International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

E.812

Amendment 1

(09/2020)

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

Quality of telecommunication services: concepts, models, objectives and dependability planning – Models for telecommunication services

---

Crowdsourcing approach for the assessment of end-to-end quality of service in fixed and mobile broadband networks

Amendment 1

Recommendation ITU-T E.812 (2020) – Amendment 1

International Telecommunication Union

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.812

Crowdsourcing approach for the assessment of end-to-end quality of service in fixed and mobile broadband networks

Amendment 1

Summary

End-user equipment, consumer premise equipment and its software have evolved to become faster, more powerful and able to perform data collection. This has enabled the crowdsourcing approach which seeks to increase the amount of technical parameters which can be collected from end-users without modification to existing hardware and software.

Increasingly, players such as regulators and service providers have started to assess end-to-end quality of service (QoS) through a crowdsourcing approach. However, assessment using data collected through the crowdsourcing approach can be deployed in multiple ways and different approaches provide different views of QoS.

Recommendation ITU-T E.812 outlines the different crowdsourcing approaches used to assess end-to-end QoS on both fixed and mobile broadband networks.

Amendment 1 to Recommendation ITU-T E.812 introduces Appendix II (Use cases for the crowdsourcing approach) and Appendix III (Practical approaches to fixed broadband crowdsourcing).

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T E.812	2020-05-29	12	11.1002/1000/14272
1.1	ITU-T E.812 (2020) Amd. 1	2020-09-11	12	11.1002/1000/14489

Keywords

Assessment, crowdsourcing, data collection, fixed broadband, mobile network, QoS.

* 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/.

© ITU 2020

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

Introduction

Providers of telecommunication services usually need to assess QoS from their networks, as it is a way to verify if the service is being adequately delivered to their end-users. The assessment of QoS may be important from the regulator's perspective as well, as quality improvements are often a critical item on the regulatory agenda. The collected data on quality issues can serve regulators as a guide to raise consumer awareness and to foster further improvements in the network infrastructure. One of the most important steps in telecommunication networks quality management is the definition and implementation of the quality monitoring and assessment approach. There are many solutions available for this task, and all of them have their own trade-offs between accuracy, temporal and geographic granularity and costs.

The crowdsourced approach is one of the possible methods for monitoring and assessing QoS in fixed and mobile broadband networks. Its basic assumption is to collect data from a large base of end-users, directly from end-user's equipment (for example, mobile devices and customer premise equipment (CPEs)). There are, however, many vendors that provide these kind of solutions, and they may differ significantly from each other. Differences include how the data is collected (if initiated by the end-user or if no intervention is required) or what type of data is collected (if a download test is performed or if the data just comes from the device/equipment regular usage). Crowdsourced data that are collected on multiple service provider's networks in the same market using the same data collection methodology will benefit from consistent data collection approaches and methodology across service providers.

The crowdsourcing approach increases the number of data points significantly as compared to the typical QoS assessment approach (e.g., drive/walk test). The large amount of data may strengthen the reliability and representativeness of the obtained results. Furthermore, it improves the use of resources and allows countries with larger geographical masses to leverage on the public to collect the data.

This Recommendation identifies advantages, disadvantages and precautions that should be considered when deploying such methods of QoS monitoring.

Recommendation ITU-T E.812

Crowdsourcing approach for the assessment of end-to-end quality of service in fixed and mobile broadband networks

Amendment 1

Editorial note: This is a complete-text publication. Modifications introduced by this amendment are shown in revision marks relative to Recommendation ITU-T E.812 (2020).

1 Scope

This Recommendation covers the end-to end QoS assessment of fixed and mobile Internet access using the crowdsourcing approach, including:

– Overview of crowdsourcing approach for fixed and mobile Internet access;

– Types of crowdsourced data collection, characteristics and requirements;

– Set-up scenarios;

– Guidelines for regulators, service providers and vendors that can be used for benchmarking and network improvement when using the crowdsourcing approach.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this document, constitute provisions of this Recommendation. At the time of publication, the indicated editions indicated were valid. All Recommendations and other references are up for revision; users of these Recommendations 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.800] Recommendation ITU-T E.800 (2008), *Definitions of terms related to quality of service*.

[ITU-T E.806] Recommendation ITU-T E.806 (2019), *Measurement campaigns, monitoring systems and sampling methodologies to monitor the quality of service in mobile networks*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

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

3.1.2 latency [b-ITU-T G.9961]: A measure of the delay from the instant when the last bit of a frame has been transmitted through the assigned reference point of the transmitter protocol stack to the instant when a whole frame reaches the assigned reference point of receiver protocol stack. Mean and maximum latency estimations are assumed to be calculated on the 99th percentile of all latency measurements. If retransmission is enabled for a specific flow, latency also includes retransmission time.

3.1.3 jitter [b-ITU-T G.9961]: A measure of the latency variation above and below the mean latency value. The maximum jitter is defined as the maximum latency variation above and below the mean latency value.

3.1.4 IP packet loss ratio (IPLR) [b-ITU-T Y.1540]: IP packet loss ratio (IPLR) is the ratio of total lost IP packet outcomes to total transmitted IP packets in a population of interest.

3.1.5 crowdsourcing [b-ITU-T P.912]: Obtaining the needed service by a large group of people, most probably an on-line community.

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

3.1.7 quality of service [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.

3.1.8 crowdsourced data collection [ITU-T E.806]: A method to gather active and/or passive quality of service measurements from a large number of end-user devices.

3.2 Terms defined in this Recommendation

None.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

API Application Programming Interface

CDN Content Delivery Network

CPE Customer Premise Equipment

CPU Central Processing Unit

GNSS Global Navigation Satellite System

HTTPS Hypertext Transfer Protocol Secure

IoT Intert of Things

IP Internet Protocol

ISP Internet Service Provider

IXP Internet Exchange Point

KPI Key Performance Indicator

MCC Mobile Country Code

MNC Mobile Network Code

OTT Over-the-top

PoP Point of Presence

QoS Quality of Service

RAM Radom-access Memory

RF Radio Frequency

RoI Return on Investment

UDP User Datagram Protocol

5 Conventions

None.

6 Overview

[b-ITU-T P.912] defines in its Annex I the term crowdsourcing as "obtaining the needed service by a large group of people, most probably an on-line community." In this sense, it is implied that the main feature of crowdsourcing is to involve a large number of people. This concept can be applied to a varied range of purposes, like raising funds, sharing tasks or pooling results from opinion tests.

When it comes to the context of monitoring the QoS for fixed and mobile broadband networks, the crowdsourced approach can be understood as the gathering of data from a large base of end-users, with the goal of assessing the QoS of the employed network infrastructure, as a means to serve the consumers' needs. This can be done through the employment of software and hardware that collect QoS-related data directly from end-user's CPEs (such as routers) or mobile devices (such as tablets and smartphones). In the case of fixed broadband, embedded software in the consumer's premises equipment, which is usually supplied by the service provider, can collect QoS-related information. In the mobile service perspective, a similar set of data may be gathered, for instance, by software that can be downloaded on mobile devices.

A wide range of data can be collected from the CPEs and mobile devices (collectively known as "data collection device" in this Recommendation) when using crowdsourcing solutions that is useful for the QoS assessment of fixed and mobile broadband connection. They may include, but are not limited to:

– Location (e.g., GNSS coordinates),
– Date and Time (e.g., duration, etc.),
– Service provider (e.g., home and visited networks for mobile and internet service provider for fixed broadband),
– Network information (e.g., access point name, signal strength, cell information),
– Type of connection (e.g., Ethernet, Wi-Fi, 3G, 4G, 5G, etc.),
– Device information (e.g., manufacturer, model, etc.),
– Device usage (e.g., data usage, central processing unit (CPU) usage, battery level, concurrent traffic, etc.).

This collected data can provide relevant QoS indicators for broadband networks, such as throughput, latency, jitter, packet loss, among others. These are basic parameters that characterise the end-to-end performance and can be used to infer service quality.

7 Types of crowdsourced data collection, technical characteristics and requirements

Crowdsourced data collection is categorized into active and passive measurements, as introduced by [ITU-T E.806]. The data collection process can be initiated by the end-user or programmatically. Different configurations can be chosen to cater to the different needs.

Crowdsourced based solutions may include active and passive measurements. Hybrid solutions can use both.

7.1 Types of crowdsourced data collection

Each type of data collection method has its own sets of advantages and challenges which will be addressed in the following clauses. While a QoS indicator can be collected using both methods, the choice of method can reflect the same indicator differently. For example, the throughput collected from an active measurement may reflect closer to the performance of the network while the throughput collected from a passive measurement may reflect closer to the end-users' actual usage.

7.1.1 Active measurements

Active data collection methods create artificial traffic with the intention to assess end-to-end QoS parameters. For example, intentional file transfer with the aim to measure the throughput, ping tests, among others. The active approach is illustrated in Figure 1.

Active tests may be specifically designed to saturate the network, and therefore produce a more accurate measure representing the peak end-to-end performance at the point of test. However, attention must be directed towards the analysis of the results to ascertain its reliability and representativeness. To this end, clauses 7.3.1 and 7.3.2 outline considerations for the proper processing and statistical analysis of data.

When measuring the end-to-end performance using an active test, many factors can be controlled, such as, the test duration, the test server location, the number of simultaneous streams, the size of packets, the sampling rate or even the type of data sent (i.e., random data might not be compressed as effectively by network routers as text, images or video).

When measuring upstream end-to-end performance (i.e., upload), an active test can be designed to measure the amount of data received at the test server, not merely the amount of data sent by the device. When measuring packet loss, an active test can be designed to compare the number of packets sent to the number of packets received by the test server. When measuring the path used through the network (e.g., to detect differences between download and upload paths), only an active test can generate adequate packets, receive responses from and build up the path utilized (i.e., using an approach such as traceroute).

Care should be taken to identify scenarios where an active test does not represent end-to-end performance. In the following scenarios a test is likely to underestimate the performance:

– Where another process is using data. For example, if the user is streaming video while a test is running, the active test will be constrained in the throughput it measures.

– Where the device operating system denies network access. Generally, this is done to preserve battery life. In such cases, an active test may fail to send any data, but this does not indicate any failure on the part of the cellular network.

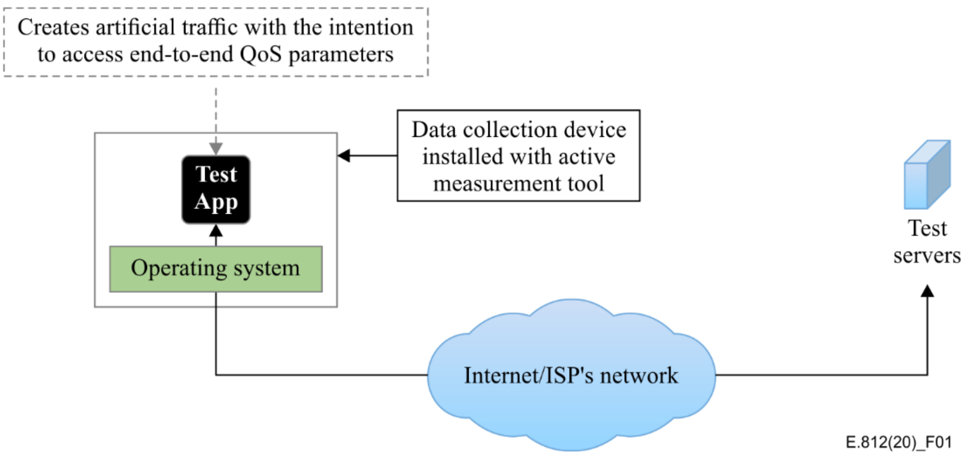

Figure 1 – Example of active data collection on mobile network

a) Benefits

Some of the benefits that can be highlighted in active methods are:

– Potential standardization of active tests.

– Active tests can be performed at application-level (e.g., in the form of an app, embedded tool, web browser).

– Active tests can be configured to emulate behaviour of different services[1] which can influence the different QoS parameters (e.g., latency, packet loss).

b) Challenges

On the other hand, some of the identified challenges are:

– Active tests utilize resources on the network (e.g., adding load to a heavily utilised network when attempting to achieve the peak end-to-end performance) and data collection devices (e.g., end-users' data quota, battery, radom-access memory (RAM), etc.).

– Potential increase in the data usage for higher speed networks, depending on the measurement design.

– Measurements designed to reflect peak throughput will be inhibited if other processes on the device are also consuming resources such as network connection, CPU and RAM.

To address the challenges, the following techniques can be considered:

– Zero-rating the data used in tests that consume data from a users' data plan would address the end-user concerns around data usage but should be used with caution. The process involves providing network operators with details of the solution's characteristics such as test servers or IP address information, which might lead operators to game the results by focusing efforts on improving the QoS parameters of the test route. However, statistical tools can be used to analyse the representativeness and reliability of the results. Clause 7.3.1 discusses the available tools and highlights considerations to perform these analysis.

– Measurement design choices can mean tests do not consume more data on faster networks. For example, tests that emulate specific services – such as streaming video (using a bandwidth of 1-1.5 Mbit/s), gaming or voice traffic (typically <100 Kbit/s) – may be designed to have a fixed data consumption. As an illustration, a test might send a stream of 200 packets to a server at a fixed bitrate and receive packets echoed back. Implemented correctly this would allow measurement of latency, packet loss, jitter and burst rate. However, the data consumed per test would remain the same no matter the network speed.

– Relying on a large pool of end-users, each running measurements relatively infrequently. The solution needs to strike a balance between collecting a useful number of samples per device, while avoiding consuming too many resources and inhibiting the size of the crowd that can be achieved. To allow flexibility, the measurement sampling strategy can be designed to be centrally configurable, allowing tuning of this balance.

– Selection of strategic timing of measurements for peak throughput, for example running brief automated throughput tests immediately after the screen is unlocked before end-users have time to start data intensive activities.

– Adopt a hybrid approach (i.e., active and passive methods) to have a holistic view of end-to-end QoS.

[1] Example: VOIP and online gaming traffic typically involve the transmission of small UDP packets at fairly constant and low bandwidth; downloading email attachments, system updates and other files typically takes place over TCP and may use much larger packets and much higher bandwidth, possibly only constrained by the network.

7.1.2 Passive measurements

Passive data measurements do not inject artificial traffic or test payload into the network for QoS assessment. Instead, they act more as an observer of radio parameters, the end-user's transferred data, and collect information regarding the actual traffic. This approach is illustrated in Figure 2.

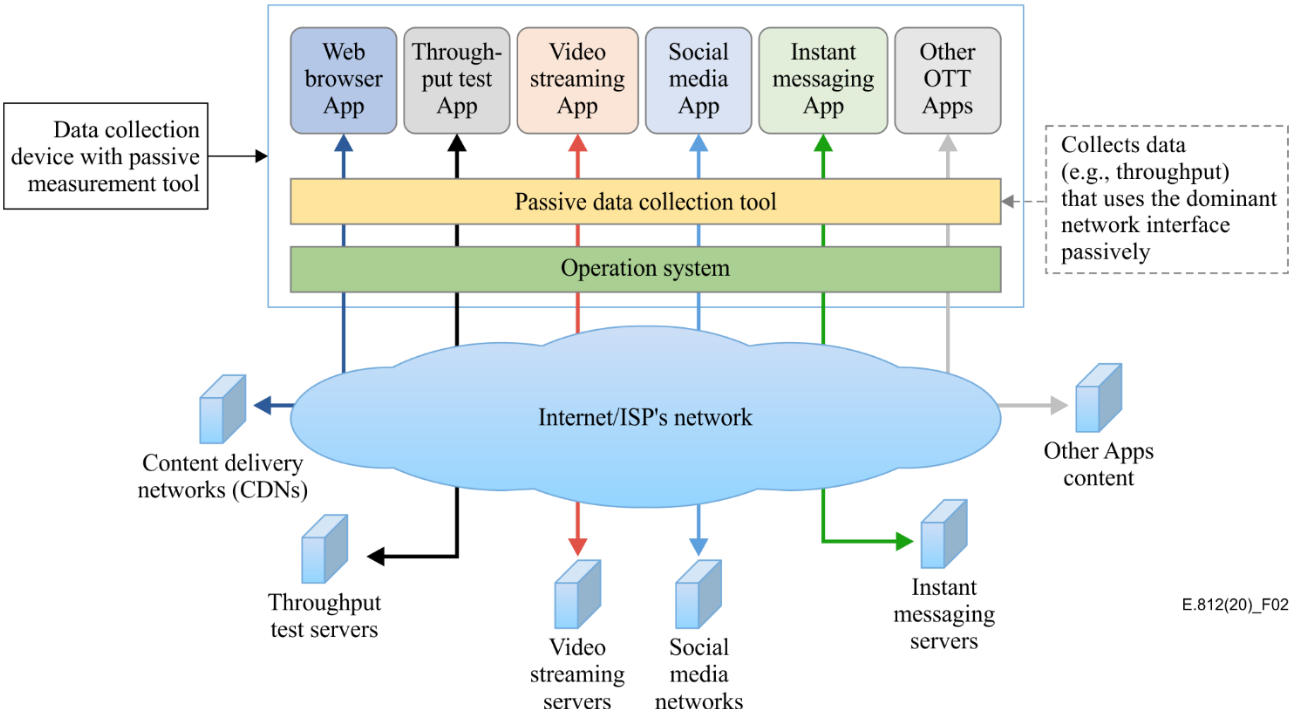

Figure 2 – Example of passive data collection on mobile network

Passive data measurements may be specifically designed not to be aware of the applications generating the traffic due to privacy concerns, thus, it is difficult to accurately tell whether the measurements made are a result of the application, network or even the end user. Therefore, passive data collection produces measurements representing the end-to-end performance experienced by users under actual users' usage at that point in time.

For instance, when measuring end-to-end performance passively, a passive test cannot tell whether the application generating the traffic is doing so at its maximum rate. The application could be sending or receiving the data only when it needs it (e.g., data might be cached locally), it could be doing computations that delay a network request, or the user's actions could affect the traffic (e.g., a user could have paused a video stream).

As such, when interpreting the results of passive tests, additional information such as the amount of data transferred, condition of the network (e.g., signal strength in the case of mobile networks) and the aggregation of the crowdsourced data collected can be used to more accurately determine the end-to-end performance experienced by users.

a) Benefits

Some of the benefits that can be highlighted in passive methods are:

– The running of measurements does not consume additional data which means more samples may be collected per end-user.

– It is not necessary to maintain a test server for artificial traffic transfer.

– It does not generate artificial traffic on network, thus it is an efficient method from a traffic point of view.

b) Challenges

On the other hand, some identified challenges are:

– Different platforms (for example, operating systems, hardware, etc.) provide different capabilities for passive measurements, thus making it difficult to compare across platforms.
– Passive measurements are not able to monitor certain QoS indicator (e.g., latency or packet loss or consistent maximum throughput between the end-user and server).
– If application-specific data is collected, the method of collection becomes intrusive.

To address the challenges, the following techniques can be considered:

– While not all platforms allow passive measurements, it is necessary to evaluate if excluding these platforms affect the QoS indicators results significantly.
– Adopt a hybrid approach (i.e., active and passive methods) to have a holistic view of end-to-end QoS.
– Ensure that end-users are informed about the types of data that will be collected and agree to participate in the crowdsourcing campaign.

7.2 Methods to initiate crowdsourced data collection

In cases where the end-user is required to deliberately start data collection, the tests are classified as end-user initiated. On the other hand, if data collection occurs programmatically by pre-established start rules, it is classified as automated.

7.2.1 End-user initiated measurements

This approach requires the end-user to initiate the test and the initiated test results are typically provided to the end-users.

a) Benefits

This type of crowdsourced approach brings the following benefits:

– End-users initiate the test which means that they agree for the data collection during the test period, making it more transparent.
– End-users' initiation provides end-users with the option to carry out more tests and to maintain data collection for longer periods under consent, which may lead to more collected information about the end-to-end performance and measurement conditions (such as device location and utilization).
– End-users are given the control on the circumstances to perform the tests (i.e., end-user's location and network condition).
– End-users can perform assessment of the networks' QoS and obtain the test results after the test, raising the end-users' awareness of the network's condition during that period of time.
– End-user initiated tests allow introducing survey questions which can provide additional information on end user's internet connection or motivation for the testing.

b) Challenges

The dependency of end-users' action to initiate the test may impose the following challenges:

– As it relies on end-users' interaction, the number of collected samples may be much lower.
– Some bias may be introduced as there may be specific group(s) of end-users interested in QoS assessment results. For example, end-users may tend to perform data collection when they experience difficulties to access some applications. Another possible bias is related to periods of day in which collection occur, as there is no scheduling.
– It may be easier for malicious parties to influence the collected data than with automated tests since the end-user has greater control over the conditions under which the measurement happens.

– When using survey questions, thought should be given to the placing of the survey and whether it is optional or a condition for running the test. Placing questions after the test may lead to answers being influenced by the test result, while placing before may deter testing or encourage random answers – particularly if the questions cannot be dismissed.

7.2.2 Automated measurements

Automated data collection can be performed without the need of end-user's intervention either as an independent application or by embedded solutions within other applications or hardware. The tests could be scheduled to run regularly or triggered based on certain algorithms or specified rules.

a) Benefits

Some of the benefits of this approaches are:

– It is possible to determine time periods and locations for data collection. This may result in a larger number of samples, which may facilitate achieving statistical validation.

b) Challenges

Automated data collection may impose the following challenges:

– End-users may not be fully aware that the data collection is on-going.

– Not all platforms allow automated background measurements.

Tests initiated autonomously provide greater control to the entity that designs the tests, making it possible to determine the frequency of tests and the geographic area where they shall take place, something that is out of reach for set-ups that depend entirely on end-user initiated tests.

Table 1 provides a summary of clauses 7.1 and 7.2.

Table 1 – Benefits and challenges of the different types and methods of initiation of crowdsourced data collection

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Methods to initiate crowdsourced data collection</td></tr>
<tr><td>End-user initiated</td><td>Automated</td></tr>
<tr><td rowspan="2">Types of crowdsourced data collection</td><td rowspan="2">Active</td><td><u>Benefits</u>
– Perceived to be more transparent
– Raise end-user awareness
– Potential standardization of active tests
– Multi-platforms
– Configurable to emulate services' behaviour
– May be designed to provide an estimate of the end-to-end performance during the test period</td><td><u>Benefits</u>
– May have larger number of samples
– Potential standardization of active tests
– Multi-platforms
– Configurable to emulate services' behaviour
– May be designed to provide an estimate of the end-to-end performance during the test period</td></tr>
<tr><td><u>Challenges</u>
– May have smaller number of samples
– May introduce biasness
– Utilize additional resources
– Potential increase in data usage
– Results may be affected by data collection devices' condition</td><td><u>Challenges</u>
– Perceived to be less transparent
– In some situations it is not supported by all platforms
– Utilize additional resources
– Potential increase in data usage
– Results may be affected by data collection devices' condition</td></tr>
</table>

Table 1 – Benefits and challenges of the different types and methods of initiation of crowdsourced data collection

		Methods to initiate crowdsourced data collection	
		End-user initiated	**Automated**
	Passive	**Benefits** – Perceived to be more transparent – Raise end-user awareness – May have larger number of samples – Does not require test server – Does not further congest the network – Provides an indication of the end-to-end performance based on end-users' actual usage	**Benefits** – May have larger number of samples – Does not require test server – Does not further congest the network – Provides an indication of the end-to-end performance based on end-users' actual usage
		Challenges – May have smaller number of samples – May introduce biasness – In some situations, it is not supported by all platforms – Limited QoS indicators to be monitored – May be intrusive	**Challenges** – Perceived to be less transparent – In some situations, it is not supported by all platforms – Limited QoS indicators to be monitored – May be intrusive

7.3 Crowdsourced data collection requirements

This clause provides requirements for the crowdsourcing solutions such as homologation procedures, sampling and scheduling methodologies and data processing rules.

7.3.1 Sampling and scheduling

It is possible to collect key performance indicator (KPI) data from a wide set of data collection devices, depending on the capacity of the test servers, storage and scheduling.

In order to provide statistical validation, it is necessary to calculate the minimum number of samples by considering a pre-defined maximum acceptable error; provided the population and probability distribution associated (for example, Gaussian for median estimates or Binomial for proportion estimates). References on how to obtain the sample size and distribute the samples geographically can be found in Annex A of [ITU-T E.806].

In addition to the minimum number of samples, a sampling plan must consider that biased results can be safely avoided. Thus, the entity responsible for the data collection is recommended to verify if the geographical origin of the samples is statistically consistent with the spread of the targeted population, representing the actual QoS delivered to end-users.

However, if necessary, the geographic and/or temporal sampling distribution can also be considered in the statistical methodology. For example, depending on the goal of the campaign, it may make sense to limit the sampling to periods of high traffic, as well as to collect a larger number of samples in geographic areas with high end-user density. Thus, it might be more suitable to define the sampling plan for each operator considering end-user base distribution. In other cases, it may be considered to collect a larger amount of samples in periods of low traffic or regions with low density to provide larger sample sizes compared to typically overrepresented sample groups.

In summary, it is recommended to establish a minimum number of end-users to be periodically and randomly monitored. The more end-users included in the data collection plan according to the temporal and geographical distribution, the more precise the evaluation and formulation of reports.

Nevertheless, as a consequence of the uncontrolled testing environment, crowdsourced collected data must be post-processed in order to discard noisy test results. Such results occur, for example, due to the effect of busy hour testing schedules, end-users' changes of contracted plan, commercial plans with restrictions, changes in access technology, among others. In case of such changes, the remaining set of samples after the application of discarding rules must comply with the statistical validation requirements (number of samples, estimation error and geographical/temporal distribution).

The sampling approach of computing the sample size (minimum number of samples) for a pre-defined estimation error can be used when there is certain control of the crowdsourced data collection. For example, in a crowdsourcing test campaign, the geographical distribution of the samples can be known to a certain extent if it was planned in advance.

However, in other cases when there is limited control over the geographical/temporal distribution, results may not be representative and may have high bias in the estimated KPI. For example, when looking at the results of an entire city, it might happen that the majority of samples are coming from a single neighbourhood. In this case, results may be representative of that neighbourhood but not the entire city.

When planning a testing campaign, one must consider the statistical representativeness of the results. To this end, there are well-defined statistical formulas that can be used to derive the estimation error of a group of samples. Clause A.1.2 of Annex A of [ITU-T E.806] presents references for the general formulas of the simple random sampling which can be used to obtain the bound on the error of estimation for a given sample size (number of samples) confidence level and considering the mean and standard deviation of the parameter of observation (download/upload speed, latency, etc.). However, to get an insight into how representative is a group of samples of a geographical region, one must calculate the estimation error of the spatial distribution of the samples. Stratification sampling provides the necessary formulas to estimate the geographical representativeness (see clause A.1.2 of Annex A of [ITU-T E.806] and [b-Scheaffer] for the formulas of stratification sampling).

In stratification, the geographical region (for example, a city) under study has to be divided into strata or groups of geographical areas. Each group will have different characteristics, and one can infer how representative a group of areas (with crowdsourced data samples) is of an entire region.

To achieve this, a binning technique can be applied to divide a particular region into smaller areas. The size of the bins can be chosen depending on the size of the region under study (tens or hundreds of meters, km). Once the binning takes place, they have to be classified into one of the strata. Finally, the formulas of stratification sampling can be used to estimate the minimum number of bins need to have statistical representativeness of the region. Appendix I presents a numerical example to illustrate this stratification approach.

By using the aforementioned techniques, one can estimate the minimum number of bins (stratification sampling) with a required sample size (simple random sampling) that ensures a certain estimation error. However, when crowdsourced data is coming from user-initiated tests or passive data collection, there is no possibility to plan in advance, thus resulting in bins with no and/or too few measurements (below the 95% confidence level) and/or in bins with highly concentrated measurements. Therefore, statistical tools are needed to compute the estimation error from a group of samples when no planning/scheduling is possible.

In this case, the opposite can also be obtained from the formulas; one can calculate the estimation error from the number of bins with samples in a specific city. An approach could be to assign different weights to the under-populated/over-populated bins so that the results will account for the different number of measurements each bin could have.

In summary, given the challenges associated with the geographical distribution of crowdsourced samples, if results will be aggregated for a city/region, a statistical analysis has to be carried out to understand the reliability of the samples in terms of the associated estimation error and reliability level, for example.

7.3.2 Data processing

In this clause the considerations for the proper handling of data to produce reliable analysis are outlined.

Datasets will differ according to how the data is collected. The objectives of analyses will also differ. For these reasons no single procedure can be proscribed. However, it is useful to look at the kinds of processing steps and some examples of each.

The following are some classes of processing steps that were identified:

– Filtering
– Categorization
– Aggregation

A typical processing method might consist of a set of filtering steps, followed by a set of categorization steps and then a final aggregation; however, steps of different classes can also interleave.

7.3.2.1 Filtering

Filtering processes may be used to remove redundant, untrustworthy or irrelevant data that should be immediately excluded from analysis. Examples include:

– Duplicated data
– Data not collected during the sample period being examined, or that which does not have a valid time reading
– Data that is not within the geographical bounds being examined, or that which does not have a location reading

Depending on the use case further filters may be required. For example, for active tests it may be advisable to remove tests where the device operating system or device user preferences would be likely to be preventing or limiting network access – for example when a power saving mode, or airplane mode are active; or where the user has manually disabled data.

7.3.2.1.1 Filtering anomalous testing behaviour

A further filtering objective merits special discussion here. As mentioned in clause 7.2.1 malicious parties may attempt to influence the results, for example deliberately running tests in spatiotemporal locations they know will produce favourable test results, or others they know to be poor. By identifying devices with unusual testing patterns, such as large numbers of tests, or producing atypical results this data may be isolated.

While automated techniques can be brought to bear here, the attempts to influence data may evolve to avoid the filters requiring adaptive analysis. In particular, where there is frequent and granular public reporting on the crowdsourced results, it may be possible for the users to tell whether, and the degree to which, their attempts at influencing results are successful. If the user sees their attempts are being filtered out, they may evolve their strategy. For long or ongoing crowdsourcing campaigns with public reporting, it is realistic to expect that measures will need to be continually developed as counter-measures are detected.

Anomalous testing behaviour is not always designed to influence aggregate results. Users may have personal objectives for using crowdsourced apps that lead to them producing unusual patterns of data. As an example: a user might want to provide evidence to release them from a contract by picking places in their home where data is very slow, running tests and taking screen shots. As another extreme, they may wish to share on social networks evidence of just how good their speeds are, leading them to favour testing in places where they have a better experience of the network.

By surveying users about their motives for testing within the crowdsourcing agent (see clause 7.2.1), it may be possible to filter out such cases. It is suggested that such survey be simple, short and optional, to encourage high participation (positioned such that many users will see and answer: "start the test by telling us why you are testing today") and prevent a reduction in the number of tests, which would be expected if the survey were made a pre-requisite of testing. Even in the case that most users do not answer this question, if sufficient responses are gathered these can be used as a training set to identify user motivations in general.

7.3.2.2 Categorization

Categorization is where the data is enhanced or corrected to prepare it for further processing. Potentially useful categorization steps include:

– **Network ID mapping**: mobile networks are sometimes able to broadcast different network names or operate with multiple mobile network codes (MNCs). In this case it may help to map them to a canonical identifier and name. Note that particular care should be paid to cases where the mobile country code (MCC) of the SIM card represents a different country to the country represented by the MCC of the connected network, depending on the analysis being performed it may be appropriate to remove these international roaming results.

– **Measurement time correction**: since the time can be manually adjusted in end user devices, the device reported time may not be reliable. Corrections can be applied where a device's account of the time data that was sent significantly differs from the collection server's authoritative value.

– **Geocoding**: location coordinates in the form of latitude and longitude may be assigned location IDs, this could be using a hierarchical spatial index, or a specific set of polygons representing, for example, cities and provinces.

7.3.2.3 Aggregation

The aggregation steps take in a set of individual test results. Usually these will have been filtered, categorized and they are used to produce final summary statistics.

The aggregation chosen will depend strongly on the analysis. Factors that should be considered:

– Should results be first aggregated to device-user level; and then an average taken across these? This "one-user, one-vote" strategy will tend to prevent results being overly influenced by user-devices that report more data samples.

– Is re-weighting of the data necessary at any other level – geographically, or in time?

On this second point, Appendix I provides an example of how geographic stratification sampling might be applied in the context of mobile service.

Note that all entities involved in the data collection and data processing, i.e., vendors, operators and regulators, must comply with the corresponding data protection legislation. This is relevant to data processing because often aggregation will require unique-user device identifiers and in some legal frameworks the presence of such information alongside location data constitutes personal data.

8 Set-up scenarios

As previously mentioned, the crowdsourced approach can be classified as active or passive, and as end-user initiated or automatic data collection. Besides the crowdsource solution approach, there are different configuration options for implementing QoS data collection based on a crowdsourcing approach. In this sense, two major questions regarding the set-up scenario must be defined: (i) the device in which the crowdsource solution will be embedded and (ii) the network point in which the test server will be allocated. It is important to highlight that passive crowdsourced data collection solutions do not require a test server, since no artificial traffic or test payload is generated. Tests that are initiated by the end-user or autonomously have specific features that also have to be taken into account in the definition of the set-up scenario.

These definitions should be made considering trade-offs related to the goals of QoS assessment. In this clause, the characteristics of different set-up scenarios for crowdsourced data collection in fixed and mobile broadband networks are further analysed.

8.1 Data collection device

Data collection in crowdsourcing approach may occur in different types of devices allocated in different elements of the service providing process which is explained in the end-to-end QoS model introduced in [ITU-T E.800], shown in Figure 3.

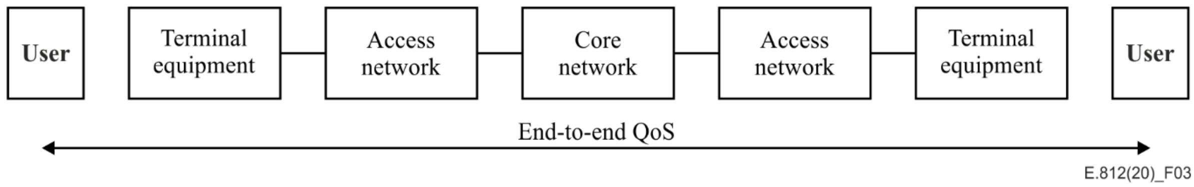

Figure 3 – End-to-end QoS assessment model [ITU-T E.800]

NOTE – [ITU-T E.800] states that the configuration illustrated above is for the conventional service with users at each end of a connection. In this Recommendation, the term end-user is used for the user indication showed in Figure 3.

Thus, when assessing end-to-end QoS, the data collection solution must be allocated in end-user's terminal equipment.

For fixed broadband networks, that means crowdsourcing solution is embedded in customer's premises equipment (CPE) or in end-user's device (such as mobile devices, personal computers, smart televisions, etc.). For mobile networks, the solution is allocated in mobile devices. Solutions in devices bring challenges such as environment variables, like the strength of the Wi-Fi signal where the test is performed, which have an impact on the collected results.

Solutions allocated in mobile devices can be, for example, apps dedicated to quality assessment, which are generally provided by specialized vendors. On the other hand, the solution can be embedded in a specific application, such as the operator's customer services app (when the operator builds its own solution) or third-party applications, such as social media, games and photo editors.

Solutions allocated to personal computers can be web pages to be accessed by end-users for testing, or background applications that run on web browsers or specific applications.

Specifically, for fixed broadband scenarios, when the crowdsourcing solution is located in the end-users' devices, the assessed QoS is related to that device, thus, there are limitations by the technical characteristics of the device and limitations in connection technology (such as Wi-Fi or Ethernet). On the other hand, if the solution is embedded directly in CPE, the QoS information reflects the provided service in an overall perspective.

8.2 Test server allocation

The servers used for hosting active tests have a strong influence on the results and thus should be selected with care. A good starting point is to consider the types of server hosting and a non-exhaustive summary of widely used server hosting listed in Table 2.

Table 2 – Comparison of test server hosting options

Type of server hosting	Characteristics	Typical services
Network point-of-presence (PoP)	Servers are located in the network infrastructure of service providers or connected entity (e.g., universities, private enterprises, or government agencies), meant to be close to the end-user. Test servers hosted here are designed to test the performance of the access network.	Hosting of: – Files – Websites
Content delivery network (CDN)	Content resources are typically distributed across hundreds or thousands of nodes across the Internet and designed to serve content. CDN is a form of server-less computing and may have several tiers, with frequently accessed content being cached in many edge nodes. Some CDNs allow the creation of certain applications, which can be used to perform tests.	Distribution of: – Video and audio on demand – Mobile application binaries – Website content (image, text, JavaScript files, etc.)
Cloud computing services	Computing resources are typically distributed across a few dozen locations across the Internet and designed to support a broad range of services that requires scalable computing resources. This type of server hosting also includes edge computing models which introduces smaller servers at a much greater number of locations typically for applications that require lower latency.	Hosting of: – Cloud-based applications – Databases – Game engines – Video OTT calls, group voice OTT calls – IoT services (Edge computing) – Cloud gaming (Edge Computing)

Additionally, the connectivity from the end-user to the test server(s) also influences the characteristics measured. Test servers can be allocated in operator's network (on-net) or externally to that network (off-net). A comparison of the types of server access is listed in Table 3.

Table 3 – Comparison of the types of server access

Type of server access	Types of server hosting	Characteristics
On-net	– Service provider of the end-user – CDN caching servers – Edge compute caching servers of cloud computing services	Useful to qualify metrics of the access network, or of popular or frequently used services
Off-net	– Connected entity (e.g., universities, private enterprises, or government agencies) – Service providers other than the provider of the end-user – CDN(s) – Cloud computing services	Useful to qualify metrics of all types of network services and locations, whether popular or not.

The profile of network and data services accessed by an end-user on a given network will inform the proportion by which the measurements to on-net or off-net servers should be used. Regulators may select network data profiles consistent with their needs. In cases where there are multiple test servers, there could be options to automatically or manually select the test server. When automatically selecting test server(s), one of the options is to select server(s) close to the end-user, which will ensure that fewer network segments are included, thus more accurately measuring the access network. On the other hand, manual server selection allows the measurement of access to a given entity.

It is apparent that the server distributions for a representative measurement of quality of service vary according to the service of interest. As mentioned in previous clauses, active tests may be designed that are specific to different services, with respect to servers this might mean that a video on demand test would stream data from a CDN used by popular video services whereas a multiplayer gaming test would communicate with a cloud server commonly used by game engines.

When considering CDNs as test servers, two further points should be noted:

– Popularly accessed content may be cached closer to end-users than less popular content. Thus, an active test may initially download a file cached on only a few CDN nodes, but once the test has been run many times the file may be distributed to more nodes and begin to download faster. By allowing a warming up period for any new crowdsourcing solution these change points can be avoided.

– Dynamic Internet protocol (IP) address allocation and hypertext transfer protocol secure (HTTPS) tests are more robust against mobile network operators unfairly treating test traffic. As an illustration, if a test that streams video uses a static IP known to the network operators, an operator that throttles video traffic could remove throttling just for that test, using HTTPS and static IPs make this much harder.

To measure end-to-end QoS, it may not be sufficient to test against servers hosted in some neutral location (for example collocated with Internet exchange points (IXPs)) as most of the network traffic may not traverse that route. This approach of measuring the end-to-end QoS may introduce factors beyond the direct control of mobile network operators. In the example of the previous paragraph, the closest servers used for popular online mobile games may be in another country to the one under study, thus the quality of the international links (and the distance they cover) will contribute to the measured QoS.

To conclude this clause, given the large variety of services accessed by end-users and the corresponding diversity in the endpoints of these services, it is advisable to focus on a few core services, particularly ones which are popular and demand good end-to-end performance. For these services, it is possible to design measurements specifically to capture their end-to-end QoS. This involves picking a server distribution (or set of server distributions) that closely matches that used by the real traffic to that service.

9 Additional guidelines for regulators

This clause identifies additional guidelines for regulators seeking to use crowdsourcing-based QoS measurements.

A regulator may perform its own data collection, require the operators to supply the data or use third party solution(s) according to the guidelines laid out in this Recommendation.

In order to allow interpretation and comparability of the results, a proper definition and the use of a common methodology is essential.

To guarantee the neutrality and comparability of the results, regulators are advised to:

a) Define regulations to ensure transparency in terms of describing the chosen sampling methodology, QoS indicators and their calculation, and data processing rules;

b) Ensure that there is no interference on data collection results, for example by setting up crowdsourcing tools to send the data, without filters, preferably in real-time, to a central platform, and monitoring any testing patterns that indicate attempts to manipulate the use of the crowdsourced tool;

c) Perform audit processes to check compliance with the requirements and established testing methodology;

d) Require the providers of the data to inform the regulator about changes in the methodology (including but not limited to set-ups, sampling, and data collection) for validation.

In the case of using crowdsourcing solutions implemented by operators, to ensure comparability of results, a regulator may decide to adopt an approval or auditing process, which may include solution validation of the chosen set-up scenarios.

Validation procedures may be used to ensure reliable comparisons of results between operators and/or regions. In this sense, the objective of validation is to evaluate the data collection solution and its ability to generate samples with a degree of accuracy and precision with respect to the established requirements such as technical standards, sampling, data processing and testing methodology for data collection.

The regulator may request that the service provider or test vendor provide an operational manual containing the testing methodology used, test set-up scenarios, technical specifications of the collected data or measured metrics, and the data processing methods used.

If there is information that cannot be obtained from the regulator's own data collection or information that could expire, regulators may require the operators to supply it on a regular basis.

NOTE – Information that is not available in the CPE may include, in the case of fixed broadband, some contracted information of the end-users (such parameters can include the maximum, minimum or guaranteed speed), and the models of devices supporting the technical standards required to perform the measurements.

Appendix I

Numerical example of the stratification sampling approach for the mobile service

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

In this appendix, a numerical example is developed to illustrate the sampling formulas following the methodology described in the Annex A of [ITU-T E.806] to estimate the geographical representativeness of crowdsourced data for the mobile service.

Suppose one has to publish results on average download/upload speed on a per city basis, this means that crowdsourced data will be used to derive the metrics. With crowdsourced data from the mobile service, it is important to talk about the spatial estimation error of the samples. In other words, how representative those samples are of a geographical region. Since there is no control of the geographical origin of the samples, one can estimate the statistical spatial representativeness.

To achieve this estimation, each city is first divided in bins of 100 m X 100 m (the size of the bins can vary and should be determined depending on the size of the cities, etc). To produce the bins, there are several approaches to be used including Geohash.

To be able to derive the degree of representativeness of the samples from a geographical/spatial perspective, each of the bins can be stratifyed into different categories (strata). An important aspect in stratification is that the elements in each category will share characteristics among them and can be different from the other categories.

So, in this example, the first step in the stratification sampling is to choose a criteria in order to build strata and then classify the bins according to this criteria. The chosen criteria is the population distribution. However it could be another relevant variable in the study such as user density, device density, etc.

The following is one example using the population density criteria to define the strata:

Number of Strata	**Range**	**Number of bins**
1	Population density <= 2 000	193 760
2	2 000 < population density <= 4 000	260 766
3	4 000 < population density <= 6 000	213 436
4	Population density <= 8 000	306 937
		974 8990

Once the bins are classified, a bound of an error of estimation (*B*) using the corresponding descriptive variables of each stratum (media, standard deviation and variance of download speed, for instance)can be defined.

Finally, the total number of bins that represents, with a certain estimation error, the area of study (in this case, a city) can be obtined using the following stratification sampling formula (described in clause A.1.1 of [ITU-T E.806]):

$$n = \frac{\left(\sum_{i=1}^{L} N_i \sigma_i\right)^2}{N^2 D + \sum_{i=1}^{L} N_i \sigma_i^2} \qquad D = \frac{B^2}{4}$$

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).

Strata	**Range**	**Number of bins**	**Media**	**Standard deviation**	**Variance**
1	Population density <= 2 000	193 760	5.60	0.66	0.43
2	2 000 < population Density <= 4 000	260 766	5.67	0.43	0.18
3	4 000 < population Density <= 6 000	213 436	5.23	0.56	0.31
4	Population density <= 8 000	306 937	5.32	0.60	0.36
		974 899			

The next step, after calculating the total number of bins necessary to obtain statistical significance of a city is to obtain the minimum number of bins to represent each stratum within the city.

As an example, the bound on the estimation error B is chosen as 0.1. This variable can also be obtained from previous results, or it can be fixed depending on the magnitude of error that is acceptable. One important thing to keep in mind is that the formula is highly sensitive to the estimation error, thus a small B implies a larger sample size. Multiple scenarios can also be studied by setting different values for B.

By applying the stratification formula, with D = 0.0025, it shows that n =124 bins (total number of bins) that is then distributed to each stratum (n_i is the number of bins necessary in each stratum) using the second formula of the stratification sampling described in clause A.1.1 of Annex A in [ITU-T E.806]:

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

Number of Strata	**Range**	$\mathbf{n_i}$
1	Population density <= 2 000	29
2	2 000 < population density <= 4 000	25
3	4 000 < population density <= 6 000	27
4	Population density <= 8 000	42
		124

In this way, each city can be characterized with at least n bins (distributed according to n_i). The results of download speed (or other variable under study) will have an estimation error B. Of course, the opposite can also be obtained from the formulas. Since the crowdsourced data will be available beforehand, one can calculate the estimation error from the number of bins with samples in a specific city.

It is important to mention that each bin can also be considered valid if it has a minimum number of samples. In this case, simple random sampling can be used to derive this sample size with an associated estimation error.

When planning a test campaign, the following three methods for calculating the strata sample size could be used: proportional allocation, optimal allocation and mixed allocation.

a) Proportional allocation:

With an allocation proportional to the number of units, the sampling rates are the same for all strata:

$$f_h = \frac{n_h}{n} = \frac{N_h}{N}$$

In other words, the larger the stratum, the larger the sample selected in this stratum.

This allocation therefore leads to a self-weighted sampling plan where all individuals have the same weight $w_k = \frac{n}{N}$.

This ensures robustness of the results when analyzing several variables simultaneously. With reference to the variances, the simple random sampling stratified with proportional allocation is more effective than the simple random sampling.

b) Optimal allocation:

Optimal allocation or Neyman's allocation optimizes the precision of the estimator of the total of this variable of interest at the level of the whole population.

Neyman's allocation assumes that a larger sample must be selected in a large strata and in a strata with high dispersion.

Let C the cost of the survey, where:

$$C = \sum_{h=1}^{H} n_h \times c_h$$

Where n_h the size of strata h and c_h is the cost of unit in the strata h

Then:

$$n_h = \frac{N_h \times S_h}{\sqrt{c_h}} \cdot \frac{C}{\sum_{h=1}^{H} N_h . S_h . \sqrt{c_h}}$$

Where N_h is the population size of the strata h and S_h is the variance of the strate.

c) Mixed allocation:

This allowance makes it possible to combine the benefits of the two previous methods therefore guarantee robustness at low cost.

With:

$$n = \alpha n_{prop} + (1 - \alpha) n_{opti}$$

Usually $\alpha = 1/2$

Appendix II

Use cases for the crowdsourcing approach

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

II.1 Examples of possible use cases

There are many possibilities to employ the crowdsourced data collection approach in the context of end-to-end QoS measurements. This includes but is not limited to the ones shown in Table II.1.

Table II.1 – Crowdsourcing approach scopes of use cases

Scopes	Type of crowdsourced data collection	Fixed/Mobile networks
Network coverage	Passive	Mobile
Performance monitoring and benchmarking	Active and passive	Fixed and mobile
Complaints verification	Active	Fixed and mobile
Checking commitment to license	Active	Fixed and mobile
Network planning	Active and passive	Fixed and mobile
Network optimization		

Generally, if needed, the deployment for the above scopes can be made in combination with other available network informational layers (e.g., frequency bands, base station locations, etc.). The below use cases are part of the possible identified uses of the crowdsourced data collection approach and also draw information from [b-CrowdWhitepaper]:

II.1.1 Mobile network coverage

The following use cases aim at identifying available mobile network coverage in both outdoor and indoor environments:

a) Outdoor

1) With sufficient data points (determined by crowd participation), it can provide coverage maps at fine resolutions per mobile operator. This can provide a detailed outlook for network coverage per operator and avoid costly drive tests in a time and efficient manner.

2) If representative data is available, it can be used for complaint validation and to identify problematic zones with respect to end-to-end QoS.

3) Indicator of the user distribution for geographical zones of interest. Since the crowdsourced data for a specific zone will have a sample distribution with more data points from operators with more users, it can be used as an indicator of how the users are distributed among the network operators.

b) Indoor

1) Indoor complaint validation.

2) Possibility to pin and identify poor indoor coverage through signal level indicators, although signal level issues will likewise affect location accuracy, or may prevent coverage completely (lower floors of underground garages).

3) Access to vital restricted buildings such as hospitals, public institutions.

4) Indicator of the user distribution for critical public zones and commercial districts. Since the crowdsourced data for a specific zone will have a sample distribution with more data points from operators with more users, it can be used as an indicator of how the users are distributed among the network operators.

II.1.2 Performance monitoring and benchmarking

a) Through KPI monitoring (both active or passive) or user interaction (e.g., user feedback after a session of a service), the crowdsourced data can be used to evaluate performance trends at different geographical levels (municipalities, cities, regions, etc.).

b) Detailed (both in time and space) crowdsourced data concerning access to the service and causes for service release can provide a good identification of the possible root cause for performance issues experienced by end users of a given version of a service at a given time and place and help finding a quick fix.

c) Combined with user interaction such as surveys and other measurement result sources such as drive tests, it can give a more reliable comparison between network operators or service providers in terms of QoS compared to only using crowdsourcing data.

II.1.3 Complaints verification

a) Empower end-users and respond to their complaints about slow Internet performance, poor or no mobile coverage (especially in low population areas), impossibility to make calls, dropped calls, no mobile Internet (despite having coverage). All can be investigated further, and possibly validated using the results from crowdsourcing data.

b) After confirmation of the issues related to a complaint, crowdsourced data can also be useful for qualifying it and proposing available solutions.

II.1.4 Checking commitment to license

Mobile coverage obligations for an operator may include rollout commitments such as a percentage of the population or territory covered in a timeframe. Such obligations can be derived from the operator's license or through a spectrum auction. With sufficient data points, crowdsourced data can be used to monitor the commitment to network rollout both in terms of infrastructure and spectrum. Another example could be when a fixed operator has the commitment to guarantee a set of minimum QoS parameters to the end-users. Crowdsourced data can be used to monitor the performance of such parameters.

II.1.5 Network planning

According to [b-CrowdWhitepaper] and inputs from a network planning and testing tool vendor, exemplary use cases for crowdsourcing in network planning are:

a) Expansions and/or rolling out new technologies. In this case, crowdsourced data can also be used to map the human activity (in terms of mobility) by looking at how the user's distribution changes over time. Identification of areas with high data usage and/or users density can be used within the scope of selecting the optimal location and density of deployment sites in the network.

b) Creation and tuning of propagation models in conjunction with topographical and geolocated data. In this case, the crowdsourced data requires measurements of radio frequency (RF) characteristics such as (but not limited to) signal level, signal quality, usage of frequencies or frequency bands. In addition, crowdsourced data used for this scope should be accurate, unbiased, free of artefacts possibility emerged from data collection and it needs to show high granularity (Guidelines for data collection requirements such as sampling, scheduling, filtering, categorization and aggregation can be found in clause 7.3).

c) Optimization of traffic models used in planning.

d) Increased profitability based on the optimization of quality (coverage, latency and throughput) vs. return on investment (RoI).

II.1.6 Network optimization

Using crowdsourced data for network optimization requires high data accuracy and granularity. If this is carefully ensured, then the following use cases can be supported by crowdsourced data:

a) On a regular (typically daily) basis, following the trends (even weak signals not visible by network-based supervision solutions) in the usage and the QoS of services delivered over the supervised network, in order to anticipate improvements such as an increase of capacity at the right location and time.

b) Establishing root causes for network issues: crowdsourced data can be used for geographical network problem detection related to service release, coverage, throughput and/or capacity.

c) Performing a dedicated collection of crowdsourced data on a specific target (e.g., a geographical zone, a group of users with the same model of mobile device) in order to detect and fix specific issues more quickly.

d) Optimizing and/or replacing blind (anywhere) and blanket (all over) drive testing with focused and on demand (when and where a problem occurs).

e) Visualize capacity usage of mobile services via indoor end-user density.

f) Monitor the efficiency of network optimization solutions after deployment.

II.2 Examples of QoS indicators and calculation for crowdsourced solutions

Examples of QoS indicators obtained from crowdsourcing approaches and their characteristics are presented in Table II.2. Indicators may be aggregated separately for different technologies, operators, regions and time frames.

It is important to highlight that conclusions based on the results from these indicators must consider the type of solution and setup scenario.

Table II.2 – Examples for QoS indicators

Indicator		Description	Service	Observations
Download and upload speed	Compliance with contracted speed (download and upload)	Percentage of tests in which measured speed reached contracted speed, calculated for downlink and uplink.	Fixed	This indicator can only be obtained when the information about contracted speed is available. Solutions applying CPE as device might give more suitable results, since the test is not influenced by Wi-Fi connection limitations.
	Compliance with minimum speed (download and upload)	Percentage of tests in which measured speed reached minimum speed established for each technology (3G, 4G), calculated for downlink and uplink.	Mobile	This indicator can only be obtained when the information about contracted speed is available.
	Typical speed (download and upload)	Central tendency result of a series of speed tests, calculated for downlink and uplink.	Fixed and mobile	Calculation may apply mean or median statistics, depending on data statistical behavior.

Table II.2 – Examples for QoS indicators

Indicator	Description	Service	Observations
Latency	Round-trip time in a series of latency tests (UDP packets transmission).	Fixed and mobile	Calculation may apply mean or median statistics, depending on data statistical behavior.
Jitter	Round-trip time variation in a series of latency tests (UDP packets transmission).	Fixed and mobile	Calculation may apply range or values at low and high percentiles as a summary statistics, depending on data statistical behavior.
Packet loss	Percentage of user datagram protocol (UDP) packets lost during a series of latency tests (UDP packets transmission).	Fixed and mobile	Calculation may apply mean or median statistics, depending on data statistical behavior.
Mobile access technology (3G, 4G, 5G, etc.)	Data points observed using different mobile access technologies.	Mobile	Calculation may apply percentage or ratio of different access technology.

NOTE 1 – 10th percentile (low) and 90th percentile (high) indicators can also be obtained for the different groups presented in previous table and might be suitable to complement average and median indicators, by giving information about worst and best results.

NOTE 2 – Compliance with minimum established latency, jitter and packet loss may also be used as complementary indicators when comparing results from different regions and operators.

NOTE 3 – For certain aggregated indicators such as but not limited to: packet lost, jitter and speed, access to detailed information such as the time partition can help to complement average values. For example, a low average packet loss session with high burstiness can give a worse user experience than a higher average packet loss session which has packet loss evenly spread out.

II.3 Examples of data filtering process rules

One of the main aspects of QoS assessment using the crowdsourcing approach is to validate the collected data and apply predefined criteria to approve or reject data records.

Examples of such criteria, also known as discarding rules, are shown in Table II.3:

NOTE – Clause 7.3.2 provides guidelines for data processing considerations such as filtering, categorization and aggregation.

Table II.3 – Data filtering rules examples

Identified situation	Description	Observations
Duplications	Records representing the same time and device.	
Measurement failure	Failure flag fields indicating there is a failure in test conduction.	

Table II.3 – Data filtering rules examples

Identified situation	Description	Observations
Measurement conditions changed	For active tests that do not complete instantly, such as download tests, where the network type or interface changes significantly (switch between cellular and Wi-Fi, or between different cellular technologies).	Applies for analyses with the objective to measure the performance for a specific scenario such as mobile-only or Wi-Fi-only. However, for certain analyses, these tests will be relevant and could be included.
Download and upload parity	Tests that present results only for download or upload.	Applies only for solutions in which each download test presents an upload test pair.
Test server identification	Test server does not correspond to accepted servers.	Applies only for active tests and cases in which test servers are pre-determined. For example, if international traffic is out of scope of measurement's strategy, tests against servers located outside the country must be excluded. Also, if the goal is to measure the on-net performance of an ISP, it is necessary to exclude tests against a server outside of the ISP's network.
Maximum measurements per time window exceeded	If the number of measurements exceed a maximum	For automatic tests, it is also possible to set a maximum number of tests to avoid collecting this kind of data. Alternatively, this data could be allowed into the pipeline so long as the statistical aggregation process prevents such readings having undue influence on the results. For example, this could be achieved by pre-aggregating to time-device bins ahead of the final calculation.
Data out of location and period scope	Tests presenting ranges of location and/or time period out of desired scope.	
Invalid or incomplete field values	Tests containing incomplete or invalid values for fields that are necessary for the calculation of indicators (such as location, battery level, device identification)	Acceptable rules for data fields must be predefined.
Conflicting field results	Results in different fields that bring conflicting information.	

Table II.3 – Data filtering rules examples

Identified situation	Description	Observations
Low battery	Results obtained from devices in conditions of low battery or battery level usage under accepted limits	Applies for solutions embedded in end-user's device. For automatic tests, it is also possible to set this as a condition for not starting the test.
Low signal level	Results obtained from devices in conditions of low signal level	Applies for solutions embedded in end-user's device and wireless connections (mobile or Wi-Fi). For automatic tests, it is also possible to set this as a condition for not starting the test. However, for certain analysis, these tests will be relevant and could be included.
Concurrent traffic	Results undermined by concurrent traffic in the device	Applies only for active tests. For automatic tests, it is also possible to set this as a condition for not starting the test.

Also, when evaluating compliance of results with contracted levels (such as the first indicator shown in Table I.2), the following list of filtering criteria may assist on the calculation of indicators in a given region/time period:

- The device IP does not correspond to the IP address pool of the assumed operator;
- Changes were made in the subscriber's contracted levels;
- The subscriber's contract was cancelled;
- The subscriber's address has changed;
- The subscriber's contract information could not be found.

Appendix III

Practical approaches to fixed broadband crowdsourcing

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

Regulators and operators face challenges in fixed-network crowdsourcing that are different from those occurring in mobile network environments. This appendix offers approaches which can be taken to implement fixed-network crowdsourcing solutions.

III.1 Standards for data collection from CPE

For a fixed-network broadband crowdsourcing approach, three standards can be considered when defining a solution allocated to CPE devices:

i) Broadband Forum Technical Report 069 [b-BBF TR-069], "CPE WAN Management Protocol"; and

ii) Broadband Forum Technical Report 143 [b-BBF TR-143], "Enabling Network Throughput Performance Tests and Statistical Monitoring";

iii) Broadband Forum Technical Report 471 [b-BBF TR-471], "Maximum IP-Layer Capacity Metric, Related Metrics, and Measurements".

TR-069 is an application layer protocol for communications between end-user's equipment and provider's central controllers, providing standardized procedure for customer premises equipment (CPE) remote management. This protocol supports functions such as auto-configuration, software or firmware image management, software module management and status. The identification mechanisms included allow CPE provisioning based either on the requirements of a specific CPE or on collective criteria, such as the CPE vendor, device model and software version.

The architecture of TR-069 enables device management with the CPE both at the customer's gateway, and with devices within the customer's office/home network.

TR-143 was originally designed for operators as a tool to enable monitoring the performance of the network continuously to prevent problems from occurring and diagnose problems when they occur. The technical report provides definitions for the CPE data model objects for performance throughput tests and data monitoring on the IP interface of a CPE using the diagnostic mechanism defined in TR-069.

In the context of TR-143, active monitoring is described as "actively transmitting or receiving data in a controlled test". In this sense, the tests conducted are network layer centric, and are agnostic to the underlying access network. Also, it can support both network and CPE initiated diagnostics. TR-471 is the latest Broadband Forum (BBF) development in the active monitoring and measurement category, improving on the TR-143 methods for high-speed Internet access assessment. TR-471 includes guidance for deployment in the BBF IP Edge and CPE, and complementary BBF work adds the necessary data model to configure and control TR-471 measurements.

The protocols also provide tools to manage the CPE-specific components of optional applications or services, for which an additional level of security is required to control, such as those involving payments.

In the context of crowdsourcing for the assessment of fixed broadband networks, it is possible to apply protocols defined in TR-069, TR-143, and TR-471 for data collection and to obtain QoS indicators. This approach enables the use of existing CPE as testing devices. Nonetheless, the recommendations regarding crowdsourced data collection, including data sampling, test scheduling and data processing can be applied.

Among the benefits and challenges of these protocols are:

- Benefits:
  - The tests are agnostic to the underlying access network.
  - The provisioning mechanism allows a straightforward future extension to allow provisioning of services.
  - Many routers already support TR-069 standard (over 1 billion worldwide) and therefore implementation can be efficient and cost-effective.
  - While TR-069 is mainly used for operators, there are already vendors who have TR-069 toolset which can be repurposed for regulatory use.
- Challenges:
  - Although TR-069 is widely available, old routers may not support it, this is especially important to consider for rural areas.
  - From a regulator's perspective, it might be challenging to implement measurement campaigns based on these protocols as it requires close collaboration from the operators and the vendors.

III.2 Providing end-user measurement tools

Regulators and operators alike may offer end users measurement tools, for example speed tests, to their end users. The tools can be accessible using a web browser or can be installed on a computer or mobile device. Mobile device in this context must be connected through Wi-Fi to the CPE which in turn is connected to the Internet using fixed access technology (such as ADSL/fibre, etc.). The tools typically rely on the user-initiated or automated. The advantage of this approach is that such tools are easy to use and understand by the end-users and can provide instant information and help. Unlike in mobile crowdsourcing scenarios the reliability of the measurements can be influenced by the existence of a wireless link between the client device and CPE. Crowdsourcing solutions solve this problem in few different ways:

- Prevent running the test on wireless links: This is possible only on computers where a speed test tool is installed in the operating system and is able to detect whether a user is connected via the Ethernet or wireless.
- Measurement tool on the end user device connects to a specialized application programming interface (API) present on the CPE to determine the quality of the wireless link and is able to inform the end user if the test has been impacted by the wireless link. The API on the CPE can also provide additional information about the bandwidth utilization from other devices connected to the same CPE. This information is useful to adjust the measurement result to compensate for the additional capacity taken by other devices or to exclude the measurement result altogether.
- Measurement tool which is running natively on the computer or mobile device is able to assess the local wireless link between the device and CPE from the parameters provided by the operating system and perform additional tests to validate the measured throughput values. Additional active measurements using the UDP protocol to measure, for example, the IP packet sending bit rate can estimate the Wi-Fi capacity and inform the end users whether the test result is correctly representing the speed of their Internet connections.

The choice of how the tool is offered to the user impacts what data can be retrieved about the device and network state during the measurement. Table III.1 compares what data can be collected based on the type of the tool offered.

Table III.1 – Data collected with different tools

	Web-based tool	Native app	API on the CPE
Device information	Limited	Yes	Limited
User's ISP	Yes	Yes	Yes
Fixed connection type (e.g., fibre/ADSL)	No	No	Yes
User's contracted plan (e.g., 100 Mbit, 1 Gbit)	No	No	Yes
Router model	No	Yes	Yes
List of connected devices to the CPE	No	Limited	Yes
Local cross-traffic bandwidth	No	Limited	Yes
Wi-Fi throughput	No	Yes	Yes
Internet throughput	Yes but may produce different results between tools	Yes but may produce different results between tools	Yes but may produce different results between tools
Wi-Fi IP-Layer Capacity	No	Yes	Yes
Internet IP-Layer Capacity	Yes	Yes	Yes
Wi-Fi latency	No	Yes	Yes
Internet latency	Yes but may produce different results between tools	Yes but may produce different results between tools	Yes but may produce different results between tools
Wi-Fi connection properties (frequency, channels etc.)	No	Yes	Yes
Possibility to detect device limitations (e.g., CPU, RAM)	No	Yes	No

Based on Table III.1 it may be concluded that an API on the CPE is the best choice for fixed crowdsourcing as it offers the most extensive collected data. However, the complexity of deploying such a solution for all the users who wish to take advantage of the system must also be considered.

From that perspective, the most suitable solution is the web-based one, as the solution can be deployed on a web page and is easily accessible for all users.

Native apps provide a good balance between the amount of collected data and deployment complexity. Native apps for various operating systems do have a small barrier for users to overcome, but generally, users know how to install new software on their devices.

An API on the CPE has the highest deployment complexity as it requires either router manufacturers or ISPs who distribute the CPEs to the users to integrate the API on every router. Networks which have the most homogenous CPE deployed base and most advanced models are the most suitable for the API approach. New CPE models will have better resources available to perform new functions as well as better options to remotely upgrade the firmware.

From Table III.1 there is a subset of the collected data which is useful for aggregate analyses and drawing conclusions about the quality of service. Additionally, there are other collected data which are mostly useful for validation of the measurement themselves and do not provide much value in being aggregated across different devices.

The following are examples for collected data which are useful for aggregated analyses:

- User's Internet service provider (ISP)
- Fixed connection type
- User's contracted plan
- Router model
- Internet throughput/latency

NOTE – Data related to Internet throughput/latency captured by different ways of collection at different layers may lead to different results. Therefore, care must be taken to avoid inconsistent results by mixing different collection methodologies.

III.3 Mass scale deployment of proprietary hardware probes to end users

Regulators and operators may decide for a large deployment of hardware probes to customer homes. Typically, a probe is a standalone low-cost hardware device which can be connected to the CPE using reliable Ethernet connection. A hardware probe runs automated tests in the background on the user's connection and reports results to the operator of the crowdsourcing system.

Among the benefits and challenges of hardware probes in a crowdsourcing scenario are:

- Ability to schedule automated measurements which are aware of the network conditions;
- Possibility to reliably execute diverse sets of measurements on the hardware that is designed specifically for running measurements;
- Relative to other methods, hardware probes have a higher cost due to the need of purchasing specific hardware and shipping;
- Given the high cost and logistical challenges, it cannot be assumed that all users will be able to take advantage of the system, which in turn could limit the sample size in terms of representativeness (clause 7.3.1 provides guidelines on the sampling and scheduling for statistical validation);
- Constant need to maintain a sustainable pool of users as users may stop measurements or hardware probes may be malfunctioning.

Bibliography

[b-ITU-T G.9961] Recommendation ITU-T G.9961 (2018), *Unified high-speed wireline-based home networking transceivers – Data link layer specification*.

[b-ITU-T P.912] Recommendation ITU-T P.912 (2016), *Subjective video quality assessment methods for recognition tasks*.

[b-ITU-T Y.1540] Recommendation ITU-T Y.1540 (2019), *Internet protocol data communication service – IP packet transfer and availability performance parameters*.

[b-CrowdWhitepaper] Tobias Hoßfeld and Stefan Wunderer (2020/03), *White Paper on Crowdsourced Network and QoE Measurements – Definitions, Use Cases and Challenges, eds., Würzburg, Germany, doi: 10.25972/OPUS-20232.*

[b-BBF TR-069] Broadband Forum TR-069 (2018/03), *CPE WAN Management Protocol.*

[b-BBF TR-143] Broadband Forum TR-143 (2015/08), *Enabling Network Throughput Performance – Tests and Statistical Monitoring.*

[b-BBF TR-471] Broadband Forum TR-471 (2020/07), *Maximum IP-Layer Capacity Metric, Related Metrics, and Measurements.*

[b-Scheaffer] Scheaffer, Richard & Mendenhall, William & Ott, Lyman (2012). *Elementary Survey Sampling, 7th Edition, Cengage Learning*.

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, 2020