD: FR1/n1, FR1/n2, FR1/n3, FR1/n5, FR1/n7, FR1/n8, FR1/n12, FR1/n13, FR1/n4, FR1/n18, FR1/n20, FR1/n24a, FR1/n24b, FR1/n25, FR1/n26, FR1/n28, FR1/n30, FR1/n65, FR1/n66, FR1/n70, FR1/n71, FR1/n74.
  - NR bands supported in TDD: FR1/n34, FR1/n38, FR1/n39, FR1/n40, FR1/n41, FR1/n48, FR1/n50, FR1/n51, FR1/n53, FR1/n77, FR1/n78, FR1/n79, FR1/n90, FR2/n257, FR2/n258, FR2/n259, FR2/n260, FR2/n261.
  - The 5G testbed shall provide a mechanism to configure a repeating pattern of different frame configurations, each of which has a different slot map that contains different slot configurations. Each slot configuration shall contain different parameters for configuring UL and DL scheduling. Configurable parameters are: the transmission periodicity, DL/UL slots, and DL/UL symbols; using a reference subcarrier spacing (SCS).

### 10.2.3 Radio frequency channel models

The testbed shall be able to emulate different channel conditions in signal levels, fading profiles, and noise and interference levels.

#### 10.2.3.1 Long-term evolution propagation conditions

Table 10-1, extracted from [b-3GPP TR 37.901], specifies *de facto* conditions to challenge the higher layers and covers the vast majority of 5G use cases experienced in field tests described in clause 7.1.

**Table 10-1 – Test points for long-term evolution [b-3GPP TR 37.901]**

| Propagation condition | SNR | Justification |
|---|---|---|
| Static | No interference | To check that upper layers do not constrain data throughput. |
| Extended pedestrian A 5 (EPA5) | 20 dB | To exhibit significant transport block size variations (see clause 5.5.4.3 of [b-3GPP TR 37.901],) and ubiquitous scenarios for high-data-rate requiring processing capability. |
| Extended vehicular A 5 (EVA5) | 10 dB | EVA frequently occurs in deployments. |
| EVA70 | 20 dB | Adds EVA70, high SNR coverage, which is common in low frequency (< 1 GHz) band networks. |
| EVA200 | 20 dB | Covers high Doppler, high SNR scenario, which is common in high frequency (≥ 1 GHz) band networks. |
| extended typical urban 70 (ETU70) | 0 dB | Fast variations and the most common high-dispersion case. |
| ETU300 | 0 dB | A high BLER scenario may trigger higher layer retransmissions and addresses the high-speed scenario in the work item objective. |
| NOTE – See clause B.2 of [b-3GPP TS 36.101] for specifications of EPA, EVA and ETU. | | |

The system shall support the emulation of the following propagation conditions, specified in clause B.2 of [b-3GPP TS 36.101]:

- EPA5;
- EVA5;
- EVA70;
- EVA200;
- ETU70;
- ETU300.

### 10.2.3.2 New-radio propagation conditions

Multiple 5G NR propagation conditions shall be supported as follows.

– Static: Static channel model, in which to choose to use 3GPP static MIMO channel models

– Tapped delay line (TDL): These propagation conditions for 5G NR are based on [b-3GPP TR 38.901]. Specifically, some models are reported based on TDL-A/B/C (non-line-of-sight models), where the number indicates the desired delay spread (DS) in nanoseconds.

  - TDLA30, TDLB100, TDLC60, TDLC300

– HST: can be based on single frequency networks (SFNs) or dynamic point selection (DPS) based on [b-3GPP TS 38.521-4], followed by an indication of the Doppler shift.

  - HST-SFN870, HST-SFN1667, HST-DPS870, HST-DPS1667

See Table 10-2.

**Table 10-2 – Propagation conditions for 5G new radio**

| Propagation condition | Justification |
|---|---|
| Static | Static channel model, in which to choose to use 3GPP static MIMO channel models |
| TDLA30 | Based on the TDL-A model from [b-3GPP TR 38.901] with a short delay spread of 30 ns |
| TDLB100 | Based on the TDL-B model from [b-3GPP TR 38.901] with a nominal delay spread of 100 ns |
| TDLC60 | Based on the TDL-C model from [b-3GPP TR 38.901] with a delay spread of 60 ns |
| TDLC300 | Based on the TDL-C model from [b-3GPP TR 38.901] with a long delay spread of 300 ns |
| HST-SFN870 | HST (custom maximum Doppler shift): High-speed train channel emulation (Doppler shift: 870 Hz) |
| HST-SFN1667 | HST (custom maximum Doppler shift): High-speed train channel emulation (Doppler shift: 1 667 Hz) |
| HST-DPS870 | HST (custom maximum Doppler shift): High-speed train channel emulation (Doppler shift: 870 Hz) |
| HST-DPS1667 | HST (custom maximum Doppler shift): High-speed train channel emulation (Doppler shift: 1 667 Hz) |

### 10.2.4 Scheduling

#### 10.2.4.1 Symbols

Each slot contains 14 orthogonal frequency division modulation (OFDM) symbols. NR enables transmission to start at any OFDM symbol and to last only as many symbols as needed for communication. This type of mini-slot transmission can facilitate very low latency for critical data and minimize interference to other links, according to the "lean design" principle in NR that aims at reducing transmissions, leading to higher network energy efficiency and lower interference.

NOTE – When using an extended cyclic prefix, the number of symbols in a slot is 12. For more information, see clause 4.3.2 of [b-3GPP TS 38.201].

#### 10.2.4.2 Resource blocks

A resource block (RB) consists of 12 consecutive subcarriers in the frequency domain. It is possible to configure the number of RBs for UL and DL, and hence the number of supported sub-carriers.

#### 10.2.4.3 Frames and slots

In LTE, DL and UL transmissions are organized into radio frames with a 10 ms duration. Two radio frame structures are supported: Type 1, applicable to FDD; and Type 2, applicable to TDD. Transmissions in multiple cells can be aggregated where up to four secondary cells can be used in addition to the primary cell.

Frame structure type 1 applies to both full duplex and half duplex FDD. Each radio frame is 10 ms long and consists of 20 slots of length 0.5 ms, numbered from 0 to 19. A subframe is specified as two consecutive slots where subframe $i$ consists of slots $2i$ and $2i + 1$. For FDD, 10 subframes each are available for both DL and UL transmissions in each 10 ms interval. UL and DL transmissions are separated in the frequency domain. In half-duplex FDD operation, UE cannot transmit and receive simultaneously, while there are no such restrictions in full-duplex FDD.

Frame structure type 2 applies to TDD. Each radio frame of length 0 ms consists of two half-frames of length 5 ms each. Each half-frame consists of five subframes of length 1 ms. Each subframe $i$ is defined as two slots, $2i$ and $2i + 1$, of length 0.5 ms each. The supported UL-DL configurations are listed in Table 4.2-2 of [b-3GPP TS 36.211].

According to 3GPP specifications, there are seven different TDD frame configurations (denoted 0 to 6), determining seven different combinations of UL and DL sub-frame sets. Table 10-3 lists the different configurations, where D stands for DL, U for UL and S is a special subframe.

**Table 10-3 – Long-term evolution TDD frame configurations**

| Uplink-downlink configuration | Downlink-to-uplink Switch-point periodicity (ms) | Subframe number | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| 0 | 5 | D | S | U | U | U | D | S | U | U | U |
| 1 | 5 | D | S | U | U | D | D | S | U | U | D |
| 2 | 5 | D | S | U | D | D | D | S | U | D | D |
| 3 | 10 | D | S | U | U | U | D | D | D | D | D |
| 4 | 10 | D | S | U | U | D | D | D | D | D | D |
| 5 | 10 | D | S | U | D | D | D | D | D | D | D |
| 6 | 5 | D | S | U | U | U | D | S | U | U | D |

Similarly, in 5G NR, a frame has a duration of 10 ms and consists of 10 subframes. The number of slots in each frame and hence their time interval, though, depends on the numerology, which corresponds to SCS in the frequency domain:

- Numerology 0 (FR1) – SCS 15 kHz = 10 slots per frame
- Numerology 1 (FR1) – SCS 30 kHz = 20 slots per frame
- Numerology 2 (FR1 and FR2) – SCS 60 kHz = 40 slots per frame
- Numerology 3 (FR2) – SCS 120 kHz = 80 slots per frame

Each slot can be configured with a different configuration of UL and DL (or flexible) symbols. Slot formats are specified in Table 11.1.1-1 of [b-3GPP TS 38.213]. For TDD, symbols in a slot can be used for either UL or DL. For FDD, symbols can be used for UL, DL or flexible (which can be used for either UL or DL).

#### 10.2.4.4 5G new-radio scheduling configuration

The testbed shall support the following 5G scheduling configurations:

- basic scheduler;
- full throughput;
- physical downlink shared channel (PDSCH) reference measurement channel (RMC) (Rx tests, [b-3GPP TS 38.521-4]);
- physical uplink shared channel (PUSCH) RMC (Tx tests, [b-3GPP TS 38.521-4]);
- physical uplink control channel (PUCCH) aggregate power control RMC (Tx tests, [b-3GPP TS 38.521-1]);
- PUCCH error vector magnitude RMC (Tx tests, [b-3GPP TS 38.521-1]).

The 5G cell shall be able to configure the bandwidth part (Initial BWP + BWP 1 + BWP 2 + …). Also, the frame and slot configuration, DL settings such as fixed MCS index, antenna ports, security request system request, RB allocation start or count or RB allocation type/bitmap, and symbol start or length can be configured.

### 10.2.5 Core network

Depending on the target of the study, most of the effort of emulating a mobile network may lie on the RAN emulator rather than on the core network. As long as the DUT can be connected to a LAN, it may be sufficient for the experimenter to perform the study. This can be achieved by simplifying the core network into an abstract software orchestrator, only implementing the necessary functions that provide connectivity to the DUT.

Essential functions shall be comprised, such as authentication, IP address allocation, and routing between DUT and LAN. Given the scope of this Recommendation, managing the service policy and QoS should be included within the capabilities of the emulated core network, although not all studies might need it.

It is possible to introduce traffic impairments to broadly emulate the effect of packets passing through the core network, even when none is emulated. Typical traffic impairments comprise IP capacity limitation, IP packet loss, IP packet transfer delay, IP packet delay variation [ITU-T Y.1540], as well as other types of impairment.

#### 10.2.5.1 Long-term evolution

This clause aims to provide a brief description of the main components of the LTE evolved packet core (EPC), while more information can be found in [b-3GPP TS 23.401] and [b-3GPP TS 36.300]:

- Mobility management entity (MME): The MME is the entity in charge of managing security functions, tracking, paging, and mobility procedures and is also involved in bearer activation and deactivation procedures.
- Serving gateway (SGW): The SGW handles mobility by acting as an anchor during handover procedures between adjacent cells. It also retains the context of the DUT.
- Packet data network gateway (PGW): The PGW is the gateway towards external packet data networks, acting as a point of entry or exit for DUT traffic. It also acts as an anchor between 3GPP and non-3GPP technologies.
- Policy and charging rules function (PCRF): The PCRF is in charge of managing the service policy and sending QoS-setting information for each user session and accounting rule information.
- Home subscriber server (HSS): The HSS is the central database for all subscriber-related data in the network. It contains subscriber-specific data such as the international mobile subscriber identity, mobile station international integrated services digital network number, subscribed access point name and priority indication.

#### 10.2.5.2 New radio

This clause aims to provide a brief description of the main 5G core network functions in no particular order, while more information can be found in [b-3GPP TS 23.501].

- Access and mobility management function (AMF): The AMF is the 5G equivalent to the MME in LTE, and includes functionalities such as registration management, connection management, reachability management, mobility management, access authentication and authorization.
- Session management function (SMF): The SMF encompasses the C-Plane functionalities of the SGW or PGW in LTE. These include session management, IP address allocation, traffic steering, and DL data notification.
- User plane function (UPF): The UPF encompasses the user plane functionalities of the SGW or PGW in LTE. These include being the anchor point for intra- or inter-RAT mobility, packet routing and forwarding, policy rule enforcement, traffic usage reporting, and QoS handling for the user plane.

- Unified data management (UDM): The UDM, together with the authentication server function and unified data repository, is the equivalent of the HSS in LTE. It includes functionalities such as the generation of authentication credentials, user identification handling, access authorization based on subscription data (e.g., roaming restrictions) and subscription management.
- Policy control function (PCF): The PCF encompasses the PCRF functionalities in LTE, making the policy and charging control decisions fed to the AMF and SMF.
- Network slice selection function (NSSF): The NSSF helps the AMF to select the network slice instances that will serve a particular device.

#### 10.2.5.3 Mobility

The testbed may have the ability to emulate different mobility scenarios:

- intra-system handover;
- inter-system handover;
- roaming.

#### 10.2.5.4 Cell load

The testbed should have the ability to emulate different cell loads.

- The system should support allocating a specific number of physical resource blocks (PRBs) to simulate different load conditions of the cell.
- The system may support the simulation of virtual allocations of the non-assigned resources to virtual users to keep full DL occupancy.

The testbed should have the ability to generate application layer traffic with the following characteristics:

- protocol layers: IPv4 and IPv6;
- transport layers: user datagram protocol and transmission control protocol (TCP);
- single and multiple streams of TCP;
- differentiated services code point marking;
- application layer service data unit (SDU) size: constant, uniform, exponential, Pareto, Cauchy, normal, Poisson, gamma, Weibull;
- application layer SDU inter-departure time: constant, uniform, exponential, Pareto, Cauchy, normal, Poisson, gamma, Weibull;
- traffic bursts with the following statistical distributions for the on and off periods: constant, uniform, exponential, Pareto, Cauchy, normal, Poisson, gamma, Weibull.

#### 10.2.5.5 Dual connectivity

The testbed may be able to emulate different scenarios with dual connectivity (DC) between 3GPP radio access nodes and between 3GPP and non-3GPP radio access nodes. In particular:

- establish a DC connection with the UE, supporting DC C-plane;
- establish a dual-bearer connection with the SGW;
- being able to perform a bearer split within the master evolved node B (eNB);
- being able to perform at least LTE/Wi-Fi interworking via the access network discovery and selection function.

When considering 5G NR testbeds, DC assumes a more specific meaning. Evolved universal mobile telecommunication system terrestrial radio access network (E-UTRAN) new radio dual connectivity (EN-DC) provides a mechanism for EN-DC-enabled devices to send and receive data to and from

connected E-UTRAN and 5G NR cells. This enables network operators to leverage existing E-UTRAN infrastructure while adding 5G NR cells to increase capacity and performance.

The E-UTRAN cells (evolved nodes B (eNBs)) are grouped into master cell groups, and the 5G NR cells (grouped nodes B (gNBs)) are grouped into secondary cell groups. Master cells (eNBs) communicate with secondary cells (gNBs) over the X2 interface for U- and C-plane data. The eNBs and gNBs communicate with the EPC using the S1 interface, with all mobility management data from the EPC transmitted to the gNBs via the eNB ([b-3GPP TS 37.340]). For a depiction of the overall EN-DC architecture, see Figure 4.1.2-1 of [b-3GPP TS 37.340].

### 10.2.6 Network impairments

The testbed shall have the ability to emulate the following network impairments in the incoming and outgoing flows of the data and C-plane:

- IP packet transfer delay;
- IP packet transfer delay variation;
- IP packet error ratio;
- IP packet reordered ratio;
- IP packet duplicate ratio;
- IP packet loss;
- IP capacity limitation.

In addition, traffic management impairments such as policing and shaping might be integrated into the testbed to replicate real-life mobile operator networks. In fact, some rate plan bandwidth caps might be used to control the operator's subscriber base.

## 10.3 Service provider

A consistent data flow originating from a designated service provider is necessary for the AUT to function properly. For instance, in the context of the CDSS use case, this provider could be a video streaming server, while for the CG use case, it could be a gaming server. These service providers may reside internally within the testbed environment or externally, accessible via the Internet.

Maintaining precise control and ensuring the reproducibility of test scenarios is paramount throughout testing procedures. Therefore, prioritizing the utilization of internal service providers, such as locally deployed app servers, is generally advisable.

However, connecting to an external service provider may be preferred if the primary objective is to emulate real-world conditions and simulate how the AUT would operate under typical user circumstances. Nonetheless, this choice comes with the caveat that relinquishing control over data transmission introduces uncertainties, as the intricacies of Internet routing and external server operations cannot be regulated. Consequently, while realism may be enhanced, the trade-off compromises the control and reproducibility of test conditions.

## 10.4 Measurements

Measurements are obtained by specific probes, either software or hardware.

Software probes can take many forms, depending on what they are trying to measure. For example, probes can attach to application program interfaces (APIs) given by the AUT (if any exist), they can be integrated within the AUT (if the source code is available), or they can monitor the DUT (e.g., record network traffic, screen, audio interfaces or interactions).

If the AUT is closed source but APIs exist, the experimenter can try to tap into the information openly available. It is limited, though, by the type, amount, frequency and precision given by app developers. If this is enough or alternatives do not exist or are hard to implement, the API should be stable (e.g.,

APIs follow a rigorous versioning system, and a stable release is available), and details about it should be reported. This enables repeatable experiments, allowing third parties to reproduce the experiment and verify that the results are consistent.

If the AUT is open source, instead, the experimenter can decide to create custom hooks to log events of interest with detailed information. One possibility is to use Triangle's instrumentation library [b-TriangleResources]. With this library, developers can dump measurements in a known format, making it easy to parse and use them within the scope of this Recommendation. Examples of events of interest could be, e.g., video streaming demand request timestamp, video streaming play timestamp, send comment or message to social media timestamp, or acknowledgment from social media comment or message timestamp.

If the AUT is closed source and no APIs are available, other types of probes are needed. Some of them might rely on the DUT OS (e.g., screen recorder or audio interface recorder), some others can be commercial applications (e.g., network traffic sniffer, or touch or click recorder), or can be custom-made if needed. This is the most flexible but cumbersome method to get the necessary measurements from closed-source applications since many types of information can be retrieved at the cost of an additional interpretation layer. Furthermore, this information is often extracted raw and needs further processing and interpretation to be usable.

If software probes run on the DUT, it is essential to test how much computational resources they need (e.g., CPU or GPU usage, RAM usage, storage throughput, or network throughput). In addition, the experimenter must evaluate the resource usage with and without the active probes, understanding whether they affect the behaviour of the DUT and AUT. Further details are discussed in clause 9.4.

In addition to software probes, external hardware probes can be used to obtain specialized measurements. For example, a power analyser connected to the UE can measure power consumption, a hardware packet sniffer can record network traffic in greater detail and with fewer limitations than a software sniffer without impacting DUT performance, and a radio access emulator can provide internal logs about the protocol exchange and radio interface low layers metrics.

The testbed should be tailored based on the measurement that the experimenter requires, and thus needs to be as flexible as possible.

### 10.5 Test coordination and automation

This clause provides the features of the control tool that can carry out the coordination of the testbed. This is needed to make the different components of the testbed properly interact with each other.

One open-source option to automate operations is [b-Opentap], which provides flexible and extensible test sequences and test plan creation. [b-Opentap] has an extensible and cross-platform architecture built on the .net platform. From its wide range of sequencing functionality and infrastructure, it is also possible to develop plugins tailored to specific automation needs. [b-Opentap] is under a weak-copyleft licence, and plugins can be shared with the [b-Opentap] community. The software includes the core engine, a command line interface (CLI) for installing plugins, and a package manager for managing installed plugins.

Further details are presented in Appendix IV.

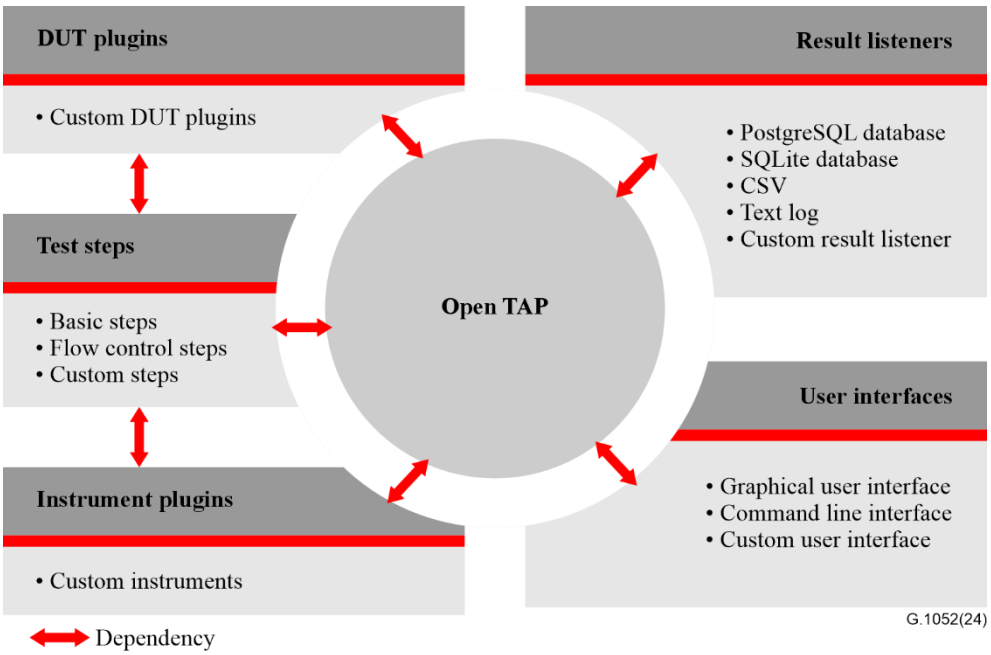



**Figure 10-4 – [b-Opentap] architecture and dependencies**

# Appendix I

## Relevant ITU-T Recommendations

(This appendix does not form an integral part of this Recommendation.)

This appendix contains a list of relevant Recommendations related to the use cases mentioned in clause 7.1. Table I.1 gives the reader useful pointers to relevant existing Recommendations and ongoing work items in ITU.

Relevant documents address different standardization objectives, specifically network requirements or planning, objective quality assessment, subjective quality assessment and influencing factors.

Table I.1 presents a snapshot in time of ITU-T Recommendations as of April 2024. Table I.1 might not be updated over time, and the reader is thus encouraged to look at more recent Recommendations on the topic, using the documents here as a starting point.

**Table I.1 – Relevant ITU-T Recommendations, 2024-04**

| Use case | Network requirements or planning | Objective quality assessment | Subjective quality assessment | Influencing factors |
|---|---|---|---|---|
| Augmented reality | – | – | – | [b-ITU-T G.1036], [b-ITU-T P.1320] |
| Cloud gaming | [b-ITU-T G.1072] | – | [b-ITU-T P.809] | [b-ITU-T G.1032] |
| Content distribution streaming | [b-ITU-T G.1071] | [b-ITU-T P.1203], [b-ITU-T P.1204], [b-ITU-T P.1211] | [b-ITU-T P.910], [b-ITU-T P.917], [b-ITU-R BT.500] | – |
| High-speed Internet | – | [ITU-T Y.1540], [b-ITU-T Y.1567] | – | – |
| Live streaming services | – | [b-ITU-T P.1201], [b-ITU-T P.1202], [b-ITU-T P.1203], [b-ITU-T P.1204] | – | – |
| Social networking | – | – | – | – |
| Telemeetings | – | [b-ITU-T P.1305] | [b-ITU-T P.920], [b-ITU-T P.1301], [b-ITU-T P.1312], [b-ITU-T P.1305] | [b-ITU-T P.1301], [b-ITU-T P.1305], [b-ITU-T P.1320] |
| Virtual reality | – | – | [b-ITU-T P.919], [b-ITU-T P.812] | [b-ITU-T G.1035], [b-ITU-T P.1320] |

# Appendix II

## Examples of key performance indicator normalization mapping functions

(This appendix does not form an integral part of this Recommendation.)

As explained in clause 6.2, KPIs can span very different values from one to another and even be of different types. For example, the number of content stalls during a video streaming can be any non-negative integer number, the time to load a reference picture on a social media can be any positive real number, a message sent either with or without success, content stalls durations over a streaming can be multiple, each one of any positive real number.

Before an objective model maps a set of KPIs to an estimated QoE, it can be helpful to normalize each KPI using a mapping function. The normalization can allow an understanding of how the different KPIs contribute to the estimated QoE and simplify its computation. For example, for the models recommended by ITU-T for video streaming, in [b-ITU-T P.1201], [b-ITU-T P.1202], [b-ITU-T P.1203] and [b-ITU-T P.1204], corresponding KPI-normalizations are included.

To make sense of each normalized KPI, they should be mapped into a bounded codomain. Possibilities include the sets of real numbers $[0,1]$ as the standard normalization, $[0,100]$ as a percent-scale normalization, $[1,5]$ as a typical five-point ACR scale normalization, $[m, M]$ in general, where $m < M \in \mathbb{R}$. Notice that these normalization functions can be easily transformed from one to another via linear or affine transformations.

NOTE – Normalization and mapping are basically two different processes. Normalization is lossless in the sense that input and output values are related unambiguously and that the process is fully reversible. Mapping, in contrast, can be lossy, e.g., when discretization or clipping is used, and therefore not reversible. Combining normalization and mapping may help to streamline data processing; however, these facts should be kept in mind.

To more easily compare how each KPI contributes to the SQS, all KPIs shall use the same range for the normalization. Furthermore, the use of the five-point ACR scale normalization for KPIs that are not directly derived by a human panel is discouraged. It is thus preferable to use other normalizations, such as the [0, 100] percent scale when KPIs do not come from human panels or when a mixture of KPIs coming or not from human panels is present.

To map unbounded domains into bounded codomains, the mapping function can be restricted using $\max_m(\cdot)$ for KPIs mapped towards $-\infty$, $\min_M(\cdot)$ for KPIs mapped towards $+\infty$, or using the clamp function $\cdot|_{[m,M]}$ for KPIs mapped into $[-\infty, +\infty]$. In general, from now on, the clamp function considering the chosen codomain $[m, M]$ is assumed for all mapping functions.

Once the domain is chosen, many mapping functions can be derived. Mapping functions describe how the original domain of a KPI is mapped into a scalar normalized KPI.

The notation in the remainder of the clause is as follows:

- $x$: real scalar KPI value, possibly unbounded, i.e., $x \in \mathbb{R}$
- $\vec{x}$: $n$-dimensional real vector KPI, with possible unbounded entries, i.e., $x \in \mathbb{R}^n$
- $v$: real scalar KPI value bounded in $[m, M]$
- $f$: the generic mapping function, such that

$$v = f(x)|_{[m,M]} \in [m, M], \text{ or } v = f(\vec{x})|_{[m,M]} \in [m, M]$$

Some mapping functions are derived by converting fixed points in the domain to fixed points in the codomain. While the fixed points in the codomain have been described $(m, M)$, fixed points of the domain are defined as follows:

- $k$: real value, such that $f(k) = m$

- $K$: real value, such that $f(K) = M$

Common mapping can be designed, e.g., linear, logarithmic, sigmoid, inverse and discrete functions. Furthermore, functions $f(\vec{x})$: $\mathbb{R}^n \longrightarrow [m, M]$ will have to operate on multiple dimensions in those cases where the target KPI is a vector. Furthermore, if data derived from human panels are available, machine learning algorithms can optimize the conversion from KPI to normalized KPI [b-ITU-T P.1402].

Normalization mapping functions can take infinitely many forms, necessarily tailored to a specific need of a given KPI. Examples of such functions can be linear, logarithmic, inverse, power law, exponential mapping, polynomial of D-degree, step functions or piecewise linear.

This appendix describes two of the simplest normalization mapping functions, specifically, linear and logarithmic mapping, specifying their analytical forms and providing examples and plots to aid their visualization.

## II.1 Linear mapping

With linear mapping, a function of the type:

$$f(x) = ax + b$$

is assumed. The parameters $a$ and $b$ can be recovered by mapping the fixed points from the domain to the codomain, i.e.,

$$\begin{cases} f(k) = ak + b = m \\ f(K) = aK + b = M \end{cases}$$

Obtaining:

$$f(x) = \frac{x-k}{K-k}(M - m) + m$$

This simplifies, for example, when the chosen codomain is $[0, M]$ as follows:

$$f(x) = \frac{x-k}{K-k}M$$

This further simplifies when the codomain is $[0,1]$ as:

$$f(x) = \frac{x-k}{K-k}$$

Furthermore, if the function is truly linear, i.e., $m = 0, k = 0$, and thus $f(0) = 0$, the following is obtained:

$$f(x) = \frac{x}{K}$$

When the chosen codomain is the typical five-point ACR scale, the mapping becomes:

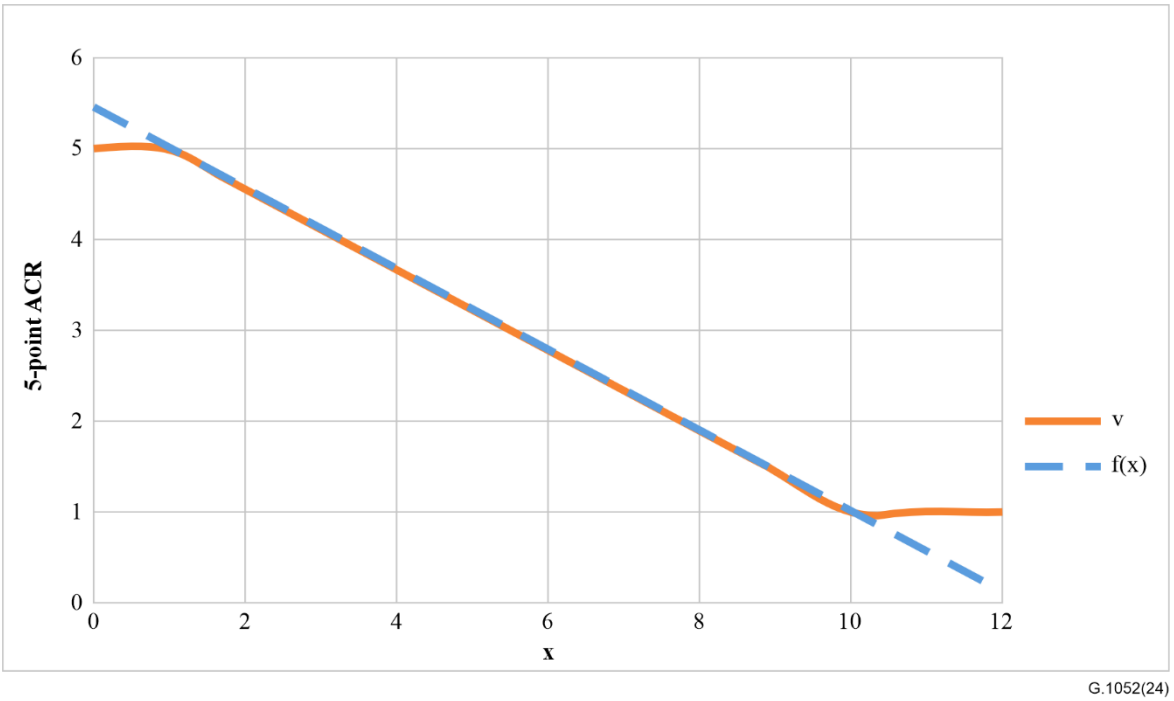
$$f(x) = \frac{x-k}{K-k}\ 4 + 1$$

As a simple example, if a KPI of the time to upload a picture on social media is under investigation, parameters $K = 1$ (i.e., if it takes 1 s or less, the mean opinion score (MOS) will be 5), and $k = 10$ (i.e., if it takes 10 s or more, the MOS will be 1) are chosen. With these parameters, the mapping function becomes:

$$f(x) = \frac{x - 10}{-9}\ 4 + 1 = -\frac{4}{9}x + \frac{49}{9}$$

Note that $k$ is the fixed point in the domain that maps into a MOS of 1, while $K$ is the fixed point in the domain that maps into a MOS of 5. In this case, $k > K$, which makes the function strictly monotonically decreasing. In the case of a monotonically increasing mapping function, the parameters follow the inequality $k < K$. An example could be the case of a KPI such as the number of times an action is taken before an error occurs.

Furthermore, note that for an instant picture upload time, i.e., $x = 0$, $f(0) = \frac{49}{9} = 5.44 > 5$, and for an extremely long upload time of, e.g., 20 s, $f(20) = -\frac{31}{9} = -3.44 < 1$. As explained, the final normalized KPI is the clamped version of the function, i.e., $v = f(x)|_{[m,M]} \in [m, M]$, turning $v = f(0)|_{[1,5]} = 5.44|_{[1,5]} = 5$, and $v = f(20)|_{[1,5]} = -3.44|_{[1,5]} = 1$. This is justified by the fact that the value $k$ is chosen such that the user is completely unsatisfied with the result, thus scoring the KPI to a value $m$ = 1, and any performance worse than that is still mapped to $m = 1$. Similarly, any performance better than $K$ is not perceived by the user due to, e.g., physiological reaction times, and thus any improvement is still scored as $M = 5$. See Figure II.1.




The five-point ACR scale is used, thus $m = 1$ and $M = 5$, and the mapping function is shown both with and without clamping. In this case, $k = 10$, and $K = 1$ are used.

**Figure II.1 – Example of the linear mapping function**

## II.2 Logarithmic mapping

Logarithmic mapping is inspired by the opinion model recommended by [b-ITU-T G.1030]. In the context of web page-loading times, it is shown that the human perception of time for such a use case is logarithmic.

The logarithmic mapping takes this form:

$$f_m(x) = a \log x + b = \log_c x + b = \log_c(dx)$$

Note that the base of the logarithm is not specified here since a change of base is implicitly obtained from the parameter $a$, and the explicit base is $c$.

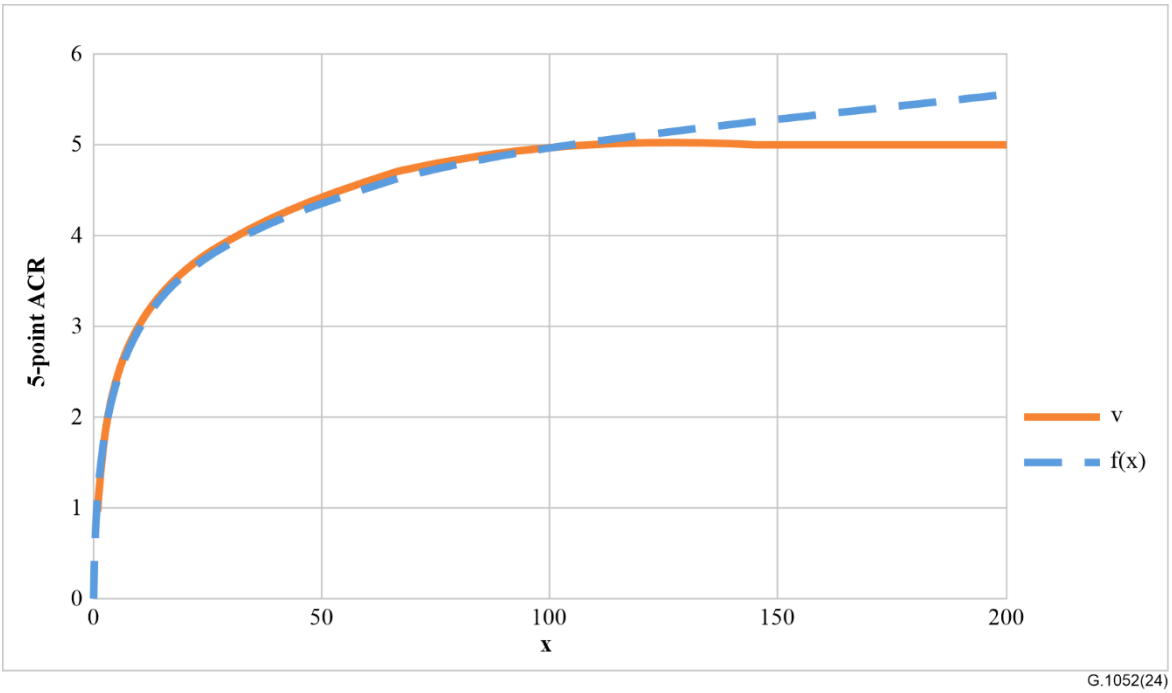
Similarly to what was done in clause II.1, the parameters can be obtained by setting, e.g.,

$$\begin{cases} f_m(k) = a \log k + b = m \\ f_m(K) = a \log K + b = M \end{cases}$$

giving

$$f_m(x) = \frac{\log x - \log k}{\log K - \log k}(M - m) + m$$

Some example values for relations between $k$ and $K$ based on human reaction times can be found in [b-ITU-T G.1030], [b-Nielsen]. Their values need updating according to the interpolation requirement if through subjective experimentation other values are found to be better adjustments for specific services. See Figure II.2.




The five-point ACR scale is used, thus $m$ = 1 and $M$ = 5, and the mapping function is shown both with and without clamping. In this case, $k = 1$, and $K = 100$, and $log_{10}(\cdot)$ is used.

**Figure II.2 – Example of the logarithmic mapping function**

With respect to linear mapping, logarithmic mapping is used when a KPI typically spans multiple orders of magnitude, and the score follows more closely the order of magnitude rather than the absolute value.

# Appendix III

## Scenario parameter settings

(This appendix does not form an integral part of this Recommendation.)

The parameters contained in this clause are based on section 3 of Appendix 8 of [b-TriangleD2.2]. These scenarios are just a suggestion. More scenarios or variations of the proposed scenarios can be used within the context of this Recommendation. Furthermore, network impairments (clause 10.2.6) can also be applied on top of the proposed scenarios, to emulate exceptionally poor network performance.

### III.1 Parameters explained

Brief explanations of the main parameters used in this appendix follow. For further information, see the original 3GPP specifications.

#### III.1.1 Long-term evolution parameters

##### III.1.1.1 Channel models

Channel model parameters can be found in [b-3GPP TS 36.101], [b-3GPP TS 36.104]. Delay profiles are specified in the form of a TDL, characterized by a number of taps at fixed positions on a sampling grid. The profile can be further characterized by the root mean squared DS and the maximum delay spanned by the taps. Three delay profiles are selected to be representative of low, medium, and high DS environments, namely:

- EPA;
- EVA;
- ETU.

Furthermore, HST and static channels models are used from [b-3GPP TS 36.104].

##### III.1.1.2 Scheduling

The number of PRBs and subframes used in DL and UL is based on the fraction of maximum throughput from Table 8.2.1.1 of [b-3GPP TS 36.104]. The values depend on the number of Tx or Rx antennas, cyclic prefix type, propagation conditions (channel models and Doppler frequency) and channel bandwidth. Considering the fraction of maximum throughput and the RMCs in [b-3GPP TR 37.901], an LTE scenario can be created to emulate a UE real condition.

#### III.1.2 5G new-radio parameters

##### III.1.2.1 Channel models

Channel model parameters can be found in [b-3GPP TS 38.101-4]. Specifically, channel models available are:

- Static propagation conditions
- TDL models, specifically, TDLA30, TDLB100, TDLC300, TDLD30
- HST scenarios, specifically, single tap, SFN and DPS channel profile.

##### III.1.2.2 Scheduling

5G/NR numerologies come in various forms (waveform configurations like subframe spacing are supported), and the radio frame structure varies slightly based on the numerology. However, regardless of numerology, one radio frame length and one subframe length are the same, respectively, 10 ms and 1 ms (see clause 4.3.1 of [b-3GPP TS 38.211]).

Within a subframe, the number of slots changes depending on the numerology, while each slot is made up of 14 OFDM symbols.

The number of slots inside a subframe changes for SCS configuration ($\mu$) value in Table 5.1-1 of [b-3GPP TS 38.300]. Each SCS configuration has associated a SCS ($\Delta f$) measured in kilohertz. Within the subframe, the duration of each DL slot ($d_{\text{slot}}$) and UL slot ($u_{\text{slot}}$) can be configured, allowing the schedule of the subframe as in LTE to be done.

The maximum transmission bandwidth configuration for each UE is specified in Table 5.3.2-1 [b-3GPP TS 38.101-1]. It is directly related to the number of RBs used. In 5G, one RB contains 12 sub-carriers in a frequency domain similar to LTE. In LTE, RB bandwidth is fixed at 180 kHz, but in NR, it is not fixed and depends on sub-carrier spacing configuration.

#### III.1.2.2.1 Downlink scheduling

For the DL, there are two different scheduling types. The first is known as dynamic scheduling, and the second as semi-persistent scheduling (SPS). Each PDSCH is planned using dynamic scheduling by downlink control information (DCI; DCI 1_0 or DCI 1_1). The radio resource control (RRC) message schedules PDSCH transfer via the SPS method.

- DCI format_1_1: Scheduling one or multiple PDSCH in one cell and triggering one-shot – hybrid automatic repeat request-acknowledgement (HARQ-ACK) codebook feedback (Table 7.3.1-1 of [b-3GPP TS 38.212]).
- PRB allocation type 1: In this type, the resource is allocated to one or more consecutive RBs. The resource allocation area is defined by two parameters $RB_{start}$ and a length in terms of contiguously allocated RBs $L_{\text{RBs}}$. Description of the DL resource allocation type 1 is described in clause 5.1.2.2.2 of [b-3GPP TS 38.214].
- $K_0$: Indicates the number of time slots between the physical downlink control channel (PDCCH)/DCI and DL data (PDSCH) transmission [b-3GPP TS 38.213], [b-3GPP TS 38.214].
- $K_1$: Indicates the number of time slots between PDSCH and HARQ ACK or negative ACK transmission [b-3GPP TS 38.213], [b-3GPP TS 38.214].
- $d_{slots}$: Duration of the DL slots in milliseconds.

#### III.1.2.2.2 Uplink scheduling

There are two methods of scheduling for UL, the same as for DL. The first is known as dynamic scheduling, and the second as configured scheduling (CS). Each PUSCH is planned using dynamic scheduling by DCI (DCI 0_0 or DCI 0_1). The RRC message schedules the PUSCH transfer via the CS mechanism.

- DCI format_0_1: Scheduling one or multiple PUSCH in one cell or indicating DL feedback information for configured grant PUSCH (configured grant-downlink feedback information) (Table 7.3.1-1 of [b-3GPP TS 38.212]).
- PRB allocation type 1: The RB assignment information indicates to a scheduled UE a set of contiguously allocated non-interleaved virtual RBs within the active bandwidth part of size specified in clause 6.1.2.2.2 of [b-3GPP TS 38.214].
- $K_2$: Indicates the number of time slots between PDCCH/DCI and UL data (PUSCH) transmission [b-3GPP TS 38.213], [b-3GPP TS 38.214].
- $u_{slots}$: Duration of the UL slots in milliseconds.

### III.1.3 Parameter translation between long-term evolution and new-radio scenarios

Most scenario parameters do not change between LTE and NR scenarios, specifically, LTE RSRP and NR synchronization signal-reference signal received power (SS-RSRP), AWGN, channel model maximum Doppler frequency, channel model MIMO correlation and additional network latency.

The received power (RSRP or SS-RSRP) by the DUT must take into account the RF loss compensation, as described in clause 9.

AWGN is assumed as noise model and it must be computed based on RF compensation (clause 9.3). All scenarios assume a constant AWGN power and it is typically computed to obtain a certain SNR.

Standardized channel models, instead, do change from LTE to NR. Models can be either static, specialized for HST, or vary based on DS based on the surrounding environment and the situation being modelled. The translation in Table III.1 should be used.

**Table III.1 – Channel model translation between long-term evolution and new radio**

| Channel model | LTE | NR |
|---|---|---|
| Static | Static | Static |
| Low DS | EPA | TDLA30 |
| Medium DS | EVA | TDLB100 |
| High DS | ETU | TDLC300 |
| HST | HST | HST-SFN |

Note that channel model maximum frequency and channel model MIMO correlation might not be available for all channel models.

Scheduling-related parameters are expressed as a percentage of the total resources allocated to the DUT both in the frequency and in the time domain. Although LTE and NR have different ways to control the amount of temporal and frequency resources allocated to a UE, both are capable of doing so.

### III.2 Scenarios

For each scenario, sub-scenarios are specified which can be concatenated together to emulate a complex temporal evolution.

#### III.2.1 Urban office

See Table III.2.

**Table III.2 – Urban office sub-scenarios parameters**

| **Sub-scenario** | | | **Default working conditions (A)** | **Daily server back-up/git clone/OS updates deployment (B)** | **After lunch break (C)** | **Videoconference calls over 5G (D)** | **Late night /weekend crunch (E)** |
|---|---|---|---|---|---|---|---|
| Serving cell | RSRP | | −90 dBm | | | | |
| | AWGN | | Constant for average SNR = 25 dB | | | | |
| | Channel model | | Low DS | | | | |
| | Max Doppler | | 5 Hz | | | | |
| | Correlation | | Medium | | | | |
| Scheduling | DL | Freq. resources | 50% | 30% | 40% | 20% | 50% |
| | | Time resources | 30% | 30% | 30% | 20% | 100% |
| | UL | Freq. resources | 50% | 30% | 40% | 20% | 50% |
| | | Time resources | 30% | 30% | 30% | 20% | 100% |
| Additional network latency (each way) | | | 0 ms | 150 ms | Uniform [0 ms, 50 ms] | 100 ms | 0ms |

For a typical work day, sub-scenarios could be concatenated as follows: ABAACDDAEE, with a duration of 30 s each.

### III.2.2 Urban pedestrian

See Table III.3.

**Table III.3 – Urban pedestrian sub-scenarios parameters**

| Sub-scenario | | | Centre to park walk (A) | Busy green area – park (B) | Park to centre walk (C) | City main square (D) |
|---|---|---|---|---|---|---|
| Serving cell | RSRP | | From −90 dBm to −105 dBm, linear decrease with 4 s per 1 dB step | −105 dBm | From −105 dBm to −90 dBm, linear increase with 4 s per 1 dB step | −90 dBm |
| | AWGN | | Constant AWGN to obtain SNR starting from 25 dB and decreasing to 10 dB | 10 dB | Constant AWGN to obtain SNR starting from 10 dB an increasing to 25 dB | 25 dB |
| | Channel model | | Low DS | | | |
| | Max Doppler | | 5 Hz | | | |
| | Correlation | | Low (RSRP ≥ −98 dBm), Medium (RSRP < −98 dBm) | Medium | Low (RSRP ≥ −98 dBm), Medium (RSRP < −98 dBm) | Low |
| Scheduling | DL | Freq. resources | 50% | 40% | 50% | 30% |
| | | Time resources | 30% | 30% | 30% | 30% |
| | UL | Freq. resources | 50% | 40% | 50% | 30% |
| | | Time resources | 30% | 30% | 30% | 30% |
| Additional network latency (each way) | | | From 0 ms to 45 ms, increase by 3 ms at each RSRP change | 45 ms | From 45 ms to 0 ms, decrease by 3 ms at each RSRP change | 0 ms |

For a typical walk in the city, sub-scenario could be concatenated as follows: ABCD. The duration of A and C is of 64 s each, while B and D are run over 30 s each.

### III.2.3 Urban driving

See Table III.4.

**Table III.4 – Urban driving normal and emergency sub-scenarios parameters**

| Sub-scenario | | | Normal driving in light traffic (A) | Normal driving in heavy traffic (B) | Emergency driving in light traffic (C) | Emergency driving in heavy traffic (D) |
|---|---|---|---|---|---|---|
| Serving cell | RSRP | | From −90 dBm to −120 dBm, linear decrease 1.5 dBm per second | From −90 dBm to −120 dBm, linear decrease 1 dBm per second | From −90 dBm to −120 dBm, linear decrease 3 dBm per second | |
| | AWGN | | Constant AWGN to obtain SNR starting from 25 dB | | | |
| | Channel model | | Medium DS | | | |
| | Max Doppler | | 75 Hz | 50 Hz | 150 Hz | |
| | Correlation | | Medium | | | |
| Scheduling | DL | Freq. resources | 100% | 50% | 100% | 50% |
| | | Time resources | 50% | 30% | 50% | 30% |
| | UL | Freq. resources | 100% | 50% | 100% | 50% |
| | | Time resources | 50% | 30% | 50% | 30% |
| Additional network latency (each way) | | | From 0 ms, with a 3 ms increase per RSRP step | From 50 ms, with a 3 ms increase per RSRP step | From 0 ms, with a 5 ms increase per RSRP step | From 50 ms, with a 5 ms increase per RSRP step |

The scenario duration is equal to 21 s for sub-scenario A, 31 s for sub-scenario B, and 11 s for sub-scenarios C and D. All sub-scenarios can be time-reversed and concatenated, emulating, for example, driving first towards the base station and then away from it, thus having a linearly increasing ramp of RSRP followed by a symmetrical linearly decreasing ramp in RSRP.

If of interest and having the capability to do so, multiple cells can be emulated with reversed scenarios run at the same time. This would emulate a DUT driving away from a cell and towards the next one. At some point, a handover will occur, testing the robustness of the app on this type of event. See Table III.5.

**Table III.5 – Urban driving traffic jam sub-scenarios parameters**

| Sub-scenario | | | Ordinary traffic jam conditions (A) | Multiple emergency response vehicles drive fast in next lane (B) |
|---|---|---|---|---|
| Serving cell | RSRP | | −100 dBm | |
| | AWGN | | Constant AWGN to obtain SNR = 10 dB | |
| | Channel model | | Medium DS | |
| | Max Doppler | | 25 Hz | 150 Hz |
| | Correlation | | Medium | |
| Scheduling | DL | Freq. resources | 20% | |
| | | Time resources | 20% | |
| | UL | Freq. resources | 20% | |
| | | Time resources | 20% | |
| Additional network latency (each way) | | | 75 ms | |

For a typical flow, sub-scenarios could be concatenated as follows: ABA. The duration of A is 30 s, the duration of B is 10 s.

### III.2.4 Urban Internet café

See Tables III.6 and III.7.

**Table III.6 – Urban Internet café during busy hours sub-scenarios**

| Sub-scenario | | | Average coffee shop (A) | Heavy car traffic in front (B) | User spills coffee on device (C) |
|---|---|---|---|---|---|
| Serving cell | RSRP | | −95 dBm | −105 dBm | −95 dBm |
| | AWGN | | 20 dB | 10 dB | 20 dB |
| | Channel model | | Low DS | | |
| | Max Doppler | | 5 Hz | 20 Hz | 5 Hz |
| | Correlation | | Low | | Medium |
| Scheduling | DL | Freq. resources | 20% | | |
| | | Time resources | 20% | | |
| | UL | Freq. resources | 20% | | |
| | | Time resources | 20% | | |
| Additional network latency (each way) | | | Uniform [0 ms, 50 ms] | | |

**Table III.7 – Urban Internet café during off-peak hours sub-scenarios**

| Sub-scenario | | | Average coffee shop (A) | Heavy car traffic in front (B) | User spills coffee on device (C) |
|---|---|---|---|---|---|
| Serving cell | RSRP | | −95 dBm | −105 dBm | −95 dBm |
| | AWGN | | 20 dB | 10 dB | 20 dB |
| | Channel model | | Low DS | | |
| | Max Doppler | | 5 Hz | 20 Hz | 5 Hz |
| | Correlation | | Low | | Medium |
| Scheduling | DL | Freq. resources | 100% | | |
| | | Time resources | 20% | | |
| | UL | Freq. resources | 100% | | |
| | | Time resources | 20% | | |
| Additional network latency (each way) | | | 0 ms | Uniform [0 ms, 50 ms] | |

Typical flow for both cases is to concatenate sub-scenarios as follows: AABABAABCB. Each sub-scenario lasts 30 s for a total duration of 5 min.

### III.2.5 Suburban

See Tables III.8 to III.11.

**Table III.8 – Suburban festival sub-scenarios parameters**

| Sub-scenario | | | Entrance tickets check (A) | Popular band playing (B) |
|---|---|---|---|---|
| Serving cell | RSRP | | −80 dBm | |
| | AWGN | | 30 dB | |
| | Channel model | | High DS | |
| | Max Doppler | | 5 Hz | |
| | Correlation | | Medium | |
| Scheduling | DL | Freq. resources | 10% | 20% |
| | | Time resources | 10% | 20% |
| | UL | Freq. resources | 20% | 10% |
| | | Time resources | 20% | 10% |
| Additional network latency (each way) | | | 150 ms | |

**Table III.9 – Suburban stadium sub-scenarios parameters**

| Sub-scenario | | | Entrance tickets check (A) | Sports game ongoing (B) |
|---|---|---|---|---|
| Serving cell | RSRP | | −95 dBm | |
| | AWGN | | 15 dB | |
| | Channel model | | High DS | |
| | Max Doppler | | 5 Hz | |
| | Correlation | | Medium | |
| Scheduling | DL | Freq. resources | 10% | 30% |
| | | Time resources | 10% | 30% |
| | UL | Freq. resources | 20% | 30% |
| | | Time resources | 20% | 30% |
| Additional network latency (each way) | | | 100 ms | 0 ms |

Typical flow for both festival and stadium scenarios would be to concatenate sub-scenarios as follows: AB. Sub-scenario A lasts 10 s, while B lasts 50 s, for a total of 1 min.

**Table III.10 – Suburban shopping mall during busy hours sub-scenarios parameters**

| Sub-scenario | | | Ordinary busy mall (A) | Online wallet payment (B) |
|---|---|---|---|---|
| Serving cell | RSRP | | −100 dBm | |
| | AWGN | | 5 dB | |
| | Channel model | | High DS | |
| | Max Doppler | | 5 Hz | |
| | Correlation | | High | |
| Scheduling | DL | Freq. resources | 20% | 10% |
| | | Time resources | 20% | 10% |
| | UL | Freq. resources | 20% | 10% |
| | | Time resources | 20% | 10% |
| Additional network latency (each way) | | | 100 ms | |

**Table III.11 – Suburban shopping mall during off-peak hours sub-scenarios parameters**

| Sub-scenario | | | Ordinary empty mall (A) | Multiple delivering trucks around the mall (B) |
|---|---|---|---|---|
| Serving cell | RSRP | | −100 dBm | |
| | AWGN | | 5 dB | |
| | Channel model | | High DS | |
| | Max Doppler | | 5 Hz | 70 Hz |
| | Correlation | | High | |
| Scheduling | DL | Freq. resources | 100% | 50% |
| | | Time resources | 20% | 20% |
| | UL | Freq. resources | 100% | 50% |
| | | Time resources | 20% | 20% |
| Additional network latency (each way) | | | 0 ms | |

Typical workflow for both cases would be to concatenate sub-scenarios as follows: ABA. All sub-scenarios last 30 s.

### III.2.6 High speed train

See Tables III.12 and III.13.

**Table III.12 – High speed train with direct passenger connection sub-scenarios parameters**

| Sub-scenario | | | High speed train in suburban area (A) | Max speed train in empty fields (B) |
|---|---|---|---|---|
| Serving cell | RSRP | | From −90 dBm to −120 dBm, with 3 dB drop per 1 s | From −90 dBm to −120 dBm, with 5 dB drop per 1 s |
| | AWGN | | Constant AWGN to obtain SNR starting from 25 dB | |
| | Channel model | | HST | |
| | Max Doppler | | 870 Hz | 1667 Hz |
| Scheduling | DL | Freq. resources | 20% | |
| | | Time resources | 20% | |
| | UL | Freq. resources | 20% | |
| | | Time resources | 20% | |
| Additional network latency (each way) | | | 50 ms | 75 ms |
| Max HARQ retx | | | 2 | |

The scenario duration is equal to 11 s for sub-scenario A and 7 s for sub-scenario B. Both sub-scenarios can be time reversed and concatenated, emulating, for example, travelling first towards the base station and then away from it, thus having a linearly increasing ramp of RSRP followed by a symmetrical linearly decreasing ramp in RSRP.

If of interest and having the capability to do so, multiple cells can be emulated with reversed scenarios run at the same time. This would emulate a DUT travelling away from a cell and towards the next one. At some point, a handover will occur, testing the robustness of the app on this type of event. Multiple such scenarios can be concatenated to emulate a long trip.

**Table III.13 – High speed train with relay passenger connection sub-scenarios parameters**

| **Sub-scenario** | | | **High speed train in suburban area (A)** | **Max speed train in empty fields (B)** |
|---|---|---|---|---|
| Serving cell | RSRP | | −90 dBm | |
| | AWGN | | Constant AWGN to obtain SNR=25 dB | |
| | Channel model | | Static | |
| Scheduling | DL | Freq. resources | 20% | |
| | | Time resources | 20% | |
| | UL | Freq. resources | 20% | |
| | | Time resources | 20% | |
| Additional network latency (each way) | | | 100 ms | 150 ms |
| Jitter | | | 50 ms | |
| Packet loss | | | 1% | 3% |

Typical flow would be to alternate scenarios 30 s at a time.

# Appendix IV

## Opentap overview

(This appendix does not form an integral part of this Recommendation.)

Opentap is an open-source project for fast and easy development, deployment, and execution of automated tests.

In the remainder of this appendix, the main concepts from the Opentap documentation are reported and explained [b-Opentap-doc]. For updated and more detailed information, the reader is encouraged to visit the tool's webpage.

Opentap consists of multiple tools, including:

- Opentap – core engine
- CLI – command line interface for installed plugins, and
- Package Manager – a tool to manage installed plugins.

This overview is dedicated to Opentap itself describing its main components, namely, test plans, test steps, verdicts, resources, result listeners, session logs, packages and plugins.

### IV.1 Test plans

A test plan is a sequence of test steps and their associated data. They are stored as XML files, and use the ".TapPlan" file extension. Test plans are created with an editor. They can be executed either in an editor, or by using the `tap run` CLI action. A test step can be a child step, a parent step or both. Therefore, a test plan is a recursive structure of test steps. This hierarchy is illustrated in Figure IV.1.

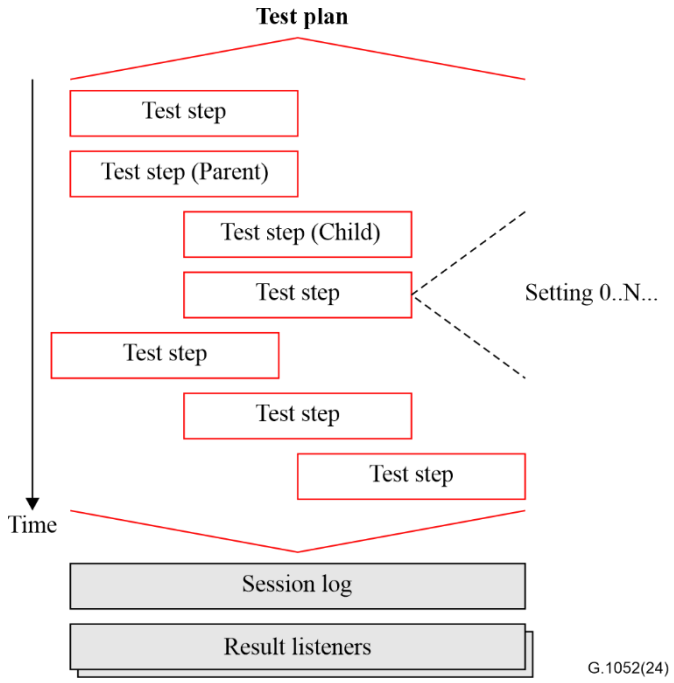



**Figure IV.1 – Opentap test steps hierarchy**

The step sequence is ordered from top to bottom, and child steps are indented to indicate that they belong to a parent step.

The execution order of child steps is decided by the parent step. For example, two typical parent steps are *Parallel* and *Sequential*, shown in Figure IV.2. *MyStep* is given as an example to illustrate that the topology of step execution can be controlled entirely by a parent step. Notice that the behaviour of MyStep can be achieved by putting a *Parallel* step in a *Sequence* step.

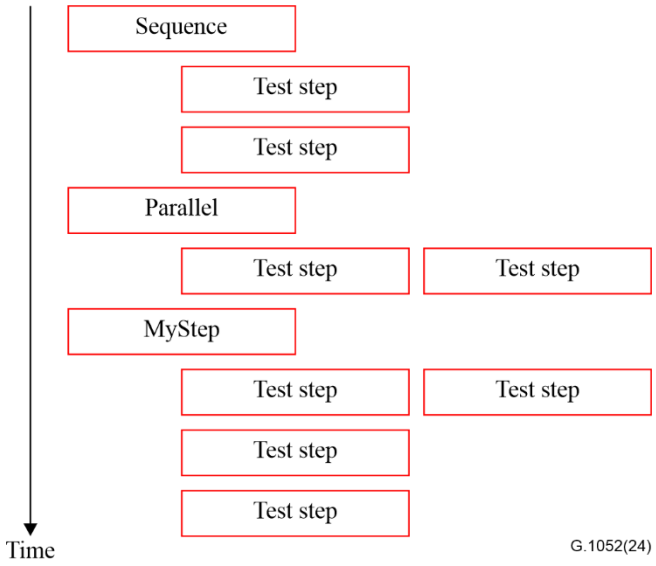



**Figure IV.2 – Opentap parallel and sequential steps**

## IV.2 Test steps

A test step is an element that encapsulates some piece of functionality. It should perform a single *step* of the test being run. The definition given is intentionally vague, as a step can perform a myriad of actions. It could make a measurement using an instrument or control a piece of hardware such as adjusting fan speed or voltage. It could also pause test execution, open a dialogue window, make a web request, run a different program or control the execution of other steps.

The aforementioned *associated data* of test steps refers to *step settings* and *resources*. Step settings affect the behaviour of a step. All steps have three *common* settings:

- *Enabled* enables or disables the step. This is useful for temporarily disabling certain steps without making destructive changes to a plan.
- *Step Name* sets the name of the step. This name will be reflected in the test plan as well.
- *Break Conditions* allows a step to override the break behaviour specified in the engine settings.

In addition to these common settings, a step may provide its own settings. These vary widely in functionality. For a *dialogue step*, which opens a dialogue box, the message in the dialogue box is controlled by a step setting. The buttons in the dialogue box (Yes/No/Cancel), and what happens when clicking those buttons, is also controlled by step settings.

On the other hand, *Session Log* and *Result Listeners* are situated outside of the test plan because they interpret the output of the test, and do not influence it directly. Ensuring this decoupling between generation and interpretation of results guarantees that, say, a new database can be added to store results without making any changes to the test steps, or their execution.

## IV.3 Verdicts

Every test step has a verdict. However, this verdict is not a part of the actual test plan. Rather, it is set during the execution of the test plan.

*NotSet* is the default verdict for all test steps. A step can set a verdict during execution to indicate success. A test plan outputs a verdict according to the verdicts of its steps. The verdict of a test plan is the most *severe* verdict of its child steps. A verdict has one of six severities, detailed in Table IV.1.

**Table IV.1 – Opentap verdicts**

| Severity | Verdict | Description |
|---|---|---|
| 1 | NotSet | No verdict was set |
| 2 | Pass | Step or plan passed |
| 3 | Inconclusive | Insufficient information to make a decision either way |
| 4 | Fail | Step or plan failed |
| 5 | Aborted | User aborted test plan |
| 6 | Error | An error occurred. Check session logs for more information |

This means that, for a test plan to output a *Pass* verdict, at least one step must set their verdict to Pass, and the rest must either set their verdict to *Pass* or *NotSet* verdict. This is also the most typical behaviour for parent steps containing child steps; a *Sequential* step passes if all of its children pass. However, this is not a rule. Parent steps decide their own verdict conditions.

*Break Conditions* operate based on step verdicts. The default *engine setting* is to abort the test plan if an *Error* verdict is encountered and continue otherwise. The following break conditions can be enabled and disabled individually:

- *Break On Error*
- *Break On Fail*
- *Break On Inconclusive*

Specifying break conditions overrides the engine settings in the given step, and all child steps of that step. Child steps can further override the value set by their ancestor if necessary. When a break condition is encountered, execution of all child steps is stopped, and the test plan resumes from the point where the break was encountered.

Figure IV.3 illustrates how break conditions can change test execution, as well as how verdicts are propagated from child steps to parent steps:

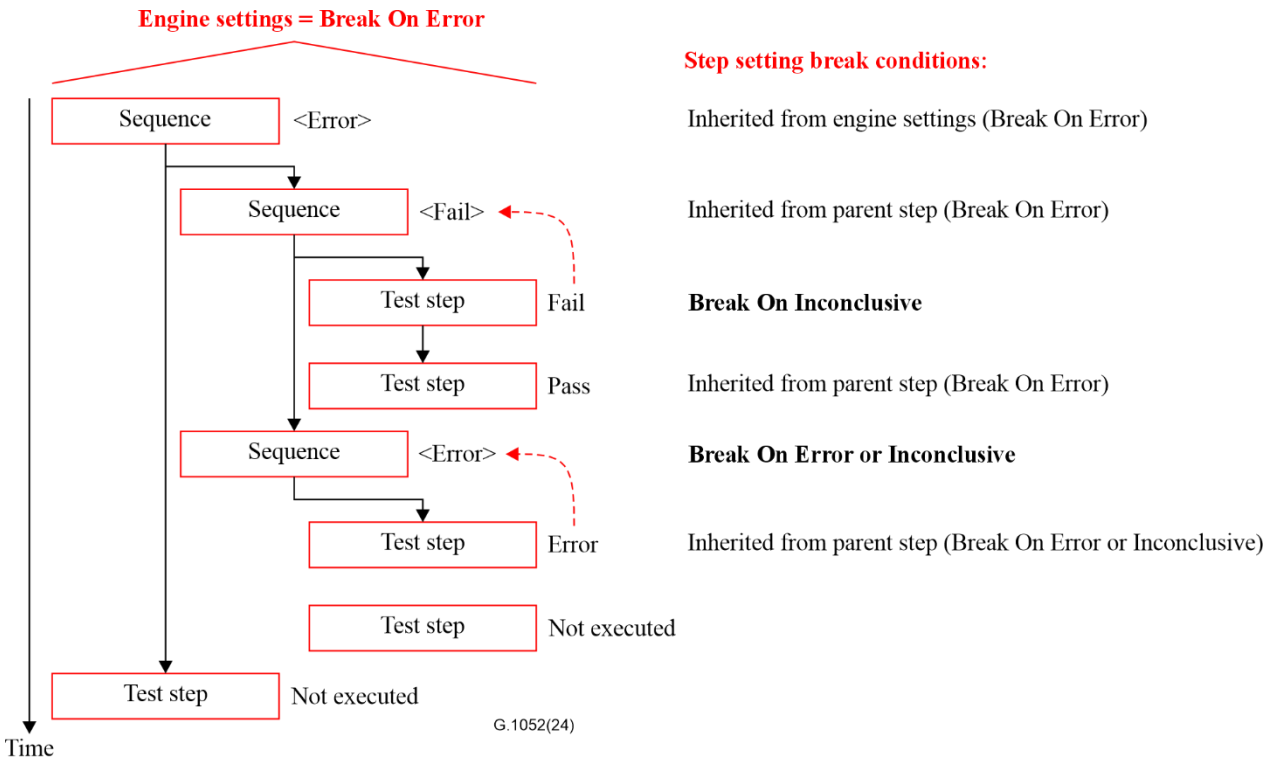



**Figure IV.3 – Opentap break conditions**

Steps are run in the sequence they appear from top to bottom, and verdicts from child steps are propagated upwards to their immediate parent. In Figure IV.3, the break condition on each step is specified on the right along with the source. **Bold text** means the break condition is explicitly set for that specific step, which is then inherited by child steps. Due to the *Break On Error* break condition, the last two steps are never executed in this example.

## IV.4 Resources

Opentap is intended for software as well as hardware testing. The concept of Instruments and DUTs are essential for Opentap. In the classical case, a DUT also requires calibration or control, and an instrument is anything that makes measurements. To integrate resources into test plans, whether to control them or read their measurements, they must be connected to test steps that know how to communicate with them. In other words, a resource driver is required. Creation of such a driver is described in the *Developer guide* [b-Opentap-dev].

Opentap is quite flexible regarding resources. Typically, they would be local, physical equipment. But they can easily be more abstract, such as a virtual resource, or even a remote resource. Depending on your use case, all of the following scenarios are valid:

- having no DUTs or instruments;
- using a single device as a DUT and an instrument simultaneously;
- using software resources as DUTs or instruments;
- using many DUTs and instruments.

Out of the box, Opentap does not provide any resources for hardware control. For that, you need plugins.

## IV.5 Result listeners

Result listeners are notified whenever a test step generates log output or publishes results.

Opentap ships with a simple result listener called *Text Log*, which saves log information from a test plan run to a file. Its behaviour can be modified via settings, including changing the file name and location of the log, as well as filtering the log information. There are four message categories: errors; warnings; information; and debug, each of which can be checked and unchecked.

Analogous to *Text Log*, there is a result listener plugin named *CSV*. Rather than logging information, this plugin saves all results published by test steps to a CSV file, which can also be changed via settings.

There are also PostgreSQL and SQLite result listeners that store all results and logs from a test plan run, as well as a copy of the test plan itself, making it possible to restore and run and old version of a test plan. The PostgreSQL plugin can be configured to use either a local database or publish to a remote database.

There is no limit to the number of active result listeners, and you can have multiple result listeners of the same type. For example, you could create two different *Text Log* listeners collecting debug and information log messages, respectively. There are more result listeners available in the package repository, and developers can make new result listener plugins to suit their needs, such as integrating with a different database technology, or uploading test results to a website.

Result listeners are tied to an Opentap installation, and not a test plan. Result listener settings are stored in `<install dir>/Settings/Results.xml`. Like test plans, it is not recommended to edit these by hand. However, they can be generated and edited using the test plan editors.

### IV.6 Session logs

Opentap keeps extensive session logs for debugging purposes. Session logs are kept from the 20 latest Opentap instances launched and are created independently of result listeners. The maximum amount of storage space used for session log files is 2 GB, with the largest size of a single file being 100 MB. When this limit is reached a new log file is created. There is a maximum number of 20 files kept at any given time. If the number of session log files exceeds 20 or the total file size surpasses 2 GB, log files will be removed starting with the oldest first.

The content of the session log files is similar to the logs collected by *Text Log*, but these also contain diagnostic messages from Opentap and any plugins you may have installed. They can be found in `<install dir>/SessionLogs`, and are named after the time and date at which they were created.

The session logs contain the same information you would see in your terminal when running Opentap with the `--verbose` flag. If you encounter errors, the logs may contain useful information for discovering what went wrong. If you think you discovered a bug in OpenTAP, please file an issue on GitHub. If applicable, please include instructions on how to replicate the issue, as well as relevant logs.

### IV.7 Packages and plugins

Opentap can be extended with plugins. A plugin provides a single piece of functionality. For example, a plugin could provide a new test step, or a new result listener. Plugins come in packages. A package is a versioned collection of plugins. Plugins vary widely in the functionality they provide.

Some examples are:

- graphical user interface editors for creating and running test plans;
- software development kit plugins to aid in developing and debugging plugins;
- tools for analysing test plans in real time to discover performance bottlenecks (timing analyser);
- representational state transfer interface to Opentap to allow you to control it remotely (REST-API).

Installing, uninstalling, upgrading, downgrading and dependencies are all managed by the Opentap package manager. A package declares its own dependencies and OS or CPU compatibility, and the package manager resolves those dependencies with the packages in your current installation, if possible.

Figure IV.4 shows the relation between plugins and Opentap. Opentap is at the centre, and plugins, providing a variety of functionality, can be added and removed painlessly.

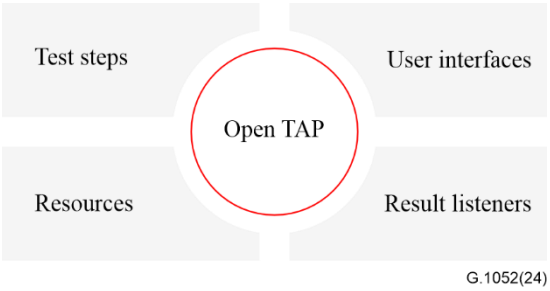



**Figure IV.4 – Opentap plugins**

# Bibliography


[b-ITU-T G.1030] Recommendation ITU-T G.1030 (2014), *Estimating end-to-end performance in IP networks for data applications*.

[b-ITU-T G.1032] Recommendation ITU-T G.1032 (2017), *Influence factors on gaming quality of experience*.

[b-ITU-T G.1035] Recommendation ITU-T G.1035 (2021), *Influencing factors on quality of experience for virtual reality services*.

[b-ITU-T G.1036] Recommendation ITU-T G.1036 (2022), *Quality of experience influencing factors for augmented reality services*.

[b-ITU-T G.1071] Recommendation ITU-T G.1071 (2016), *Opinion model for network planning of video and audio streaming applications*.

[b-ITU-T G.1072] Recommendation ITU-T G.1072 (2020), *Opinion model predicting gaming quality of experience for cloud gaming services*.

[b-ITU-T P.809] Recommendation ITU-T P.809 (2018), *Subjective evaluation methods for gaming quality*.

[b-ITU-T P.812] Recommendation ITU-T P.812 (2024), *Principles of subjective test methods for interactive virtual reality (VR) applications*.

[b-ITU-T P.910] Recommendation ITU-T P.910 (2023), *Subjective video quality assessment methods for multimedia applications*.

[b-ITU-T P.917] Recommendation ITU-T P.917 (2019), *Subjective test methodology for assessing impact of initial loading delay on quality of experience*.

[b-ITU-T P.919] Recommendation ITU-T P.919 (2020), *Subjective test methodologies for 360º video on head-mounted displays*.

[b-ITU-T P.920] Recommendation ITU-T P.920 (2000), *Interactive test methods for audiovisual communications*.

[b-ITU-T P.1201] Recommendation ITU-T P.1201 (2012), *Parametric non-intrusive assessment of audiovisual media streaming quality*.

[b-ITU-T P.1202] Recommendation ITU-T P.1202 (2012), *Parametric non-intrusive bitstream assessment of video media streaming quality*.

[b-ITU-T P.1203] Recommendation ITU-T P.1203 (2017), *Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport*.

[b-ITU-T P.1204] Recommendation ITU-T P.1204 (2023), *Video quality assessment of streaming services over reliable transport for resolutions up to 4K*.

[b-ITU-T P.1211] Recommendation ITU-T P.1211 (2023), *Derivation procedure for contribution values to quality degradation of adaptive audiovisual streaming services*.

[b-ITU-T P.1301] Recommendation ITU-T P.1301 (2017), *Subjective quality evaluation of audio and audiovisual multiparty telemeetings*.

[b-ITU-T P.1305] Recommendation ITU-T P.1305 (2016), *Effect of delays on telemeeting quality*.

[b-ITU-T P.1312] Recommendation ITU-T P.1312 (2016), *Method for the measurement of the communication effectiveness of multiparty telemeetings using task performance*.

[b-ITU-T P.1320] Recommendation ITU-T P.1320 (2022), *Quality of experience assessment of extended reality meetings*.

[b-ITU-T P.1401] Recommendation ITU-T P.1401 (2020), *Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models*.

[b-ITU-T P.1402] Recommendation ITU-T P.1402 (2022), *Guidance for the development of machine-learning-based solutions for QoS/QoE prediction and network performance management in telecommunication scenarios*.

[b-ITU-T Y.1567] Recommendation ITU-T Y.1567 (2023), *Latency under load metrics and methods of measurement*.

[b-ITU-R BT.500] Recommendation ITU-R BT.500 (2023), *Methodologies for the subjective assessment of the quality of television images*.

[b-3GPP TR 37.901] 3GPP TR 37.901 V15.1.0 (2018), *User equipment (UE) application layer data throughput performance (version 15)*.

[b-3GPP TR 38.901] 3GPP TR 38.901 V18.0.0 (2024), *5G; Study on channel model for frequencies from 0.5 to 100 GHz (Release 18)*.

[b-3GPP TS 23.401] 3GPP TS 23.401 V18.6.0 (2024), *General packet radio service (GPRS) enhancements for evolved universal terrestrial radio access network (E-UTRAN) access (Release 18)*.

[b-3GPP TS 23.501] 3GPP TS 23.501 V19.0.0 (2024), *System architecture for the 5G System (5GS); Stage 2 (Release 19)*.

[b-3GPP TS 36.101] 3GPP TS 36.101 V18.6.0 (2024), *Evolved universal terrestrial radio access (E-UTRA); User equipment (UE) radio transmission and reception (Release 18)*.

[b-3GPP TS 36.104] 3GPP TS 36.104 V18.5.0 (2024), *Evolved universal terrestrial radio access (E-UTRA); Base station (BS) radio transmission and reception (Release 18)*.

[b-3GPP TS 36.211] 3GPP TS 36.211 V18.0.1 (2024), *Evolved universal terrestrial radio access (E-UTRA); Physical channels and modulation (Release 18)*.

[b-3GPP TS 36.300] 3GPP TS 36.300 V18.2.0 (2024), *Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); Overall description; Stage 2 (Release 18)*.

[b-3GPP TS 37.340] 3GPP TS 37.340 V18.1.0 (2024), *NR; Multi-connectivity; Overall description; Stage-2 (Release 18)*.

[b-3GPP TR 37.901-5] 3GPP TS 37.901-5 V16.12.0 (2024), *5G; Study on 5G NR user equipment (UE) application layer data throughput performance /Release 16)*.

[b-3GPP TS 38.101-1] 3GPP TS 38.101-1 V18.6.0 (2024), *NR; User equipment (UE) radio transmission and reception; Part 1: Range 1 standalone (Release18)*.

[b-3GPP TS 38.101-4] 3GPP TS 38.101-4 V18.4.0 (2024), *5G; NR; User equipment (UE) radio transmission and reception; Part 4: Performance requirements (Release18)*.

[b-3GPP TS 38.201] 3GPP TS 38.201 V18.0.0 (2023), *NR; Physical layer; General description (Release18)*.

[b-3GPP TS 38.211] 3GPP TS 38.211 V18.3.0 (2024), *NR; Physical channels and modulation (Release 18)*.

[b-3GPP TS 38.212] 3GPP TS 38.212 V18.3.0 (2024), *NR; Multiplexing and channel coding (Release 18)*.

[b-3GPP TS 38.213] 3GPP TS 38.213 V18.3.0 (2024), *NR; Physical layer procedures for control (Release 18)*.

[b-3GPP TS 38.214] 3GPP TS 38.214 V18.3.0 (2024), *NR; Physical layer procedures for data (Release 18)*.

[b-3GPP TS 38.300] 3GPP TS 38.300 V18.2.0 (2024), *NR; NR and NG-RAN overall description; Stage 2 (Release 18)*.

[b-3GPP TS 38.521-1] 3GPP TS 38.521-1 V18.2.0 (2024), *NR; User equipment (UE) conformance specification; Radio transmission and reception; Part 1: Range 1 standalone (Release 18)*.

[b-3GPP TS 38.521-4] 3GPP TS 38.521-4 V18.3.0 (2024), *NR; User equipment (UE) conformance specification; Radio transmission and reception; Part 4: Performance requirements (Release 18)*.

[b-ETSI TR 103 559] ETSI TR 103 559 V1.2.1 (2023), *Speech and multimedia transmission quality (STQ); Best practices for robust network QoS benchmark testing and scoring*.

[b-5G PPP] Maternia, M., El Ayoubi, S.E., editors (2016). *Living document on 5G PPP use cases and performance evaluation models.* Heidelberg: 5G-PPP. 39 pp. Available [viewed 2024-07-18] at: https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-use-cases-and-performance-evaluation-modeling_v1.0.pdf

[b-Kozamernik] Kozamernik, F., Steinmann, V., Sunna, P., Wyckens, E. (2005). SAMVIQ – A new EBU methodology for video quality evaluations in multimedia. *SMPTE Motion Imaging J.* 114, pp. 152-60.

[b-NGMN] El Hattachi, R., Erfanian, J., editors (2015). *NGMN 5G white paper*Düsseldorf: NGMN Alliance. 125 pp. Available [viewed 2024-07-18] at: https://ngmn.org/wp-content/uploads/NGMN_5G_White_Paper_V1_0.pdf

[b-Nielsen] Nielsen, J. (1993). Response times: The three important limits. In: *Usability engineering*, chapter 5. San Francisco, CA: Morgan Kaufmann.

[b-Opentap] Opentap (2024). *opentap*. Chennai: Opentap. Available [viewed 2024-07-18] at: https://github.com/opentap/opentap

[b-Opentap-dev] Opentap (Internet). *Developer guide*. Chennai: Opentap. Available [viewed 2024-08-10] at: https://doc.opentap.io/Developer%20Guide/Introduction/Readme.html

[b-Opentap-doc] Opentap (Internet). *Overview*. Chennai: Opentap. Available [viewed 2024-07-19] at: https://doc.opentap.io/User%20Guide/Introduction/

[b-TriangleD2.2] DEKRA Testing and Certification, editors (2017). *Formalization of the certification process, requirements and use*, Deliverable D2.2, Project: H2020-ICT-688712, 5G Applications and Devices Benchmarking (Triangle). Chennai: Opentap. 87 pp. Available [viewed 2024-07-19] at: https://www.triangle-project.eu/wp-content/uploads/2017/07/TRIANGLE_D2.2_FINAL.pdf

[b-TriangleResources] Triangle Project. *Resources*. Available [viewed 2024-07-18] at: https://www.triangle-project.eu/resources/

# SERIES OF ITU-T RECOMMENDATIONS

| | |
|---|---|
| Series A | Organization of the work of ITU-T |
| Series D | Tariff and accounting principles and international telecommunication/ICT economic and policy issues |
| Series E | Overall network operation, telephone service, service operation and human factors |
| Series F | Non-telephone telecommunication services |
| **Series G** | **Transmission systems and media, digital systems and networks** |
| Series H | Audiovisual and multimedia systems |
| Series I | Integrated services digital network |
| Series J | Cable networks and transmission of television, sound programme and other multimedia signals |
| Series K | Protection against interference |
| Series L | Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant |
| Series M | Telecommunication management, including TMN and network maintenance |
| Series N | Maintenance: international sound programme and television transmission circuits |
| Series O | Specifications of measuring equipment |
| Series P | Telephone transmission quality, telephone installations, local line networks |
| Series Q | Switching and signalling, and associated measurements and tests |
| Series R | Telegraph transmission |
| Series S | Telegraph services terminal equipment |
| Series T | Terminals for telematic services |
| Series U | Telegraph switching |
| Series V | Data communication over the telephone network |
| Series X | Data networks, open system communications and security |
| Series Y | Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities |
| Series Z | Languages and general software aspects for telecommunication systems |

Published in Switzerland
Geneva, 2024