ETSI TR 103 639 V1.1.1 (2021-03)

Speech and multimedia Transmission Quality (STQ); Timeslicing KPIs for RTP based speech transmission

Reference
DTR/STQ-281

Keywords
RTP, timeslicing

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the prevailing version of an ETSI deliverable is the one made publicly available in PDF format at www.etsi.org/deliver.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

Copyright Notification

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

Foreword

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

The present document presents a classification for systems analysing speech on the basis of fixed time slices and describes methods for generating key performance indicators using time slice data to assess voice service performance.

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Introduction

Methods for characterizing the quality of interactive speech transmission based on the Real-Time Transport Protocol (RTP), e.g. for Voice over IP (VoIP) services, are standardized by ETSI, ITU-T, ITU-R and the IETF. These methods typically provide fundamental metrics such as packet loss, packet delay variation and packet reordering as well as derived metrics such as user experience estimates in terms of the Mean Opinion Score (MOS).

Most existing methods aim to characterize the quality of entire calls. Such data can hardly be aggregated to determine the quality of a set of variable length RTP flows, a route or an entire telephony service. This is particularly problematic for passive (one-sided, in-service) methods measuring live traffic, since durations and other call parameters are not under control and can vary significantly.

The use of fixed duration sample intervals facilitates the creation of meaningful statistics through temporal aggregation. Timeslicing measurement methods generate metrics for fixed duration sample intervals. Statistics and Key Performance Indicators (KPIs) based on such timeslice metrics allow to condense information and adequately characterize the quality of a set of RTP flows.

In an attempt to structure existing approaches to timeslice-based analysis of RTP speech transmission, the present document presents a framework for timeslicing methods and metrics. It builds on the Framework for IP Performance Metrics (IPPM) [i.2] developed by the IETF and essentially applies the IPPM to RTP-based speech transmission and uses it to define the concept of timeslice KPIs.

1 Scope

The present document describes a framework for measurement methodologies and metrics assessing characteristics of RTP-based speech transmission for fixed duration time intervals. This approach can be used to evaluate aspects of speech transmission based on the observed media volume in terms of time units. This facilitates temporal aggregation of metrics and calculation of key performance indicators in a more meaningful way compared to aggregation of conventional call-based metrics.

The present document presents a classification of methods obtaining RTP flow characteristics per fixed time unit and provides examples for actual timeslice metrics as well as aggregation schemes to obtain key performance indicators summarizing metric data related to a set of timeslices.

The focus is on interactive speech transmission in IP-based networks, i.e. Voice over IP (VoIP) communication. Fundamental concepts are potentially also applicable to interactive video communication, video streaming and other forms of continuous RTP-based communication.

The framework introduces a common foundation to exchange information on timeslice metrics for RTP-based speech transmission performance. The intended audience for the present document can be found among service providers, vendors, and users of telephony services.

The reader is assumed to be familiar with the Framework for IP Performance Metrics (IPPM) [i.2] developed by the IETF. The terminology of the IPPM will be used wherever possible and extended when necessary.

2 References

2.1 Normative references

Normative references are not applicable in the present document.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] Recommendation ITU-T Y.1540 (12/2019): "Internet protocol data communication service - IP packet transfer and availability performance parameters".

[i.2] IETF RFC 2330 (May 1998): "Framework for IP Performance Metrics". V. Paxson, G. Almes, J. Mahdavi, M. Mathis.

[i.3] IETF RFC 3550 (July 2003): "RTP: A Transport Protocol for Real-Time Applications". H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson.

[i.4] ETSI EG 202 765-3: "Speech and multimedia Transmission Quality (STQ); QoS and network performance metrics and measurement methods; Part 3: Network performance metrics and measurement methods in IP networks".

[i.5] IETF RFC 6076 (January 2011): "Basic Telephony SIP End-to-End Performance Metrics". D. Malas, A. Morton.

[i.6] TM Forum GB934 (May 2013): "VoIP Application Notes Release 3 Version 1.1".

[i.7] IETF RFC 3432 (November 2002): "Network performance measurement with periodic streams". V. Raisanen, G. Grotefeld, A. Morton.

[i.8] IETF RFC 5835 (April 2010): "Framework for Metric Composition". A. Morton, S. Van den Berghe.

[i.9] IETF RFC 7799 (May 2016): "Active and Passive Metrics and Methods (with Hybrid Types In-Between)". A. Morton.

[i.10] IETF RFC 3611 (November 2003): "RTP Control Protocol Extended Reports (RTCP XR)". T. Friedman, R. Caceres, A. Clark.

[i.11] Recommendation ITU-T P.563 (05/2004): "Single-ended method for objective speech quality assessment in narrow-band telephony applications".

[i.12] Recommendation ITU-T Y.1541 (12/2011): "Network Performance Objectives for IP-based services".

[i.13] Recommendation ITU-T P.564 (11/2007): "Conformance testing for voice over IP transmission quality assessment models".

[i.14] IETF RFC 7679 (January 2016): "A One-Way Delay Metric for IP Performance Metrics (IPPM)". G. Almes, S. Kalidindi, M. Zekauskas, A. Morton.

[i.15] IETF RFC 7680 (January 2016): "A One-Way Loss Metric for IP Performance Metrics (IPPM)". G. Almes, S. Kalidindi, M. Zekauskas, A. Morton.

[i.16] IETF RFC 3393 (November 2002): "IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)". C. Demichelis, P. Chimento.

[i.17] IETF RFC 3551 (July 2003): "RTP Profile for Audio and Video Conferences with Minimal Control". H. Schulzrinne, S. Casner.

[i.18] IETF RFC 5560 (May 2009): "A One-Way Packet Duplication Metric". H. Uijterwaal.

[i.19] Recommendation ITU-T P.863 (03/2018): "Perceptual objective listening quality prediction".

[i.20] Recommendation ITU-T G.107 (06/2015): "The E-model: a computational model for use in transmission planning".

[i.21] Recommendation ITU-T G.107.2 (06/2019): "Fullband E-Model".

[i.22] ETSI TS 102 250-2 (V2.4.1): "Speech and multimedia Transmission Quality (STQ); QoS aspects for popular services in mobile networks; Part 2: Definition of Quality of Service parameters and their computation".

[i.23] Recommendation ITU-T P.800.1 (07/2016): "Mean opinion score (MOS) terminology".

3 Definition of terms, symbols and abbreviations

3.1 Terms

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

Key Performance Indicator (KPI): metric aggregating a set of sample statistics

metric: specified quantity characterizing the performance and reliability of observed communication

sample: set of singleton metrics measured in a specified period as defined in IETF RFC 2330 [i.2]

sample statistic: statistical measure computed using the values defined by the singleton metric on the sample as defined in IETF RFC 2330 [i.2]

singleton: atomic metric as defined in IETF RFC 2330 [i.2]

timeslice: fixed duration sample

3.2 Symbols

Void.

3.3 Abbreviations

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

CCR	Critical Call Ratio
CMR	Critical Minute Ratio
CSR	Critical Stream Ratio
DSCP	Differential Service Code Point
GCR	Good Call Ratio
GMR	Good Minute Ratio
GSR	Good Stream Ratio
IETF	Internet Engineering Task Force
IP	Internet Protocol
IPDV	Inter-Packet Delay Variation
IPPM	IP Performance Metrics
ITU-T	International Telecommunication Union - Telecommunication standardization sector
KPI	Key Performance Indicator
MOS	Mean Opinion Score
MOS_{LQE}	Mean Opinion Score Estimated Listening Quality
MOS_{LQO}	Mean Opinion Score Objective Listening Quality
PLC	Packet Loss Concealment
RFC	Request For Comments
RTCP	RTP Control Protocol
RTP	Real-Time Transport Protocol
SER	Session Establishment Ratio
SIP	Session Initiation Protocol
SS7	Signalling System 7
TS	TimeSlice
UDP	User Datagram Protocol
VLAN	Virtual Local Area Network
VoIP	Voice over Internet Protocol

4 Framework for timeslice metrics and measurement methods

4.1 Overview

A Key Performance Indicator (KPI) measures the success of a business activity against an operational goal. KPIs are often stated as averages, ratios or percentages relative to the quantified objective. In the context of telecommunications, operational goals include high service availability, network connectivity, speech quality and performance of individual network elements.

There are many different standards for measuring the performance of telephony signalling protocols, such as Signalling System 7 (SS7) or the Session Initiation Protocol (SIP). Likewise, there are many standards specifying metrics for speech transmission performance. Most of the metrics underlying telephony performance measurements are based on observations pertaining to individual calls. For example, the widely used KPI Session Establishment Ratio (SER) [i.5] measures the ability of a network to successfully establish a SIP session as the percentage of successfully established calls over all call attempts.

For signalling performance measurements such call-based metrics are fundamental since the ability to setup, control and tear down calls is the objective of telephony signalling. Under certain conditions call-based metrics are also applicable to measurements of speech transmission performance. For example, in circuit-switched landline telephony systems, speech quality is typically very stable over time and hence a single metric per call has historically been considered sufficient.

In packet-switched telephony over IP networks, using RTP for media transport, speech transmission performance is more volatile and a single metric per call does not allow to accurately assess the performance over time. For example, the packet loss rate is a key metric for RTP-based speech transmission, but the loss distribution is needed to understand its impact on speech quality. Single metrics per call also prevent meaningful aggregation for higher-level statistics and KPIs, when the call duration varies. This applies in particular to passive measurement techniques, which analyse user-initiated calls.

The issues of volatility in speech transmission performance and varying call durations are addressed by timeslicing measurement methods, which analyse RTP flow characteristics for fixed time segments. Such timeslicing methods generate metrics that allow to calculate speech transmission KPIs based on media volume, e.g. the percentage of time intervals where speech transmission performance did not meet a stated objective [i.6]. Commercial tools implementing timeslicing measurement methods exist, but the concept lacks a general model and terminology. The following clauses define a framework for timeslicing measurement methods and performance metrics of RTP-based speech transmission. The framework is based on the Framework for IP Performance Metrics [i.2].

NOTE: Timeslicing could alternatively be defined using Recommendation ITU-T Y.1540 [i.1] or ETSI EG 202 765-3 [i.4].

4.2 Modelling timeslice metrics

Timeslice metrics can be modelled using the Framework for IP Performance Metrics (IPPM) [i.2] as a basis. The IPPM distinguishes between "singleton metrics", "samples" and "sample statistics" as illustrated in Figure 1. A singleton metric corresponds to a single observation. A "sample" is a collection of singleton measurements, and a sample statistic is an aggregation of singleton measurements over a sample period.

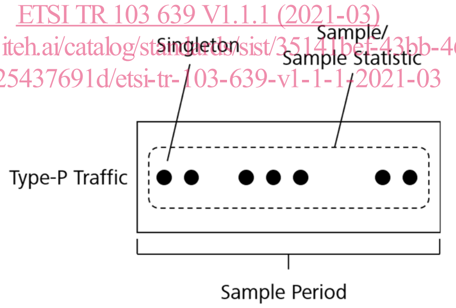

Figure 1: Illustration of basic IPPM concepts

In addition, IPPM defines the notion of a packet of type P. The idea is that metrics may depend on the type of the observed packet stream, as IP networks can treat packets differently depending on the used protocol, ports and other packet flow characteristics. The names of IPPM metrics therefore include either the specific type of the packet stream or a phrase such as type-P.

The IPPM framework provides language and a structured approach to defining metrics, such as one-way delay, one-way packet loss and packet duplication.

The present document builds on the IPPM to define a framework for timeslice metrics and KPIs qualifying the performance of RTP-based speech transmission. The basic idea of RTP timeslicing is to segment RTP flows into consecutive samples with fixed sample periods and to define fundamental and composed metrics on these samples. To this end, the general IPPM framework is parameterized for VoIP timeslicing as described in the following:

- Singleton metrics are of type "RTP-VoIP", i.e. they refer to packet streams transporting speech using RTP over UDP.

- Sample periods are of fixed duration and shorter than the typical length of a call [i.6].

- All RTP packets in a sample period contribute to singleton metrics to fully characterize the timeslice.
- Samples are consecutive and continuous for the duration of each RTP stream.

Based on this, timeslicing can be defined as the consecutive execution of continuous sampling on RTP streams using fixed duration sample periods. The application of the basic IPPM concepts to timeslicing measurements is illustrated in Figure 2.

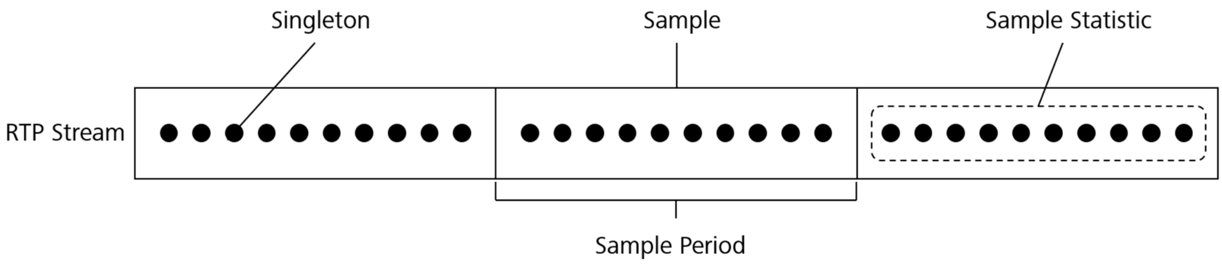

Figure 2: Application of IPPM concepts to timeslicing measurements

NOTE: In the present document, only methods with fixed time unit timeslicing are considered. Methods with adaptive interval lengths, for example methods using sample periods defined by a fixed number of packets instead of a fixed duration, are out of scope.

The Type-P of traffic addressed by the present document is similar to periodic streams defined in IETF RFC 3432 [i.7] as similar sized packets transmitted through a network at regular intervals. For this kind of traffic [i.7] - which is also based on the IPPM - proposes a periodic sampling methodology and sample metrics. Timeslicing as proposed in the present document is related to periodic sampling, but differs in three ways. First, sample periods are shorter than a typical call, whereas [i.7] suggests to use sample periods corresponding to a typical call duration. Second, timeslicing uses consecutive fixed duration samples, i.e. a zero interval between uniform length samples, whereas [i.7] considers multiple samples a special case and allows arbitrary intervals. Third, timeslicing considers all packets in a sample period, whereas [i.7] introduces a random offset from the beginning and end of each sample period.

4.3 Classification of timeslicing measurement methods

4.3.1 Classes

Timeslicing as defined in clause 4.2 applies to active and passive measurements methods as specified in IETF RFC 7799 [i.9]. Active methods depend on packet streams, e.g. from artificial test calls, generated for the purpose of measurement and observation. Passive methods measure and observe existing packets streams, e.g. RTP packet streams transmitting speech of actual calls.

Passive methods either directly measure streams of RTP packets or they use data contained in RTCP, IETF RFC 3550 [i.3] or RTCP-XR, IETF RFC 3611 [i.10] VoIP metrics reports, which summarize measurements performed by the communication endpoints. These two approaches need to be distinguished, because they yield different measurement results. Direct measurement of RTP packets assesses RTP-VoIP flow characteristics at the measurement point, i.e. along the path from source to destination. In contrast, RTCP and RTCP-XR packets report on measurements performed by the endpoints, i.e. the data provides an end-to-end view. Another relevant difference is that passive methods by definition assess existing traffic and the devices sending RTCP/RTCP-XR reports may not be known and trusted. In contrast, direct measurement of RTP packets is typically performed by known devices, often referred to as probes, whose measurement characteristics are known. Table 1 classifies the different measurement approaches.

Table 1: Classification of timeslicing measurement methods

Class	Type	Description
Active, Embedded	A	The system uses measurement data obtained from segmented objective algorithmic testing.
Passive, Embedded	B	The system uses data from RTCP reports representing RTP media flow characteristics as measured by the reporting endpoints.
	C	The system uses data from RTCP-XR reports representing RTP media flow characteristics as measured by the reporting endpoints.
Passive, Midpoint	D	The system uses data from continuous RTP flow measurements at one or more points in the network.

NOTE: Only methods assessing the actual stream of RTP packets are considered in the present document. Other timeslicing measurement methods, e.g. decoding the transmitted audio and performing segmented single-ended analysis, similar to Recommendation ITU-T P.563 [i.11], are out of scope. Likewise, hybrid methods as described in IETF RFC 7799 [i.9] are not considered in the present document.

4.3.2 Reference measurement setup

This clause describes the reference measurement setup for the three classes. Methods differ in measurement and observation points, scope, control over sample periods and modelling of the communication endpoints.

Endpoint models are needed when assessing the impact of RTP-based speech transmission impairments on the application level, i.e. the VoIP service performance. This impact assessment requires knowledge or basic assumptions about the communication endpoints, related to dejitter buffer configurations and PLC mechanisms. Models of hypothetical endpoints are described e.g. in Recommendation ITU-T Y.1541 [i.12] and Recommendation ITU-T P.564 [i.13].

NOTE: Endpoint models are less relevant when solely the RTP packet transmission performance on the network level, e.g. when pure one-way packet loss is of interest.

The reference measurement models for the three defined classes are shown in the figures below. For simplicity, the models only show unidirectional RTP packet transmission, whereas VoIP calls generally transmit RTP packets in both directions.

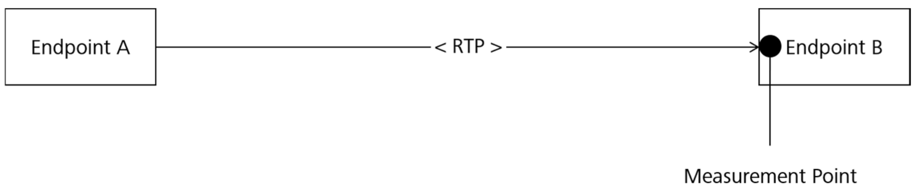

Figure 3: Reference model for Type A active measurement methods

Type A active timeslicing measurement methods generate test calls between known endpoints, i.e. the metrics delivered by a type A system deliver an end-to-end view. Because the measurement method is co-located with the call generator, all relevant information about the endpoints is known and the VoIP application level performance can be determined for a specific measurement setup. Active timeslicing methods can generally control the sample period or at least the period is known, however specific methods may impose limitations on the minimum and maximum sample duration and other test parameters.

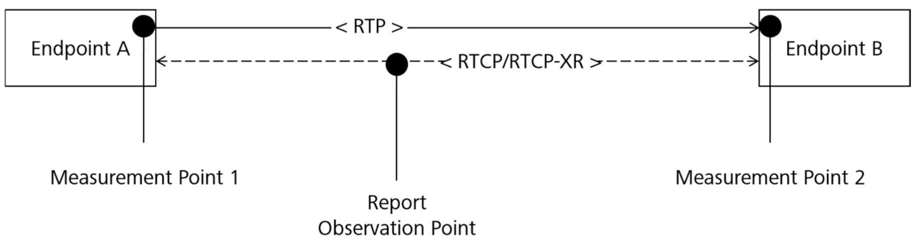

Figure 4: Reference model for Type B/C passive measurement methods