International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

E.806

(06/2019)

SERIES E: OVERALL NETWORK OPERATION, TELEPHONE SERVICE, SERVICE OPERATION AND HUMAN FACTORS

Quality of telecommunication services: concepts, models, objectives and dependability planning – Terms and definitions related to the quality of telecommunication services

Measurement campaigns, monitoring systems and sampling methodologies to monitor the quality of service in mobile networks

Recommendation ITU-T E.806

ITU-T E-SERIES RECOMMENDATIONS

OVERALL NETWORK OPERATION, TELEPHONE SERVICE, SERVICE OPERATION AND HUMAN FACTORS

INTERNATIONAL OPERATION	
Definitions	E.100–E.103
General provisions concerning Administrations	E.104–E.119
General provisions concerning users	E.120–E.139
Operation of international telephone services	E.140–E.159
Numbering plan of the international telephone service	E.160–E.169
International routing plan	E.170–E.179
Tones in national signalling systems	E.180–E.189
Numbering plan of the international telephone service	E.190–E.199
Maritime mobile service and public land mobile service	E.200–E.229
OPERATIONAL PROVISIONS RELATING TO CHARGING AND ACCOUNTING IN THE INTERNATIONAL TELEPHONE SERVICE	
Charging in the international telephone service	E.230–E.249
Measuring and recording call durations for accounting purposes	E.260–E.269
UTILIZATION OF THE INTERNATIONAL TELEPHONE NETWORK FOR NON-TELEPHONY APPLICATIONS	
General	E.300–E.319
Phototelegraphy	E.320–E.329
ISDN PROVISIONS CONCERNING USERS	E.330–E.349
INTERNATIONAL ROUTING PLAN	E.350–E.399
NETWORK MANAGEMENT	
International service statistics	E.400–E.404
International network management	E.405–E.419
Checking the quality of the international telephone service	E.420–E.489
TRAFFIC ENGINEERING	
Measurement and recording of traffic	E.490–E.505
Forecasting of traffic	E.506–E.509
Determination of the number of circuits in manual operation	E.510–E.519
Determination of the number of circuits in automatic and semi-automatic operation	E.520–E.539
Grade of service	E.540–E.599
Definitions	E.600–E.649
Traffic engineering for IP-networks	E.650–E.699
ISDN traffic engineering	E.700–E.749
Mobile network traffic engineering	E.750–E.799
QUALITY OF TELECOMMUNICATION SERVICES: CONCEPTS, MODELS, OBJECTIVES AND DEPENDABILITY PLANNING	
Terms and definitions related to the quality of telecommunication services	**E.800–E.809**
Models for telecommunication services	E.810–E.844
Objectives for quality of service and related concepts of telecommunication services	E.845–E.859
Use of quality of service objectives for planning of telecommunication networks	E.860–E.879
Field data collection and evaluation on the performance of equipment, networks and services	E.880–E.899
OTHER	E.900–E.999
INTERNATIONAL OPERATION	
Numbering plan of the international telephone service	E.1100–E.1199
NETWORK MANAGEMENT	
International network management	E.4100–E.4199

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

Recommendation ITU-T E.806

Measurement campaigns, monitoring systems and sampling methodologies to monitor the quality of service in mobile networks

Summary

Recommendation ITU-T E.806 describes a baseline framework of best practices for measuring quality of service (QoS) in mobile networks. It provides a high-level overview of measurement campaigns, monitoring systems characteristics and requirements, post-processing general recommendations and sampling methodologies to monitor mobile electronic services.

This Recommendation is technology-neutral, but may state different requirements depending on the services being measured.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T E.806	2019-06-29	12	11.1002/1000/13924

Keywords

KPI, measurement campaigns, mobile networks, monitoring systems, QoS, sampling methodologies.

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

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, 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 TSB patent database at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Recommendation ITU-T E.806

Measurement campaigns, monitoring systems and sampling methodologies to monitor the quality of service in mobile networks

1 Scope

This Recommendation describes a baseline framework of best practices for measuring quality of service (QoS) throughout the industry, and covers mobile network QoS measurement campaigns, characteristics and requirements for monitoring systems, post-processing scenarios, as well as sampling methodologies used by regulators, test equipment vendors, companies that deliver network measurements, data analysts and service providers to monitor QoS on a national level.

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 E.802] Recommendation ITU-T E.802 (2007), *Framework and methodologies for the determination and application of QoS parameters*.

[ITU-T E.804] Recommendation ITU-T E.804 (2014), *Quality of service aspects for popular services in mobile networks.*

[ITU-T E.807] Recommendation ITU-T E.807 (2014), *Definitions, associated measurement methods and guidance targets of user-centric parameters for call handling in cellular mobile voice service.*

[ITU-T E.840] Recommendation ITU-T E.840 (2018), *Statistical framework for end-to-end network performance benchmark scoring and ranking.*

[ITU-T G.1031] Recommendation ITU-T G.1031 (2014), *QoE factors in web-browsing*.

[ITU-T P.863.1] Recommendation ITU-T P.863.1 (2019), *Application guide for Recommendation ITU-T P.863*.

[ITU-T P.1401] Recommendation ITU-T P.1401 (2012), *Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models*.

[ITU-T Y.1540] Recommendation ITU-T Y.1540 (2016), *Internet protocol data communication service – IP packet transfer and availability performance parameters*

[ITU-T Y.1545.1] Recommendation ITU-T Y.1545.1 (2017), *Framework for monitoring the quality of service of IP network services*.

[ETSI TR 125 942] ETSI Technical Report 125 942, V15.0.0 (2018), *Universal mobile telecommunications system (UMTS); Radio frequency (RF) system scenarios*.

[ETSI TR 138 900] ETSI Technical Report 138 900, V15.0.0 (2018), *LTE; 5G; Study on channel model for frequency spectrum above 6 GHz.*

[IETF RFC 7799] IETF RFC 7799 (2016), *Active and passive metrics and methods (with hybrid types in-between)*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 A-party [ITU-T E.804]: Initiating part of a connection (also: Mobile Originating, MO) or in direct transactions, the party initiating the transaction (calling party).

NOTE – In store-and-forward transactions, the party sending the content.

3.1.2 B-party [ITU-T E.804]: In direct transactions, the termination or counterpart of a transaction.

NOTE – In store-and-forward transactions, the party receiving the content.

3.1.3 confidence coefficient; confidence level [b-ITU-T E.800]: The value of the probability associated with a confidence interval or a statistical tolerance interval.

3.1.4 download [ITU-T E.800]: Transfer of data or programs from a server or host computer to one's own computer or device.

3.1.5 end-to-end quality [b-ITU-T E.800]: Quality related to the performance of a communication system, including all terminal equipment.

3.1.6 measure [b-ITU-T E.800]: A unit by which a parameter may be expressed.

3.1.7 network operator [ITU-T E.804]: Organization that provides a network for the provision of a public telecommunication service.

3.1.8 network performance [b-ITU-T E.800]: The ability of a network or network portion to provide the functions related to communications between users.

3.1.9 probe [ITU-T Y.1545.1]: Is an end-point test tool which uses probing packets to collect measurements.

3.1.10 quality of service (QoS) [b-ITU-T E.800]: Totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service.

3.1.11 streaming (in multimedia services) [ITU-T E.800]: Multimedia data (usually combinations of voice, text, video and audio) transferred in a stream of packets that are interpreted and rendered, by a software application as the packets arrive.

NOTE 1 – Streaming is a technique for transferring multimedia data.

NOTE 2 – Streaming may or may not be in real time.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 crowdsourced data collection: A method to gather active or passive quality of service measurements from a large number of end user devices.

3.2.2 drive testing: A testing method using vehicles equipped with network device-testing equipment.

3.2.3 monitoring system: A tool or equipment able to perform an evaluation of network performance.

3.2.4 monitoring system user: A user that can be a regulator, test equipment vendor, company that delivers network measurements, data analyst or service provider.

3.2.5 unattended probe: Remotely managed mobile or fixed monitoring system (smartphone, single-sim device, dual-sim device, virtual-sim device) that can be installed in indoor or outdoor locations, or in vehicles.

3.2.6 walk testing: A testing method using device, carry-on testing equipment, normally employed where vehicles cannot be used.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

2G	second Generation
3G	third Generation
4G	fourth Generation
BS	Base Station
BTS	Base Transceiver Station
GIS	Geographic Information System
GPS	Global Positioning System
HTTP	Hypertext Transfer Protocol
KPI	Key Performance Indicator
MMS	Multimedia Messaging Service
MNO	Mobile Network Operator
OMC	Outdoor Measurement Campaign
QoS	Quality of Service
SINR	Signal to Interference plus Noise Ratio
SMS	Short Message Service
UPS	Uninterruptible Power Supply

5 Conventions

None.

6 Measurement campaigns to monitor quality of service

QoS measurement campaigns aim to gather information to characterize the QoS of mobile networks in terms of key performance indicators (KPIs), perform benchmarking and evaluate conformance to existing legislation commitments.

The intended scope of QoS measurements related to geographical footprint, timeframe and testing frequency depends mainly on the purpose of the measurement campaign and the measurement environment.

6.1 Measurement environment

For mobile networks, measurement environments can be divided into two main categories: indoor and outdoor. Measurements can include both indoor and outdoor, stationary and mobile scenarios thus covering all the different environments in which end users make use of their mobile service. Currently, there is a high rate of mobile traffic generated from indoor environments; hence, it is advisable to measure indoor QoS performance, in addition to outdoor.

6.1.1 Indoor testing

The following recommended methodologies can help the monitoring system user to identify the indoor QoS status:

- walk testing;
- unattended probes;
- crowdsourced data collection.

6.1.2 Outdoor testing

The following recommended methodologies can help the monitoring system user to identify the outdoor QoS status:

- drive/walk testing;
- unattended probes;
- crowdsourced data collection.

Clause 6.2 provides a high-level description of the four different measurement methodologies that a monitoring system user can perform for indoor and outdoor testing.

6.2 Measurement methodologies

6.2.1 Walk testing

Locations such as train stations, underground train platforms, airports, sports stadiums, shopping malls, university campuses and pedestrian zones are becoming hotspots for wireless communication. In outdoor measurements, walk testing is normally employed where vehicles cannot be used. Multi-device, carry-on testing equipment provides a convenient way to work when benchmarking multiple operators or testing coverage for multiple services and radio access technologies.

However, walk testing measurement campaigns have some limitations in terms of number of samples, mobile network operator (MNO) and service due to physical parameters such as spatial dimensions, equipment weight and technical requirements, like antenna isolation, which ultimately can influence sample resolution. The following actions are recommended before initiating a walk testing measurement.

1 Draft blueprints of building layouts, in the case of an indoor location.

2 Schedule and draft a list of targeted spots or locations.

3 The targeted sample size per service or MNO depends on the number of smartphones that can be accommodated by the equipment. However, the sample size and distribution should be chosen according to, among others, the type of variable under consideration and the statistical representation being targeted. For example, covering a wider area or measuring over a longer period could provide more statistically representative results. See Annex A and [ITU-T E.840] for references.

4 Adopt and conduct a general routine procedure to check hardware or software functionality.

5 Consider a light portable control unit to report real-time results or status of the measurement equipment.

6 The automatic network selection feature should be blocked and set to the on-net feature.

6.2.2 Drive testing

Drive testing measurement campaigns involve preplanning procedures to satisfy the goal and scope of the campaign. Targeted services, demographic distribution and socio-economic factors can impose different campaign design parameters. Essentially, a successful drive testing campaign shall consider a sample size so that the measured data are representative of the targeted region's population.

In order to perform drive testing measurement campaigns, monitoring system users should consider the following recommendations:

1 The sample size should be chosen so that the results are representative of the behaviour of the mobile networks in the area under study. For this purpose, monitoring system users must define a sampling methodology. References can be found in Annex A and [ITU-T E.802].

2 An initial radio coverage footprint or MNO/technology is required for planning.

3 The antennas used to perform the measurements should be installed at average human height.

4 In the case of a regulator's benchmarking, measurements should be performed randomly, for all access technologies and for all MNOs simultaneously.

5 Measurement profiles should be established (technical references can be found in [ITU-T E.804]).

6 The routes should cover areas with human activity, avoiding route repetition if drive testing is used to derive KPIs for a wide area.

7 For services that are evaluated in movement, the speed of the vehicle should be set considering that one device can be located at a fixed point while another device will be in motion.

8 The automatic network selection feature should be blocked and set to the on-net feature.

9 The targeted region's population distribution should be studied prior to the campaign to guarantee that:

   a. the collected sample may consider residential and commercial concentrations;

   b. development factors of sub-regions may be considered.

10 Business days and daytime testing intervals should be considered preferentially.

6.2.3 Unattended probes

Unattended probes may provide near real time and historical end-to-end QoS performance and can be used to collect granular data that can help to detect QoS degradations. Monitoring system users are advised to consider the following operating and design recommendations to deploy a probe-based measurement campaign.

1 Unattended probes should be installed securely in chosen locations or vehicles with power supply; the choice of these locations or vehicles depends on the measurement objective.

2 Indoor unattended probes should be installed in locations with adequate radio coverage conditions and include high network usage. Depending on the measurement scope, this may be a location with the best possible coverage if peak network performance is monitored. If unattended probes are intended to monitor minimum radio coverage, a location with non-optimal coverage conditions can be considered.

3 All unattended probes should operate under a global positioning system (GPS) (or another suitable global navigation satellite system) connection.

4 The number of unattended probes per targeted location depends on the sampling requirements, the number of services or operators to be evaluated, user density and occurring events (concerts, sports tournaments, etc.).

Knowing that network usage can change dramatically in high human activity locations when considering the need for efficient use of their measurement resources, monitoring system users are recommended to use manageable unattended probes that can be remotely activated once a location is considered an active hotspot.

6.2.4 Crowdsourced data collection

Crowdsourced data collection campaigns can be used to obtain QoS measurement data that can only be meaningful if a representative number of samples from different end users are gathered. To gain a deeper insight into the results, environmental conditions can be checked by using data-enabled end user devices. It is recommended that measurements results be used for any conclusion only if they are representative for the respective measurement goal. Data gap limitations have to be taken into account. Crowdsourced data collection solutions must be in compliance with national data protection legislation, guaranteeing that no personal data is mishandled.

The sample collection frequency can be continuous unlike the conventional (simulating collection) QoS sample collection, therefore crowd-sourced measurements can be a source of 24 hours by 7 days QoS measurements.

Crowdsourced data collection can be largely categorized into two categories namely, active and passive measurements (see [IETF RFC 7799]):

NOTE – Alternatives sources to collect continuous information on network QoS are discussed in [ITU-T E.804]. Depending on the legal framework in each country, regulators may have access to network performance counters information. In accordance to data protection legislation and where both parties so desire, operators and regulators may want to enter into non-disclosure agreements as per the foregoing.

6.2.4.1 Active Measurements

Active data measurements are typically data speed tests and application-specific tests that are end user initiated or scripted. The active approach measurement creates artificial traffic to determine the network/application capability.

When measuring the capacity, capability or performance of the network using active crowdsourcing approach, monitoring system users are recommended to perform QoS measurements using automated scheduled scripts against targeted servers located on-net and off-net (e.g., in the closest Internet exchanging point). Also, provision for available bandwidth in the servers and links has to be considered to prevent measurements from being affected adversely by the degradation attributed to the servers or test links.

Moreover, for regulator developed crowdsourcing application which may be integrated with operator SIM or an independent downloadable application, it is important to avoid scheduling the measures of all end users at the same time to avoid possible congestion.

6.2.4.2 Passive measurements

Passive data measurements are either application-specific or application-agnostic and do not require any form of end user intervention. Passive measurements do not inject artificial traffic or test payload into the network and aim at measuring QoS based on the end users' activity.

This approach removes the dependency on the servers and links and may leverage on the popularity of the host mobile applications (e.g., transport app, government app, MNO app) to provide a representative geographical distribution of the measurement data points.

It is recommended that monitoring system users consider more samples (end users), but from unique identifiers as they provide monitoring system users with higher network data accuracy.

6.3 Guidelines for measuring quality of service parameters on mobile networks

This section provides further details on the measurement campaigns for monitoring some relevant parameters. The parameters listed here can be measured by any of the methodologies described in clause 6.2 and should be considered as examples only; more references can be found in [ITU-T E.804].

6.3.1 Radio coverage

The coverage measurements consist of taking the received signal level at a given geographical position for each radio access network technology separately. These measurements are carried out automatically using a frequency scanner.

Regulators are advised to take into consideration coverage maps for the target area using MNO data submitted, acquired or generated prior to the drive test campaign. Coverage maps reflect the area where MNOs provide service. In this sense, a regulator can use them to plan a measurement campaign and select areas of interest. For example, the regulator can choose to measure areas that have the highest rates of complaint or, for the sake of comparison, select areas where all MNOs have coverage. Coverage maps can also be validated during measurement campaigns. For 4G and above networks, service coverage is another relevant parameter to measure. References for measurement methodologies for coverage and service coverage can be found in [b-ECC Report 103], [b-ECC Report 256], [b-ECC Report 118], [b-ECC Report 231] and [b-ECC Rec (12) 03].

A monitoring system user may use different approaches to represent and visualize mobile network coverage. Clauses 6.3.1.1 and 6.3.1.2 describe two approaches that can be used independently or combined depending on the monitoring system user needs and available information.

6.3.1.1 Theoretical or analytical modelling measurement methodology

This method is mostly based on formulation of mathematical and statistical computations based on simulation tools for predicting network coverage and performance. Some of the parameters involved in such measurements include the location of base station (BS), signal to interference plus noise ratio (SINR), BS power, service SINR threshold, attenuation of signals for indoor coverage and antenna information or sensitivity per service.

A monitoring system user could use such models if information about the MNO infrastructure is available and should be used only as a reference since it is based on statistical predictions. For more information on propagation models, see [ETSI TR 125 942], [ETSI TR 138 900] and the references therein.

6.3.1.2 Geographic information systems

A geographic information system (GIS) allows advanced mapping and spatial analytics to visualize geographic information on a map. With this kind of tool, it is possible to integrate different kinds of data layer using spatial location, as it is the coverage service maps of the MNO. Such information layers can be:

- base transceiver station (BTS) layer for the MNO;
- output results (scaled samples) of QoS measurements;
- specific coverage or network indicator from crowdsourcing system;
- population and administrative layers;
- terrain and elevation layer.

Moreover, the GIS-generated layer can be easily shared through websites and is accessible via web browsers, mobile phones and tablets.

For this reason, it is advisable that monitoring system users have access to coverage maps compatible with GIS tools to provide clear and robust information to end users and that the information be presented in a clear, friendly and understandable way.

Using GIS systems, a monitoring system user may run analysis to determine, for example, the geographical regions to measure, plan routes for drive testing, and perform population coverage analysis. Examples of GIS analysis are given in [b-ECC Report 103], [b-ECC Report 256], [b-ECC Report 118], [b-ECC Report 231] and [b-ECC Rec (12) 03].

A GIS system should be able to perform union, intersection, overlapping and areal definition (points, polygons), generate heat maps and thematic maps. Moreover, it should allow running grid analysis in different resolutions to characterize geographical regions, to show the variability of a KPI and allow the analysis of population and clutter maps.

6.3.2 Voice call measurement

Voice service measurements include the launch of a series of call attempts for automatic selection of radio access technologies. Call attempts are generated from different scenarios, can be mobile to mobile, fixed line to mobile, or mobile to fixed line. An appropriate sample for voice measurements should be chosen according to the scenario. Information can be found in [ITU-T P.863.1] and [ITU-T E.807].

The timeline scenario in Figure 1 for an active voice call test is just an example and does not lay down specifications. Besides the radio network characteristics, the associated time interval parameters and the overall window test time will affect the total acquired number of samples.

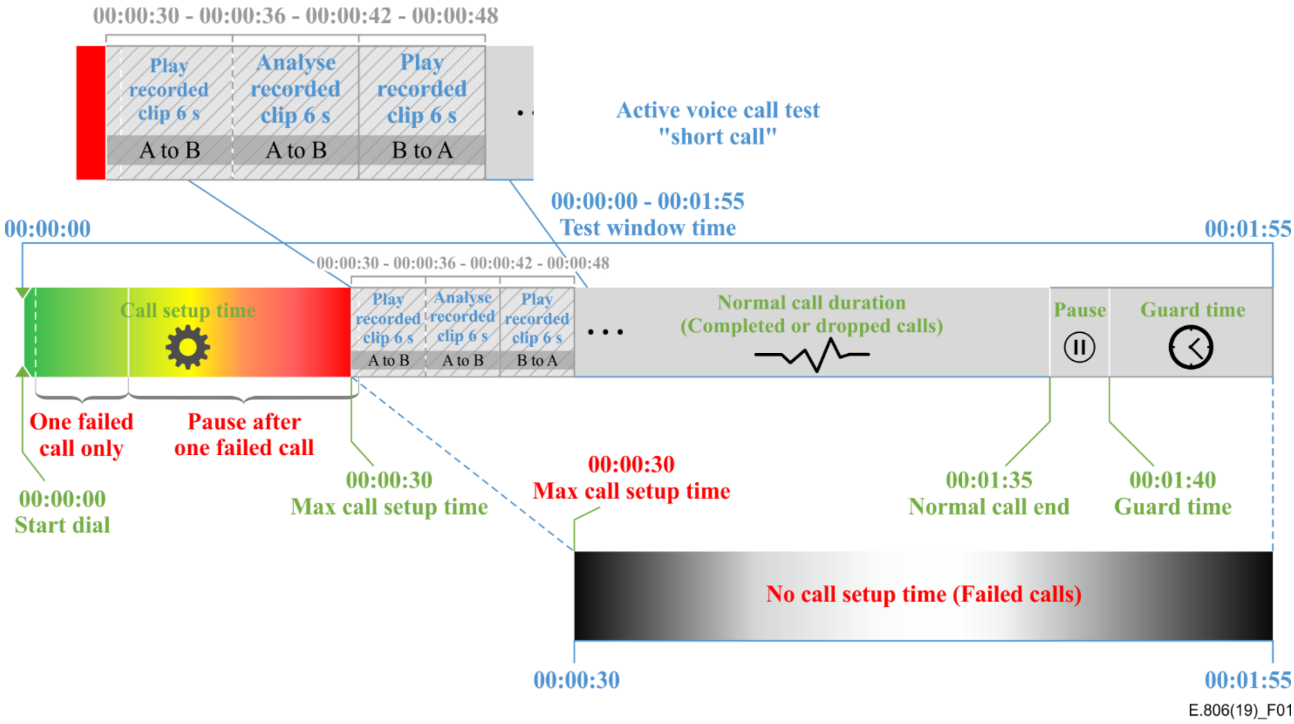

Figure 1 – Timeline for call measurements

The scenario depicted represents an example of a voice call testing procedure; while the normal call duration may vary depending on the purpose of the test (short call, long call), almost all other properties are the same for each case faced during the measurement test.

Additionally, specific characteristics per country behaviour should be considered as it can impact the scenario design, for example, if statistically the average call time for country A is 2 min while country B is 4 min, the minutes of usage per dropped call would be higher.

Before the test starting point, the calling party (A) starts dialling the preconfigured called party (B), in that case a window of fixed duration is allowed as the maximum call setup time to establish a connection to party B, in which the test is waiting for party B to answer the test call. If party B does not answer or the calling party has issues during the test call setup phase, the test call procedure enters the "no call setup time", where the test call will be flagged as a "failed call".

6.3.3 Short and multimedia message measurement

Short message service (SMS) and multimedia messaging service (MMS) measurements can be executed without forcing the mobile terminal equipment to a particular access technology to simulate a similar scenario when the end user mobile terminal equipment is continually changing access technology.

Measurements consist of sending an SMS with a fixed number of alphanumeric characters and fixed size for MMS from a mobile probe simulating a mobile subscriber to a fixed one simulating another mobile subscriber belonging to the same operator. The SMS/MMS is considered as received if the delivery time is less than the maximum time established.

Figure 2 shows a reference timeline diagram for an SMS measurement.

Figure 2 – Timeline for short message service measurements

Both, voice service and SMS/MMS measurements can be done in a simultaneous way to better manage time resources.

6.3.4 Broadband data measurement

Broadband data measurement campaigns should be balanced on different monitoring systems to reduce the window test frame and increase the number of samples; this will enable greater coverage of the testing map area.

The setup for the scenarios will depend on the service application being measured. Figure 3 shows an example of a timeline diagram for the evaluation of data services. More information can be found in [ITU-T Y.1540], [ITU-T Y.1545.1].

NOTE – When measurements involve specific server resources, user equipment or unattended probes, or multiple networks to reach the desired content or test server, then their contribution to results cannot be distinguished from mobile network performance. See clause 6 of [ITU-T G.1031] for an example list of such influence factors contributing to results.

Clauses 6.3.4.1 to 6.3.4.3 are examples of broadband data measurements.

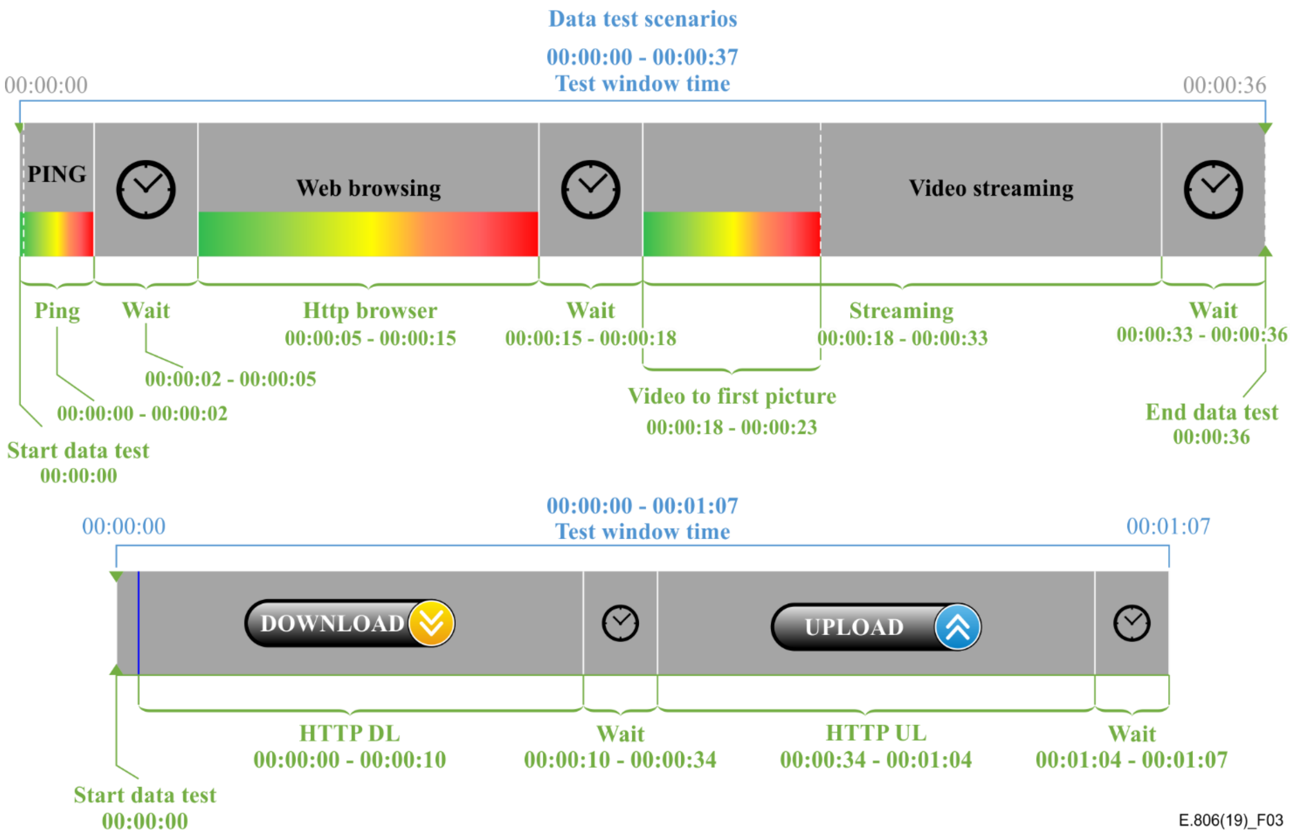

Figure 3 – Example of broadband data measurements

6.3.4.1 Web-browsing measurement

Measurements related to web browsing can be characterized, among others, by measuring the performance of navigation in terms of page load time and the overall success rate for page download. Web-browsing tests, with an average fixed duration, are launched from mobile equipment or a probe simulating a mobile subscriber in a static or mobile situation to a predetermined set of recommended websites. Access attempts are randomly distributed between the different sites tested.

6.3.4.2 File transfer measurement

File transfer measures consist of sending and receiving files of a fixed size for uplink and downlink channels between mobile devices or a probe simulating a mobile subscriber in a fixed location and a server having resources dedicated to this measurement. KPIs such as download session success rate, mean download time or mean upload time can be measured (with limits placed on the time allowed for completion in either direction.

6.3.4.3 Audio or video streaming measurement

For audio or video streaming services, the success rate of access to a set of audio or video files hosted on streaming servers could be measured through average fixed duration tests that are launched from a mobile probe simulating a mobile subscriber in a fixed location to a predetermined series of streaming servers. Furthermore, the initial video loading duration, the frequency and duration of stalling events, and the overall throughput can be used as indicators for streaming performance.

Access attempts are randomly allocated between the different servers tested.

7 Characteristics and requirements for monitoring systems

Monitoring systems should be able to perform an evaluation on how degraded network performance affects service quality to the end user. It therefore gives an indication of network coverage performance, capacity and end-to-end QoS.

The following monitoring systems characteristics apply to all measurement methodologies described in clause 6.2, unless otherwise stated, except for crowdsourced data collection.

1 Ability to execute simultaneous measurements in all access technologies.

2 Ability to execute simultaneous measurements, e.g., voice, SMS and data, if needed.

3 It should be possible to carry out the measurements in movement, at static locations or combining both modalities.

4 It should automatically store data from measurements and should count with additional external storage for backup.

5 It should have a GPS (or another suitable global navigation satellite system) to identify and register the geographical location and speed for each measurement.

6 It should be kept in optimal operation conditions through maintenance programmes considering all applicable regulation certification requirements. For example, scanner equipment should be calibrated. The validity of the calibration should be within the time period of the measurement campaigns.

7 Antennas must be placed at an appropriate distance to avoid interference.

8 It should be able to execute measurements for all access technologies and therefore be updated to meet the needs of technological evolution.

9 It should allow the setup and reuse of measurement script templates.

10 It should use certified mobile equipment similar to those regularly purchased in MNO shops by end users.

11 It should support voice codecs for all access technologies.

12 It should be able to produce visual or audio alarms to notify any failure in the measuring equipment; measurements performed during this scenario should be discarded.

13 Measuring equipment should provide a feature, as an added functionality or additional equipment, to register the day, hour and geo-referenced position of each measurement.

14 It is desirable that the measurement equipment provide the ability to register, at least, the start and the end of the measurement campaign, breaks, equipment failures and extraordinary situations such as atypical end user concentration, vandalism, interference or natural disasters.

15 Any situation that could potentially affect the results and is not attributed to the network of the MNO should be registered.

16 All activities registered by the measurement equipment, mentioned in the foregoing, should be correlated with the log files and measurements taken during these time intervals and should be discarded during the post-processing stage.

17 Monitoring systems should generate encrypted log files.

18 Unattended probes should be equipped with self-monitoring functionality (e.g., temperature sensors or adequate power supply level) to detect unusual operating conditions. There should be corresponding post processing to ensure data quality and to detect and exclude data artefacts.

19 Unattended probes should be robust against loss of external power supply, e.g., have power supply buffering to ensure secure turn-off processes and reliable turn-on processes. If unattended probes do not support this, an uninterruptible power supply (UPS) should be considered to ensure stable operating conditions. This is critical to avoid log file corruption generation.

20 Depending on the way measured data are transferred (e.g., via the monitored network), and the type of information that is expected to be created, unattended probes and back-end infrastructure shall provide adequate monitoring functionality. For instance, unattended

probes can be set to regularly transmit measured data, or send status reports at regular intervals, with back-end processes that generate alarms if such information is overdue or measured data indicate that selected network parameters are outside predefined ranges.

NOTE – The characteristics for crowdsourced data collection monitoring systems need further study.

8 General recommendations for post processing

The first step in the post processing of results is to determine the radio coverage limit for which the electronic service of a mobile communications network is considered to be provided. This limit should be set for each technology (e.g., 2G, 3G, 4G) separately. These limits will determine which tests will be taken into consideration when exporting the final results. Also, rules should be laid down as to when a measurement should be taken into account in the results. Such rules are, for example, if during a measurement there is a period of time that does not meet the radio coverage limit, how long should that period be for the test to be taken into account in the results.

The following general recommendations should be practised by all monitoring system users during post-processing action for all measurement methodologies.

1 In order to examine the events that occurred during the measurements, signalling information should be taken into account. Post processing of measurements results should consider and manage all log files obtained during the measurement campaign. It is recommended that all software errors be considered. The final number of measurements gathered in the measurement campaign will have an impact on the estimation error of the results. The estimation error can be calculated based on the sample size. If it is larger than desired, it is recommended that the measurement campaign for the target area be repeated. More information is provided in Annex A and [ITU-T E.802], [ITU-T E.840] and [ITU-T P.1401].

2 Monitoring system users conducting QoS measurements should check measurement data integrity and cleanse data. Outlier samples can be generated for many reasons, e.g., equipment alarms due to equipment or software failure or malfunction, network element upgrade, incidents or maintenance.

3 All log files gathered from the testing equipment should be checked, if missing or corrupted.

4 It is essential to set the accepted percentage threshold for fake or false samples collected in a measurement campaign. Except for crowdsourced data collection methodologies, it is important to determine the need to re-measure so as to maintain the accuracy and integrity of results.

9 Sampling methodologies

Measurements may be distributed between indoor and outdoor scenarios depending on each telecommunications market subscriber density or behaviour. The number of measurements must be determined in such a way that the statistical relative accuracy is less than the maximum value established and for at least a 95% level of confidence. More information is provided in [ITU-T E.802].

A two-step methodology that can be used to obtain representative samples involves stratification and simple random sampling.

Stratification can be used to calculate the number of geographical areas (e.g., cities, municipalities or districts) to be covered during a measurement campaign to get results that represent the network QoS at a national level. In this sense, stratification is a tool useful to obtain representative results when it is not possible to measure a large area (e.g., large countries).

Simple random sampling can then be used to calculate the number of measurements to perform at each of the geographical areas that were selected through sampling methods. More details are provided in Annex A.

It is important to consider the results obtained from each measurement campaign to derive the behaviour of the network in terms of performance. The mean and standard deviation values can be calculated from these results and used together with statistical formulas to monitor QoS parameters.

To perform measurements over a large area, monitoring system users are advised to divide the base zone into parts, and allocate a weight to each part in accordance with criteria that will identify where it is more relevant to carry out a greater number of measurements, depending on the purpose of the measurement campaign. Some examples of criteria are population weight, telecommunications density, traffic density and end user density; in order to define these criteria, it is highly advisable that monitoring system users obtain updated data from an official source.

Publication of benchmarking performance results lies outside the scope of this Recommendation; however, a statistical framework aimed to score and rank network performance can be found in [ITU-T E.840].

Annex A

Statistical guidelines to obtain representative results

(This annex forms an integral part of this Recommendation.)

A monitoring system user can obtain representative results of the QoS performance of the network based on sampling applied during measurement campaigns together with the use of statistical inference by following a procedure based on hypothesis testing.

A.1 Representative samples to monitor quality of service at a national level

The statistical modelling approach recommended for a monitoring system user who wants to monitor the QoS at a national level using outdoor measurement campaigns (OMCs) is based on a two-step model:

- stratified random sampling;
- simple random sampling.

The first step is used to select the geographical areas to be measured in the country (in this context, geographical area can refer to cities, municipalities, towns or roads), while the second step determines the sample size for each area selected in the first step, e.g., how many voice calls are done in a city. The distribution of measurements within each area can be chosen according to the purpose of the measurement, e.g., measurements can be uniformly distributed if no differentiation is needed or they can be weighted to give more importance to urban areas.

A.1.1 Stratified random sampling

From a statistical point of view, stratification is used to produce a smaller bound on the error of estimation than would be produced by a simple random sample of the same size alone. When monitoring QoS, stratification is a tool that can be used to reduce the number of regions to measure and still have representative results, e.g., it would not be possible to drive across countries that are large in geography, so the area can be divided into groups (strata) and a sample measured of each group.

In this sense, stratification allows distribution of geographical areas into groups or strata that are homogeneous or different. The groups are defined based on characteristics that are common to the group and have a direct influence over the KPI that is being measured. Examples of homogeneous strata can include, but are not limited to, geographical areas (cities or towns) with similar population or BSs density, traffic demand, interference levels, inter-site distance, etc. Then, the strata are constructed by selecting non-overlapping groups from the geographical regions in the country (e.g., if strata are chosen according to population density, localities in the country can be classified into urban or rural, and each locality is assigned to only one group, i.e., a locality cannot be urban and rural at the same time). Finally, geographical areas in each stratum to be measured have to be randomly selected to ensure validity of the results.

The following are statistical formulas for stratification sampling, including an example of urban and rural localities in a country.

The number of geographical areas to be measured during a campaign can be derived from the following formula (number of localities in a country that are going to be measured):

$$n = \frac{\left(\sum_{i=1}^{L} N_i \sigma_\mathrm{i}\right)^2}{N^2 D + \sum_{i=1}^{L} N_i \sigma_i^2}$$

where

L is the total number of strata (groups used to divide the country; in the case of urban and rural classification, L = 2);

σ_i is the expected standard deviation for stratum i;

N_i is the number of geographical areas in each stratum (number of localities classified as urban or rural);

N is the total number of geographical areas (total number of localities in a country);

$D = \frac{B^2}{4}$, where B is the bound on the error of estimation

$$n_i = n\left(\frac{N_i\sigma_i}{\sum_{i=1}^{L} N_i\sigma_i}\right) \quad i = 1,2,3$$

where

n_i is the number of geographical areas to be measured in stratum i (for i = 1 (urban), n_i would be the number of urban localities to be measured);

n is the total number of geographical areas to be measured (number of urban and rural localities to be measured).

A.1.2 Simple random sampling

After selecting the number of geographical areas in the country to be measured, the second step consists of a simple random sampling to determine the sample size for each of the geographical regions previously selected, i.e., how many voice calls, data measurements or SMSs are to be made or sent.

During the second step of the statistical modelling, simple random sampling is used to define the number of events needed to measure a certain KPI with a defined confidence level and error of estimation.

If the parameter under study is a mean value, the sample size is calculated as follows

$$m_i = \frac{z^2_{1-\alpha/2}}{a^2} \cdot \left(\frac{\sigma_i}{\bar{x}_i}\right)^2$$

where

m_i is the sample size for stratum i;

$z_{1-\alpha/2}$ is the percentile $1 - \alpha/2$ of a standard normal distribution;

$1 - \alpha$ is the confidence level;

a is the bound on the error of estimation;

$\bar{x}_i$ is the mean value for the parameter of observation in stratum i;

σ_i and $\bar{x}_i$ are calculated from previous measurement campaigns.

If the parameter under study is a proportion, the sample size for every KPI should be calculated using the following formula:

$$n \geq \frac{k^2 * P * (1 - P)}{d^2}$$

where

n is the sample size;

k is the confidence level;

P is the proportion target value (threshold for the KPI under study);

d is the maximum acceptable estimation error (difference between the actual value P and its estimation p.

For every stratum, the standard deviation σ_i and the mean value $\bar{x}_i$ corresponding to stratum i obtained from the measurement campaign can be obtained as follows:

$$\sigma_i^2 = \frac{\sum_{k=1}^{n_i}(x_k - \bar{x}_i)^2}{n_i - 1}$$

$$\bar{x}_i = \frac{\sum_{k=1}^{n_i} x_k}{n_i}$$

$$x_k = \frac{\sum_{j=1}^{m_i} p_j}{m_i}$$

where

σ_i^2 is the variance for the parameter under observation in stratum i;

x_k is the mean value for each of the n_i samples corresponding to stratum i;

P_j is the value for measurement i within the sample of size m_i.

In order to determine whether the operator achieves a certain threshold, it is recommended that hypothesis testing based on the measurement results for all strata be used.

A.2 Hypothesis testing

Hypothesis testing is a statistical inference tool to determine whether the results obtained from the selected sample are meaningful from a statistical perspective, implying that the MNO follows a minimum QoS threshold.

To apply statistical hypothesis testing, the mean and standard deviation of the weighted values of the measurement campaign results are determined:

A weight is defined based on the population on each stratum N_i, with respect to the total population N:

$$w_i = {}^{N_i}\!/_{N}$$

Weighted values for mean and standard deviation for the KPI based on the mean values for each stratum i are given by:

$$\bar{x} = \sum_{i=1}^{L} w_i \bar{x}_i$$

$$\sigma^2 = \sum_{i=1}^{L} w_i^2 \, \sigma_{\bar{x}_i}^2$$

After calculating the mean and standard deviation for the measurement results, the first step in hypothesis testing is to define a null and an alternative hypothesis. Then, a statistical hypothesis test defines whether the null hypothesis is correct, with a certain probability, or the alternative hypothesis is accepted.

Considering the measurement results, the hypothesis test can be performed based on the total mean value and the KPI threshold that has been established. The null hypothesis (H_0) states that the MNO follows the minimum KPI threshold (μ), meanwhile, the alternative hypothesis (H_a) states that the MNO does not comply with such a threshold. The statistical hypothesis test has a significance level of α.

In this way, the hypotheses can be expressed as:

$$H_0: x_{st} \geq \mu$$

$$H_a: x_{st} < \mu$$

Based on the data from the measurements samples obtained from the measurement campaign, the statistical hypothesis test can determine whether the null hypothesis should be accepted or rejected.

An example follows of how to apply hypothesis testing when KPI thresholds are established by a regulator in the form of proportions. For more information and references on different metrics and procedures for statistical evaluation, see [ITU-T E.840], [ITU-T P.1401] and the references therein.

The decision is taken by comparing the test statistic (x_{st}), that can be computed from a critical value $z_{1-\alpha}$:

$$x_{st} = \frac{\bar{x} - \mu}{\sigma}$$

where

μ is the KPI threshold established by the regulator (mean value);

$$X = \frac{p_{st} - p}{\sqrt{\frac{p\,(1-p)}{n}}}$$

where

p is the KPI threshold established by the regulator (proportion);

$Z_{1-\alpha}$ is the z value which corresponds to a standard normal distribution with a significance level of α;

n is the total number of samples obtained from the measurement campaigns for each QoS parameter and considering all strata.

Finally, if the test statistic (x_{st}) is greater than or equal to the critical value $z_{1-\alpha}$, then, statistically, there is not enough information to reject the null hypothesis with a significance level of α (which means the MNO is in compliance with the threshold established by the regulator and no penalty would be imposed); otherwise, the null hypothesis is rejected, and the alternative hypothesis is accepted (MNO does not comply with the regulator's threshold).

Bibliography

[b-ITU-T E.800] Recommendation ITU-T E.800 (2008), *Definitions of terms related to quality of service*.

[b-ECC Report 103] Electronic Communications Committee within the European Conference of Postal and Telecommunications Administrations (CEPT) Report 103 (2007), *UMTS coverage measurements.*

[b-ECC Report 256] Electronic Communications Committee within the European Conference of Postal and Telecommunications Administrations (CEPT) Report 256 (2016), *LTE coverage measurements*.

[b-ECC Report 118] Electronic Communications Committee within the European Conference of Postal and Telecommunications Administrations (CEPT) Report 118 (2008), *Monitoring methodology to assess the performance of GSM networks*.

[b-ECC Report 231] Electronic Communications Committee within the European Conference of Postal and Telecommunications Administrations (CEPT) Report 231 (2015), *Mobile coverage obligations*.

[b-ECC Rec (12) 03] Electronic Communications Committee within the European Conference of Postal and Telecommunications Administrations (CEPT) Recommendation (12) 03 (2013), *Determination of the radiated power through field strength measurements in the frequency range from 400 MHz to 6000 MHz*.

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

Printed in Switzerland
Geneva, 2019