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

TECHNICAL REPORT

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

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.

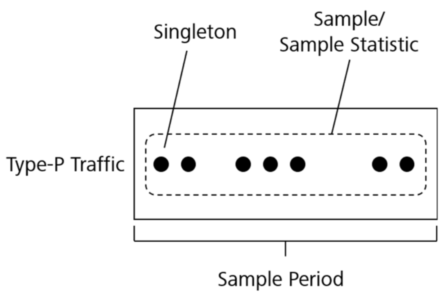

Figure 1: Illustration of basic IPPM concepts

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.

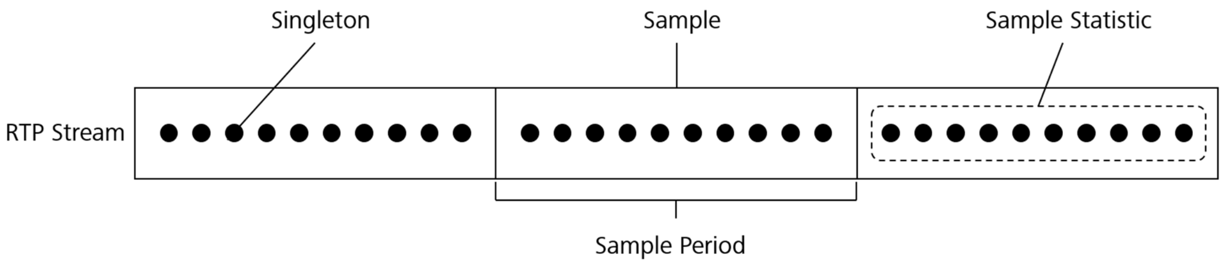

Figure 2: Application of IPPM concepts to timeslicing measurements

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.

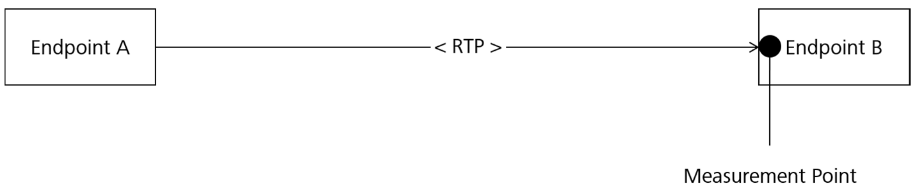

Figure 3: Reference model for Type A active measurement methods

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.

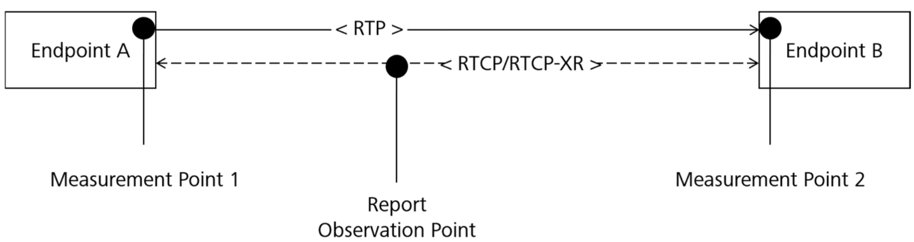

Figure 4: Reference model for Type B/C passive measurement methods

Type B and C passive measurement methods provide metrics based on measurement data exchanged between endpoints via the RTCP (type B) or RTCP-XR (type C) protocols. The reported measurements are performed by the endpoints. The measurement scope is end-to-end, i.e. from endpoint to endpoint, but reports containing the measurement data can be observed anywhere along the path. It is possible that RTCP packets do not take the same path as their corresponding RTP streams.

Typically, when existing RTP traffic is measured, the endpoints will be unknown and the reported data may be inaccurate. In addition, the sample period for timeslicing of existing traffic may not be known a priori, may differ between different endpoint pairs and may even change over time. For this reason, Type B and C measurement methods require well-known and controlled environments to generate timeslice metrics.

The difference between type B and type C models is the reporting protocol. Type B methods use RTCP which provides basic information on the RTP stream transmission performance and no information relevant to the VoIP application level performance. Type C methods use optional RTCP-XR report blocks which enable an endpoint to provide information about its configuration and internal state. In particular, specific RTCP-XR report blocks include basic information on the dejitter buffer and the Packet Loss Concealment (PLC) algorithm, which allows to estimate the impact of jitter and packet loss on the user experience. The ability of type C methods to calculate specific singleton metrics depends on the availability of relevant report blocks, i.e. on the sender of RTCP-XR reports. The following clauses assume that report blocks required to calculate specific metrics are available.

Figure 5: Reference model for Type D passive measurement methods

Type D passive methods perform direct measurements of RTP streams at one or more distinct points in the network. Measurements reflect the RTP stream transmission performance from the source up to the respective measurement point. This mid-point view on transmission performance is useful when assessing traffic at network borders, e.g. at interconnections between VoIP service providers.

Type D passive methods measuring existing traffic will generally have no knowledge about the endpoints of RTP communication. Such methods therefore likely require a hypothetical endpoint model to estimate the application level performance.

In contrast to type B and C methods, type D methods do not rely on measurements performed by the endpoints, but perform direct measurements along the transmission path. Type D methods therefore have control over the actual sample period duration.

4.3.3 Qualification of timeslicing measurement methods

A qualifier for a timeslicing measurement method consists of the type and an indication of the sample period, i.e. timeslice duration, in seconds. For example, Type D-5 signifies a passive monitoring system providing timeslice metrics based on five second sample durations. The unit seconds is chosen because timeslicing aims to segment RTP streams into multiple fixed units and call durations are typically on the order of seconds or a few minutes.

When the sample period is unknown or variable an *x* is added. This is typically the case for measurement methods of type B and C measuring existing calls, because the endpoints control when RTCP packets are sent, not the measurement method.

EXAMPLE: Type B-x signifies a method using data from observed RTCP packets to generate timeslice metrics.

Methods that provide application level performance metrics require a description of the endpoint model to complement method qualification, as the model has a significant impact on application level metrics.

4.4 Errors and uncertainties

4.4.1 Incomplete samples

The IPPM provides general guidance on potential error sources and uncertainties related to different measurement methodologies. On top of general error sources like clock inaccuracies, timeslicing methods are also affected by specific effects related to the slicing process. In particular, border cases are of interest, e.g. when the duration of a measured RTP stream is not a multiple of the sample period.

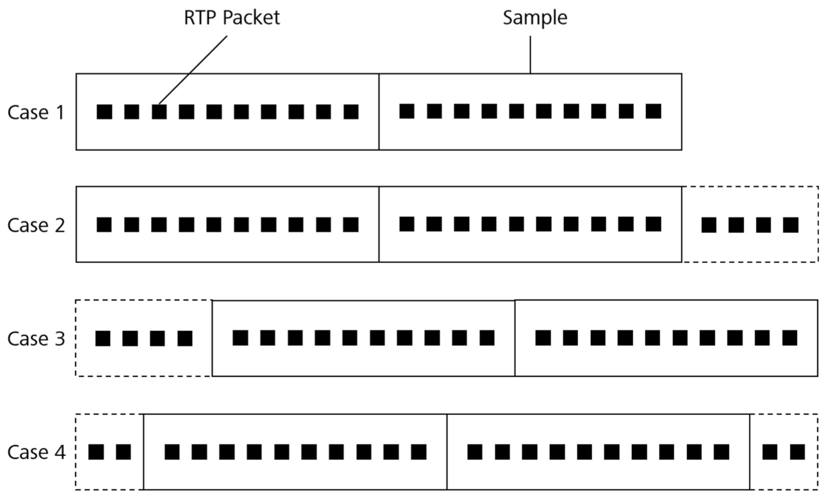

Figure 6: Potential slicings of an RTP stream

If the observed RTP stream has a duration that is a multiple of the sample period and slicing starts with the first RTP packet, then the samples cover the entire RTP stream and each sample has the maximum number of singletons. This scenario is illustrated in case 1 of Figure 6. This can be the result of deliberate configuration of Type A active testing or it can happen by chance when performing passive measurements.

Incomplete samples at the end of an RTP stream, as illustrated in case 2, occur when a timeslice starts with the first RTP stream packet, but the stream is not sufficiently long to complete the final period. This could be deliberate, although there is little meaning to it from a measurement perspective. More likely, this case occurs when a passive method starts measurement with the first packet of the observed RTP stream.

case 3 shows the situation, when the first sample is incomplete, i.e. its duration is less than the fixed sample period. This can happen, when slicing is not performed individually for each observed RTP stream - in which case the sample period would start with the first packet -, but simultaneously for all streams observed at a measurement point.

Case 4 illustrates the situation, when neither the first nor the last sample is complete. As in case 3 this can occur, e.g. when observing existing traffic with a passive measurement method that simultaneously slices all streams at a measurement point.

The above discussion focuses on systematic errors caused by the slicing process, but all illustrated cases can also be the result of packet loss at the beginning or end of an RTP stream.

Incomplete samples can lead to a lack of singleton metrics. For example, if only one packet is observed during a sample, then it is not possible to calculate the inter-packet delay variation.

A measurement method needs to define how incomplete samples contribute to metrics calculated from multiple samples. Discarding results from incomplete samples can lead to a complete lack of measurement data. Including results from incomplete samples can have negative effects on multi-sample metrics.

4.4.2 Varying singleton count

The cases discussed in clause 4.4.1 lead to incomplete samples caused by the slicing of RTP packet streams. Even if a sample is complete, in the sense that the observed RTP stream extends from the start to the end of the sample period, it may still not yield the maximum possible number of singletons, that can be expected for a timeslice. Reasons for this include desired VoIP media stream characteristics, such as silence suppression, or network events, such as packet loss preventing the meaningful calculation of inter-packet delay singletons.

As a consequence, the number of singletons per fixed timeslice is not necessarily constant, and it is possible that timeslices have varying singleton counts. This potentially leads to uncertainties regarding comparability of timeslice sample metrics and aggregation of such metrics. Full qualification of timeslice measurement methods includes information about how varying singleton counts are treated.

4.4.3 Sliced singletons

Singleton metrics may require the observation of multiple packets, e.g. when detecting sequences of lost packets or burst loss. Erroneous measurements can occur if the observation period of such a multi-packet singleton metric crosses over the boundary between consecutive samples. If carry-over of information between samples is not considered, then, in this example, instead of one burst loss two single packet losses will be observed. The way singleton slicing is handled is part of the singleton definitions for timeslicing measurement methods.

4.4.4 Incomplete stream slicing

Incomplete stream coverage occurs, when a consecutive series of samples does not fully capture the characteristics of an observed RTP stream. This is a common issue with RTCP-based methods, which exchange reports on measurements performed by the endpoints. If an RTP stream ends before the end of a reporting interval (sample period) and the endpoint does not send partial reports then the final part of the RTP stream is not characterized. This is a source of error and uncertainty particularly when trying to characterize the user experience, as the final part of an RTP stream is most relevant due to the recency effect.

5 Some timeslice metrics

5.1 Overview

A number of metrics have been described based on the Framework for IP Performance Metrics (IPPM) [i.2], such as a one-way delay metric [i.14], a one-way loss metric [i.15], and a packet delay variation metric [i.16]. These metrics are generally also applicable to timeslicing of type RTP-VoIP packet streams. This holds in particular for the definitions of singletons as they are based on individual packets or sets of packets.

The main differences between generic IPPM metrics and RTP-VoIP timeslice metrics are the duration of the sample period, and the selection function for singleton measurements within a sample period. Even though the IPPM framework allows for other approaches such as IETF RFC 3432 [i.7], the IPPM metric definitions initially assumed comparatively long sample periods and sample at times determined by a Poisson process [i.2]. In contrast, timeslice metrics for type RTP-VoIP streams are based on short successive sample periods and all possible singletons within a sample period. A further difference is that it is possible to define RTP-VoIP singletons using known protocol characteristics, such as RTP header information, thereby eliminating the need for synchronized clocks.

The following clauses define basic RTP-VoIP timeslice metrics which are based on respective IPPM metrics. The definitions reflect the intended meaning of a metric and specific implementations depend on the type of timeslice measurement method.

5.2 A one-way loss timeslice metric

5.2.1 A singleton definition for one-way packet loss

The following defines a one-way packet loss singleton that is based on the specification of the respective generic IPPM metric [i.15]. Subsequent clauses will use this singleton definition to define sample metrics and statistics.

Metric name

RTP-VoIP-One-way-Packet-Loss.

Metric parameters

- Src, the IP address of a sender of RTP-VoIP packets.
- Observer, a receiver of type RTP-VoIP packets.
- Flow, a sequence of RTP-VoIP packets sent by Src belonging to an RTP session.
- SeqNo, a sequence number contained in the header of an RTP-VoIP packet.

Metric units

The value of an RTP-VoIP-One-way-Packet-Loss is either a zero (signifying successful transmission of the packet) or a one (signifying loss).

Definition

The RTP-VoIP-One-way-Packet-Loss from Src to Observer for Flow at SeqNo is 0 means that Src sent an RTP-VoIP packet of RTP session Flow with sequence number SeqNo and that Observer received that RTP packet.

The RTP-VoIP-One-way-Packet-Loss from Src to Observer for Flow at SeqNo is 1 means that Observer did not receive a packet of RTP session Flow with sequence number SeqNo.

Discussion

The following issues need to be considered:

- If an RTP packet arrives but is corrupted, then it is counted as lost.
- If an RTP packet is received two or more times by Observer, then the packet is counted as received.
- It needs to be defined how to conclude that an expected packet has not been received. This particularly applies to passive type D methods, which have little information on if and when a packet can be expected at all. One approach is to wait for each packet for a defined period of time. Another option is to wait for a defined number of packets, if the expected packet is received. For measurement methods implementing a dejitter buffer to assess the effect of impairments on the application level, the latter approach should be preferred. This ensures consistency in loss measurement on the network and application level.

Methodologies

The metric definition is applicable to all timeslicing measurement types defined in clause 4.3.1. The Observer parameter is either:

- the RTP destination of an active test call (type A);
- the destination of RTCP/RTCP-XR packets associated with and reporting on the RTP stream (types B and C); or
- a monitoring device performing direct measurements of RTP streams on a network link (type D).

Pure passive measurement methodologies of type D do not explicitly know if a Src is sending a Flow of RTP packets and consequently also do not know the sequence number SeqNo to expect. However, according to the RTP specification IETF RFC 3550 [i.3] the sequence number is incremented by one for each RTP packet sent. When an Observer detects an RTP packet with sequence number SeqNo the Observer will expect the next RTP packet to have the sequence number SeqNo+1 (or higher in case of loss).

NOTE: The RTP sequence number is 16 bit and its initial value is random [i.3]. Therefore, a potential wraparound of RTP sequence numbers needs to be considered by a given timeslicing methodology.

Errors and uncertainties

The IPPM Framework [i.2] provides general guidance on sources of errors and uncertainties relating to measurements. Specifically, for RTP-VoIP-One-way-Packet-Loss, the following sources of error apply:

- The sequence number of an observed RTP stream may not be valid, i.e. not conforming with the RTP specifications [i.3]. This issue is particularly relevant for measurement methods of type B, C and D, where RTP senders are not necessarily known and trusted.
- Conceptually, type D passive measurement methods perform direct measurements along the transmission path. In practice, type D systems receive a copy of the traffic on a network link, e.g. through a network element's mirroring port connected to the passive measurement device. If the traffic copy is corrupted, then a type D system reports packet loss, which is not present in the observed RTP streams.

5.2.2 A definition for one-way packet loss timeslice samples

Metric name

RTP-VoIP-TS-One-way-Packet-Loss.

Metric parameters

- Src, the IP address of a sender of RTP-VoIP packets.
- Flow, a sequence of RTP-VoIP packets sent by Src belonging to an RTP session.
- Observer, a receiver of type RTP-VoIP packets.
- Ta, sample start time.
- D, fixed sample period.

Metric units

A sequence of pairs (S,L) with:

- S, an RTP sequence number, and
- L, either a zero or a one.

The values of S in the sequence are monotonically increasing and consecutive.

NOTE 1: S would be a valid parameter SeqNo to RTP-VoIP-One-way-Packet-Loss and that L would be a valid value of RTP-VoIP-One-way-Packet-Loss.

Definition

Given Ta and D, RTP-VoIP-One-way-Packet-Loss is determined for every packet of an RTP session Flow received by Observer within the interval [Ta,Ta+D). The value of the sample is the sequence composed by the resulting <sequence number, loss> pairs. If there are no such pairs, the sequence is of length zero and the sample is said to be empty.

Discussion

The timeslice sample metric summarizes the values of all possible singletons during a sample period. All the RTP-VoIP-One-way-Packet-Loss singletons in the sample have the same values of Src, Observer, and Flow.

To fully characterize an entire RTP stream requires a sequence of samples with consecutive intervals [Ta,Ta+D), [Ta+D,Ta+2D), [Ta+2D,Ta+3D) etc., accounting for all of the stream's packets and all possible singletons.

NOTE 2: The fixed sample period D corresponds to the chosen measurement method's timeslice duration.

The discussions on methodologies, errors and uncertainties for the singleton metric in clause 5.2.1 also apply to the sample metric.

5.2.3 Some statistics definitions for one-way packet loss

Given the sample metric RTP-VoIP-TS-One-way-Packet-Loss, statistics of that sample can be defined. The following statistics illustrate potential definitions.

RTP-VoIP-One-way-Packet-Loss-Count

Given an RTP-VoIP-TS-One-way-Packet-Loss, the sum of all the sample's L values is the number of packets lost in that timeslice. In addition, the RTP-VoIP-One-way-Packet-Count is undefined if the sample is empty.

RTP-VoIP-One-way-Packet-Loss-Ratio

Given an RTP-VoIP-TS-One-way-Packet-Loss, the average of a sample's L values is the ratio of losses to total packets in the timeslice. In addition, the RTP-VoIP-One-way-Packet-Loss-Ratio is undefined if the sample is empty.

RTP-VoIP-One-way-Burst-Loss-Event-N

Given an RTP-VoIP-TS-One-way-Packet-Loss, the RTP-VoIP-One-way-Burst-Loss-Event-N is 0 means that the timeslice has no sequence of at least N consecutive losses. The RTP-VoIP-One-way-Burst-Loss-Event-N is 1 means that the timeslice has at least one sequence of N or more consecutive losses.

For example, let a timeslice consist of an (S,L)-sequence as follows:

Timeslice = <(1,0), (2,1), (3,0), (4,1), (5,1), (6,1), (7,0), (8,0), (9,1), (10,1), (11,1), (12,0)>.

For this timeslice the RTP-VoIP-One-way-Burst-Loss-Event-1 is 1, the RTP-VoIP-One-way-Burst-Loss-Event-2 is 1 and the RTP-VoIP-One-way-Burst-Loss-Event-3 is 1 - all other burst loss event types are 0.

NOTE 1: This statistic requires a mechanism for carrying over information from the previous timeslice to account for cross-timeslice burst losses.

RTP-VoIP-One-way-Loss-Gap-N

Given an RTP-VoIP-TS-One-way-Packet-Loss, the RTP-VoIP-One-way-Loss-Gap-N is 0 means that the timeslice has no sequence of exactly N consecutively received packets between two loss events. The RTP-VoIP-One-way-Loss-Gap-N is 1 means that the timeslice has at least one sequence of exactly N consecutively received packets between two loss events.

For example, let a timeslice consist of an (S,L)-sequence as follows:

Timeslice = <(1,0), (2,1), (3,0), (4,1), (5,1), (6,1), (7,0), (8,0), (9,1), (10,1), (11,1), (12,0)>.

For this timeslice the RTP-VoIP-One-way-Loss-Gap-1 is 1 and the RTP-VoIP-One-way-Loss-Gap-2 is 1 - all other received event types are 0.

NOTE 2: This statistic requires a mechanism for carrying over information from the previous timeslice to account for cross-timeslice burst losses.

5.3 A one-way-IPDV timeslice metric

5.3.1 A singleton definition of a one-way-IPDV metric

The following defines a one-way Inter-Packet Delay Variation (IPDV) singleton, that is based on the specification of the respective generic IPPM metric [i.16], but exploits the characteristics of RTP-VoIP traffic to avoid the requirement for clock synchronization between sender and receiver. Subsequent clauses use this singleton definition to define sample metrics and statistics.

Metric name

RTP-VoIP-One-way-IPDV.

Metric parameters

- Src, the IP address of a sender of RTP-VoIP packets.
- Flow, a sequence of RTP-VoIP packets sent by Src belonging to an RTP session.
- Observer, a receiver of type RTP-VoIP packets.
- Pi, Pj, consecutive, same-size RTP-VoIP packets of Flow.
- SeqNo, the RTP sequence number of Pi.

Metric units

The value of an RTP-VoIP-One-way-IPDV is either a real number of seconds or an undefined number of seconds.

Definition

Consider an RTP-VoIP packet stream Flow sent by Src with consecutive packets Pi and Pj, such that the RTP sequence numbers differ by one, i.e. SeqNo(Pi) = SeqNo(Pj)-1. The packets are sent by Src at times Ts(i) and Ts(j) with a time difference of dTs(i,j) = Ts(j) - Ts(i). The packets are received by Observer at times Tm(i) and Tm(j) with an interarrival time of dTm(i,j) = Tm(j) - Tm(i).

The RTP-VoIP-One-way-IPDV from Src to Observer for Flow at SeqNo is ddT means that Src sent two successive packets Pi and Pj with time difference dTs(i,j) and that these packets were received in order by Observer with a time difference of dTm(i,j), that dTm(i,j) - dTs(i,j) = ddT.

The RTP-VoIP-One-way-IPDV from Src to Observer for Flow at SeqNo is undefined means that Observer did not receive one or both packets Pi, Pj or that dTm(i,j) < 0, i.e. packet reordering took place.

Figure 7 illustrates this definition.

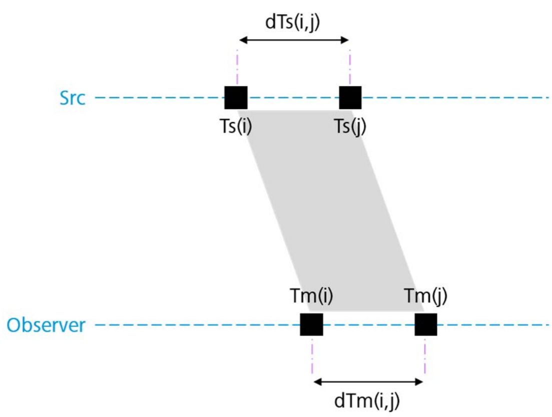

Figure 7: Illustration of the IPDV metric definition

Discussion

The following issues need to be considered:

- If an RTP packet arrives at Observer but is corrupted, then it is counted as lost and the IPDV is undefined.
- If an RTP packet is received two or more times by Observer, then the packet is counted as received and the earliest measurement is used to calculate dTm(i,j).
- The time difference dTs(i,j) between the sending of the two packets Pi and Pj is often constant for RTP-VoIP streams corresponding to the packet rate of typically 20 milliseconds.
- The size of a packet determines its serialization delay on network links and therefore also the one-way delay. The packets used to calculate *RTP-VoIP-One-way-IPDV* therefore need to be of the same size.

Methodologies

The metric definition is applicable to all timeslicing measurement types defined in clause 4.3.1, except for RTCP-based methods of type B. The reason is that RTCP does not provide information about packet interarrival times.

The RTP sequence number is 16 bit and its initial value is random [i.3]. Therefore, methodologies need to consider a potential wraparound of RTP sequence numbers, when determining the next packet to use for the metric calculation.

NOTE: Passive measurement methodologies (type B, C and D) do not explicitly know the time when a packet was sent. However, packets of type RTP-VoIP contain RTP timestamps, which reflect the sampling instant of the first octet in the RTP payload [i.17]. According to IETF RFC 3550 [i.3] the RTP timestamp increments monotonically and linearly in time and can therefore be used to calculate transmit time differences in seconds. The benefit of this approach is that clock synchronization between Src and Observer is not required.

Errors and uncertainties

The IPPM Framework IETF RFC 2330 [i.2] provides general guidance on sources of errors and uncertainties relating to measurements. Specifically, for RTP-VoIP-One-way-IPDV, the following sources of error apply:

- The sequence number of an observed RTP stream may not be valid, i.e. not conforming with the RTP specifications IETF RFC 3550 [i.3]. This issue is particularly relevant to measurement methods of type B, C and D, where RTP senders are typically unknown. The singleton definition is based on successive packets with consecutive sequence numbers, and the effect of invalid RTP sequence numbers therefore is obtaining an undefined IPDV.
- The RTP timestamp of an observed RTP stream may also be incorrect, and wrong timestamps lead to erroneous values for the IP packet delay variation. Given that the intended packet rate of RTP-VoIP streams is often constant, it is possible to avoid wrong IPDV measurements by disregarding packets with inconsistent timestamps.
- While clock synchronization between Src and Observer is not required, clock drift is a potential error source. Measurement errors can build up over time, if the actual packet rate of Src determined by dTs systematically deviates from the rate indicated by the RTP timestamps.
- Conceptually, type D passive measurement methods perform direct measurements along the transmission path. In practice, type D systems receive a copy of the traffic on a network link, e.g. through a network element's mirroring port connected to the passive measurement device. If the traffic copy does not exhibit the same inter-packet timing, then a type D system will report wrong IPDV measurements.

5.3.2 A definition for one-way IPDV timeslice samples

Metric name

RTP-VoIP-TS-One-way-IPDV.

Metric parameters

- Src, the IP address of a sender of RTP-VoIP packets.
- Flow, a sequence of RTP-VoIP packets sent by Src belonging to an RTP session.
- Observer, a receiver of type RTP-VoIP packets.
- Ta, sample start time.
- D, fixed sample period.

Metric units

A sequence of pairs (S,ddT) with:

- S, an RTP sequence number, and
- ddT, a real number or an undefined number of seconds.

The values of S in the sequence are monotonically increasing and consecutive.

NOTE 1: S would be a valid parameter SeqNo to RTP-VoIP-One-way-IPDV and that ddT would be a valid value of RTP-VoIP-One-way-IPDV.

Definition

Given Ta and D, RTP-VoIP-One-way-IPDV is determined for every pair of consecutive packets of an RTP session Flow received by Observer within the interval [Ta,Ta+D). The value of the sample is the sequence composed by the resulting <sequence number, IPDV> pairs. If there are no such pairs, the sequence is of length zero and the sample is said to be empty.

Discussion

The timeslice sample metric summarizes the values of all possible singletons during a sample period. All the RTP-VoIP-One-way-IPDV singletons in the sample have the same values of Src, Observer, and Flow.

To fully characterize an entire RTP stream typically requires a sequence of samples with consecutive sample intervals [Ta,Ta+D), [Ta+D,Ta+2D), [Ta+2D,Ta+3D) etc. accounting for all of the stream's packets and all possible singletons.

NOTE 2: The fixed sample period D corresponds to the chosen measurement method's timeslice duration.

The discussions on methodologies and errors and uncertainties for the singleton metric in clause 5.3.1 also apply to the sample metric.

5.3.3 Some statistics definitions for one-way IPDV

Given the sample metric RTP-VoIP-TS-One-way-IPDV, statistics of that sample can be defined. The following statistics illustrate potential definitions.

RTP-VoIP-One-way-IPDV-Min

Given an RTP-VoIP-TS-One-way-IPDV, the minimum of all the sample's ddT values is the RTP-VoIP-One-way-IPDV-Min of that timeslice. In addition, the RTP-VoIP-One-way-IPDV-Min is undefined if the sample is empty.

RTP-VoIP-One-way-IPDV-Max

Given an RTP-VoIP-TS-One-way-IPDV, the maximum of all the sample's ddT values is the RTP-VoIP-One-way-IPDV-Max of that timeslice. In addition, the RTP-VoIP-One-way-IPDV-Max is undefined if the sample is empty.

RTP-VoIP-One-way-IPDV-Avg

Given an RTP-VoIP-TS-One-way-IPDV, the average of all the sample's ddT values is the RTP-VoIP-One-way-IPDV-Avg of that timeslice. In addition, the RTP-VoIP-One-way-IPDV-Avg is undefined if the sample is empty.

RTP-VoIP-One-way-IPDV-BufferUnderrunEvent-N

Endpoints of RTP-VoIP communication typically use a dejitter buffer to smooth out expected packet delay variations and ensure continuous playout. Packet delay variations exceeding a receiver's dejitter buffer cause gaps in the decoded speech and are therefore of particular interest for analysis of speech transmission performance.

Given an RTP-VoIP-TS-One-way-IPDV, the RTP-VoIP-One-way-IPDV-BufferUnderrunEvent-N is 0 means that the timeslice has no IPDV of more than N milliseconds. The RTP-VoIP-One-way-IPDV-BufferUnderrunEvent-N is 1 means that the timeslice has at least one IPDV of N milliseconds or more. In addition, the RTP-VoIP-One-way-IPDV-BufferUnderrunEvent-N is undefined if the sample is empty.

For example, let a timeslice consist of an (S,ddT)-sequence as follows:

Timeslice = <(1,0), (2,10), (3,-10), (4,40), (5,-20), (6,-20), (7,60), (8,-20), (9,-20), (10,-20), (11,0)>.

For this timeslice the RTP-VoIP-One-way-IPDV-BufferUnderrunEvent-40 is 1 and the RTP-VoIP-One-way-IPDV-BufferUnderrunEvent-60 is 1 - all other buffer underrun event types are 0.

NOTE: This statistic requires a mechanism for carrying over information from the previous timeslice to account for cross-timeslice IPDV measurements.

RTP-VoIP-One-way-IPDV-Alternation

One-way IPDV timeslice samples can also be used to detect specific patterns of packet delay variations and the following provides one example.

Given an RTP-VoIP-TS-One-way-IPDV, the RTP-VoIP-One-way-IPDV-Alternation is 1 means that all of the sample's absolute IPDV values are constant (or within defined boundaries) and consecutive values alternate between positive and negative values. The RTP-VoIP-One-way-IPDV-Alternation is 0 means that the absolute IPDV values are not constant or consecutive values do not alternate between positive and negative values. In addition, the RTP-VoIP-One-way-IPDV-Alternation is undefined if the sample has less than two consecutive elements.

For example, let a timeslice consist of an (S,ddT)-sequence as follows:

Timeslice = <(1,20), (2,-20), (3,20), (4,-20), (5,20), (6,-20), (7,20), (8,-20), (9,20), (10,-20)>.

For this timeslice the RTP-VoIP-One-way-IPDV-Alternation is 1.

5.4 A one-way packet duplication timeslice metric

5.4.1 A singleton definition of a one-way packet duplication metric

The following defines a packet duplication singleton that is based on the specification of the respective generic IPPM metric [i.18]. The singleton provides a count of received packet copies, which can be used to determine the amount of duplicates. Subsequent clauses use the singleton to define different sample metrics and statistics.

Metric name

RTP-VoIP-One-Way-Received-Packet-Count.

Metric parameters

- Src, the IP address of a sender of RTP-VoIP packets.
- Observer, a receiver of type RTP-VoIP packets.
- Flow, a sequence of RTP-VoIP packets sent by Src belonging to an RTP session.
- SeqNo, the sequence number contained in the header of an RTP-VoIP packet.
- L, lowest protocol layer considered for packet comparison.

Metric units

The value of an RTP-VoIP-One-Way-Received-Packet-Count is a natural number including zero, signifying the number of times a packet has been received.

Definition

For the purpose of this metric two packets are considered identical, if and only if, the packets are bitwise identical down to and including layer L of the protocol stack. Such duplicates are called L-copies in the following. Corrupted packets are not counted as L-copies.

The *RTP-VoIP-One-Way-Received-Packet-Count from Src to Observer for Flow at SeqNo is 0* means that Observer did not receive an RTP-VoIP packet of session Flow with sequence number SeqNo. The *RTP-VoIP-One-Way-Received-Packet-Count from Src to Observer for Flow at SeqNo is n* means that Observer received n L-copies of an RTP-VoIP packet of session Flow with sequence number SeqNo.

Discussion

This singleton counts the number of identical packets - in the sense of L-copies - with sequence number SeqNo received by Observer. If the singleton is zero then no packet with this sequence number was observed, which corresponds to a packet loss.

The metric definition is very similar to the one-way packet loss singleton definition and all considerations regarding methodologies and errors from the clause 5.2.1 apply.

5.4.2 A definition for one-way packet duplication timeslice samples

Metric name

RTP-VoIP-TS-Packet-Duplication.

Metric parameters

- Src, the IP address of a sender of RTP-VoIP packets.
- Flow, a sequence of RTP-VoIP packets sent by Src belonging to an RTP session.
- Observer, a receiver of type RTP-VoIP packets.
- Ta, sample start time.
- D, fixed sample period.

Metric units

A sequence of pairs (S,C) with:

- S, an RTP sequence number; and
- C, a natural number including zero.

The values of S in the sequence are monotonically increasing and consecutive.

NOTE: S would be a valid parameter SeqNo to RTP-VoIP-One-Way-Received-Packet-Count and that C would be a valid value of RTP-VoIP-One-way-Received-Packet-Count.

Definition

Given Ta and D, RTP-VoIP-One-way-Received-Packet-Count is determined for every packet of an RTP session Flow received by Observer within the interval [Ta,Ta+D). The value of the sample is the sequence composed by the resulting <sequence number, packet count> pairs. If there are no such pairs, the sequence is of length zero and the sample is said to be empty.

Discussion

The timeslice sample metric summarizes the values of all possible singletons during a sample period. All the RTP-VoIP-One-way-Received-Packet-Count singletons in the sample have the same values of Src, Observer, and Flow.

The discussions on methodologies, errors and uncertainties for the singleton metric defined in clause 5.2.1 also apply to this sample metric.

5.4.3 Some statistic definitions for packet duplication

Given the sample metric RTP-VoIP-TS-Packet-Duplication, statistics of that sample can be defined. The following statistics illustrate potential definitions.

RTP-VoIP-One-way-Duplicate-Count

This statistic provides the number of packet duplicates observed in the timeslice.

Given an RTP-VoIP-TS-Packet-Duplication, one sums up all the sample's C values larger than one to obtain the RTP-VoIP-One-way-Duplicate-Count of that timeslice. (If C is zero, the packet is lost, if C is one, only one copy was observed.) In addition, the RTP-VoIP-One-way-Duplicate-Count is undefined if the sample is empty.

RTP-VoIP-One-way-Replicated-Packet-Rate

This statistic calculates the fraction of packets, which were observed more than once.

Given an RTP-VoIP-TS-Packet-Duplication, one counts the number of C values larger than one. This number is divided by the number of C values larger than zero to obtain the RTP-VoIP-One-way-Replicated-Packet-Rate of that timeslice. In addition, the RTP-VoIP-One-way-Replicated-Packet-Rate is undefined if the sample is empty.

RTP-VoIP-One-way-Duplication-Event

This statistic determines if any packet duplication took place in a timeslice.

Given an RTP-VoIP-TS-Packet-Duplication, the RTP-VoIP-Packet-Duplication-Event is one, means that at least one of the sample's C values is larger than one. The RTP-VoIP-Packet-Duplication-Event is zero, means that all of the sample's C values are at most one. In addition, the RTP-VoIP-Packet-Duplication-Event is undefined if the sample is empty.

5.5 More example timeslice metrics

5.5.1 Overview

Clauses 5.2 to 5.4 provide examples of key metrics that illustrate the basic concept of RTP timeslice metrics. Metrics for packet reordering, delay and other IPPM-type metrics can be defined in a similar way.

Apart from these generic network-focused metrics, VoIP service monitoring also requires metrics to describe traffic characteristics more specifically related to RTP and its application to real-time speech transmission. The following clauses briefly discuss examples of such metrics.

5.5.2 Codec timeslice metric

One important factor in determining the user experience of VoIP services is the codec used to encode speech. This is not a static property of an RTP-VoIP stream, but can change with every packet. A codec singleton therefore requires to inspect every packet of an RTP-VoIP stream to determine the used encoding. A timeslice sample for such a codec singleton then consists of a sequence of pairs of sequence number and codec type.

Based on this the following codec timeslice statistics can be defined.

RTP-VoIP-Codec-List

This statistic provides a list of distinct codecs used during a timeslice and the number of packets per codec (or the duration of encoded speech carried by these packets).

RTP-VoIP-Codec-Change-Event

This statistic indicates whether more than one codec was used during the timeslice. RTP-VoIP-Codec-Change-Event is 0, means that the same codec was used to encode speech in all packets covered by the timeslice. RTP-VoIP-Codec-Change-Event is 1, means that at least two distinct codecs were used during the timeslice.

5.5.3 Conformance timeslice metrics

The conformance to standards and service policies is one important way to characterize RTP-VoIP packet flows. While typical transmission performance metrics mainly reflect the dynamic behaviour of the network and service, conformance metrics put a focus on the configuration of the RTP-VoIP sender and active network elements along the transmission path.

The following provides some examples for timeslice metrics relating to conformance.

RTP-VoIP-DSCP-Class-X

This statistic states whether an RTP-VoIP packet of the timeslice was marked with the DSCP class X. RTP-VoIP-DSCP-Class-X is one means that at least one RTP-VoIP packet in the sample period used DSCP class X. RTP-VoIP-DSCP-Class-X is zero means that no RTP-VoIP packet in the sample period used DSCP class X.

RTP-VoIP-Silence-Suppression

This statistic indicates whether silence suppression was active during a timeslice. Silence suppression can be detected by a singleton that considers RTP header timestamps of successive packets. The transmission of RTP packets was discontinued if the increase in timestamp ticks of two successive packets is larger than the duration of the second packet's audio payload.

RTP-VoIP-Silence-Suppression is one means that packet transmission was discontinued at some point during the timeslice. RTP-VoIP-Silence-Suppression is zero means that packet transmission was continuous during the timeslice.

RTP-VoIP-Sequence-Number-Reset

This statistic indicates whether the sequence numbers of RTP-VoIP packets of a timeslice failed to continuously increase as required by IETF RFC 3550 [i.3]. RTP-VoIP-Sequence-Number-Reset is one means that second packet of a pair of successive packets during a timeslice had a sequence number of zero, excluding the cases of wrap-around and packet reordering. RTP-VoIP-Sequence-Number-Reset is zero means that the timeslice has no pair of successive packets where the sequence number of the second packet is zero, excluding wrap-around and reordering.

NOTE: This definition is a special case of backward jumping in RTP-VoIP stream sequence numbers.

5.5.4 Service quality timeslice metrics

Information about decoded speech segments, the used codec, packet loss, jitter and other relevant data can be used to calculate complex metrics characterizing the overall network performance or user experience for a timeslice. These complex metrics are composed from timeslice samples for different singleton metrics. The following provides examples for such complex timeslice metrics.

RTP-VoIP-Critical-Network-Performance

This statistic provides a high-level assessment of the network's ability to forward packets of an RTP-VoIP stream timeslice in a timely fashion.

For an RTP-VoIP timeslice RTP-VoIP-Critical-Network-Performance is 0 means that the timeslice's RTP-VoIP-One-way-Burst-Loss-Event-2 is 0 and the timeslice's RTP-VoIP-One-way-IPDV-BufferUnderrunEvent-40 is 0. RTP-VoIP-Critical-Network-Performance is 1 means that either the timeslice's RTP-VoIP-One-way-Burst-Loss-Event-2 is 1 or the timeslice's RTP-VoIP-One-way-IPDV-BufferUnderrunEvent-40 is 1.

In effect, this timeslice statistic states if packet burst loss or excessive packet delay variation occurred, both of which can lead to gaps in speech.

RTP-VoIP-One-Way-MOS$_{LQO}$

This metric provides an objective Mean Opinion Score (MOS) for a timeslice, for example according to Recommendation ITU-T P.863 [i.19]. This means that the MOS prediction is performed using a type A active timeslicing measurement method, but only for the RTP stream segment covered by the timeslice. Care needs to be taken, that the specific requirements for using a given method, e.g. sample selection and sample length, are met for each timeslice.

RTP-VoIP-One-Way-MOS$_{LQE}$

This metric provides a MOS listening quality estimate for a timeslice, e.g. based on the E-Model defined in Recommendation ITU-T G.107 [i.20] or Recommendation ITU-T G.107.2 [i.21]. The factors contributing to the R-factor, which is then mapped to a MOS$_{LQE}$ value, are determined by sample statistics for the timeslice. Examples of statistics defined in clause 5 that are relevant for the R-factor calculation include:

- RTP-VoIP-Codec-List - the used codec(s) define the maximum R-factor that can be achieved.
- RTP-VoIP-One-way-Burst-Loss-Event-N - various forms of packet burst loss are key contributors to the transmission impairment factor.
- RTP-VoIP-One-way-IPDV-BufferUnderrunEvent-N - packets that arrive too early or too late lead to gaps in speech output and thus jitter buffer underrun and overflow events are relevant to the transmission impairment factor.

The actual calculation of the MOS$_{LQE}$ depends on the specific timeslice measurement method and available sample statistics.

6 Timeslice metric aggregation and composition

6.1 Overview

Existing VoIP service metrics often aim to summarize and characterize entire calls. While this approach can provide relevant information, it also has some drawbacks:

- Loss of temporal detail: the temporal distribution of events, such as packet loss, greatly influences the resulting user experience. It makes a difference if the loss of 200 packets is evenly spread across a ten-minute call or if the last 200 packets are lost.
- Lack of comparability: it is not well-defined how to compare media quality metrics of calls with different durations. For example, given a MOS for a 10-second call and a MOS for a five-minute call, one cannot state with confidence that one is better than the other, because the scores relate to different durations.
- Problematic aggregation: elementary measurements are not sufficient to characterize an entire service or network and often some form of aggregation is needed to create high-level statistics for a specific purpose. Using per-call metrics in this context can be problematic, because each call is given the same weight, regardless of its duration.

Particularly passive (one-sided, in-service) methods observing live traffic are affected by these drawbacks. Active test campaigns can circumvent this by defining uniform call durations, yet the fundamental problems remain.

The use of fixed duration sample intervals facilitates the creation of meaningful statistics through aggregation and composition. Timeslice sample metrics, such as the examples defined in clause 5, provide metrics for fixed duration sample intervals, which can be aggregated in a natural way to generate statistics and Key Performance Indicators (KPIs) for sets of RTP-VoIP traffic. The traffic may relate to different aggregation dimensions and metrics serve to characterize individual calls, sets of calls exchanged between interconnection partners or an entire telephony service, to name some examples.

The following clauses discuss the composition and aggregation of timeslice metrics, provide examples of aggregation dimensions and analyse the impact of a measurement method's timeslice duration on the aggregation results.

6.2 Aggregation and composition

6.2.1 IPPM aggregation framework

The uniform duration of samples provided by a timeslice measurement method simplifies aggregation and composition to high-level metrics and statistics. IETF RFC 5835 [i.8] describes a framework of generic composition and aggregation mechanisms for the IPPM.

Specifically, IETF RFC 5835 [i.8] describes the following mechanisms:

- Temporal aggregation is defined as *the composition of metrics with the same type and scope obtained in different time instants or time windows*. For example, given a series of RTP-VoIP-One-way-IPDV-Max values (see clause 5.3.3) representing all timeslices of an RTP-VoIP stream at a measurement point, the series' maximum is the maximum IPDV for the entire stream at that point in the network.

- Spatial aggregation is defined as t*he combination of metrics of the same type and different scope*. According to [i.8] the purpose is to estimate the overall performance of a network, although the concept is also applicable to smaller spatial scopes. For example, consider the geo-redundant interconnection between two VoIP service providers. Spatial aggregation can be used to characterize the overall performance of the interconnection, by composing measurements of RTP-VoIP flows traversing each of the points of interconnection.

- Spatial concatenation is defined as *the composition of metrics of same type with (ideally) different spatial scope, so that the resulting metric is representative of what the metric would be if obtained with a direct measurement over the sequence of the several spatial scopes*. For example, a path between two call parties in a network can be sub-divided into different segments. Given measurements of the maximum one-way packet loss in each segment, the sum of these maximums is the maximum one-way loss for the network path between the calling parties. Spatial concatenation also applies to concatenation of the two media directions of a call to a two-way call summary.

Due to the short duration of timeslice samples, temporal aggregation is needed to assess the performance over typical observation periods. Indeed, any composed metric based on timeslice samples will most certainly require at least temporal aggregation to be of practical use. This composition is trivial since the samples provided by a timeslice measurement method have the same duration. Furthermore, the samples are consecutive and using temporal aggregation allows to comprehensively characterize entire RTP streams, VoIP calls or statistic periods.

6.2.2 Contextual aggregation

The defining property of timeslicing is that sample durations are short and uniform, creating basic building blocks to generate statistics. This property allows to define a generic aggregation mechanism that is purely based on the context or semantics of the timeslices of interest. Technically, it makes no difference whether timeslices are aggregated over space, time or some other characteristic.

Contextual aggregation is defined as the composition of timeslice metrics within a context of interest. A context may span different time instants or time windows and different measurement points or network paths under observation. A context may also include specific characteristics of timeslices, e.g. used codecs, DSCP classes or source addresses.

For example, given a set of RTP-VoIP-One-way-IPDV-Max values (see clause 5.3.3) representing the maximum IPDV for each timeslice of RTP-VoIP flows at two measurement points in a given time period, the maximum is the maximum IPDV for this set of timeslices. Another example is the packet loss rate of RTP-VoIP timeslices that used best effort forwarding on a given day.

6.2.3 Higher-order composition

More complex timeslice statistics require what [i.8] refers to as higher-order composition, meaning further aggregation or composition of composed timeslice metrics. The first-order composed timeslice metric is typically a temporal aggregate of timeslices to obtain information on relevant statistical periods. Higher-order composed metrics are calculated by applying any composition mechanism to the set of composed metrics.

For example, data on individual RTP streams is calculated using temporal aggregation of the streams' timeslices observed at a measurement point. In a second step, temporal aggregation is applied a second time, e.g. to obtain a statistic on the performance of all RTP streams that ended in a 5-minute interval. Another example for a second composition step, is to aggregate RTP stream metrics from different measurement points using spatial aggregation.

6.3 Some examples for aggregation

6.3.1 RTP flow aggregation

RTP flow aggregation, i.e. the aggregation of a specific metric of all timeslices associated with an RTP flow, is the simplest and most obvious form of aggregation. It can be defined in the form of temporal aggregation of timeslice metrics spanning the period from the first packet of an RTP flow to its last packet, measured at a single measurement point.

Aggregating a specific metric of all timeslices associated with an RTP flow serves to characterize an RTP flow as a whole. For example, given a series of RTP-VoIP-One-way-Packet-Loss-Count values (see clause 5.2.3) representing the packet loss counts for all timeslices of an RTP-VoIP stream at a measurement point, the sum is the packet loss for the entire stream at that point in the network. Another example is the average MOS_{LQO} of an RTP-VoIP stream, calculated as the average of all RTP-VoIP-One-Way-MOS_{LQO} (see clause 5.5.4) values associated with an RTP flow.

In theory sample statistics yielding ratios can also be aggregated as a method's timeslices are of the same duration. However, the number of singletons per timeslice can still vary - see clause 4.4 - and aggregating absolute values should be preferred to reduce errors.

6.3.2 Call aggregation

Call aggregation is the aggregation of timeslice metrics associated with RTP-VoIP flows of a call with the aim to characterize the call. This type of aggregation is specifically relevant to call detail records.

Call aggregation can take different forms, depending on the measurement method, timeslice metric and desired aggregate metric. Call aggregation needs to consider the following aspects:

- Calls typically involve two parties, both sending and receiving RTP-VoIP flows. It may be useful to aggregate the timeslice metrics of both media directions of a call into one value. For example, [i.22] defines the aggregate speech quality of a call as the minimum MOS_{LQO} for each call party. Similar call aggregate definitions for speech quality and other timeslice metrics are conceivable.

- The media sent from one call party to the other might be split up into several consecutive RTP-VoIP flows, e.g. due to early media. A call aggregate needs to define if it considers all RTP-VoIP flows or only a subset, such as the last flow.

- RTP-VoIP flows of a call might be measured at multiple measurement points, e.g. by a type D measurement method. A call aggregate needs to define which measurement points to consider for an RTP-VoIP flow.

Figure 8 illustrates the impact of the measurement points on sample statistics and thus on the estimated user experience. An impairment source between two measurement points impacts the RTP-VoIP flows from a call's A-party to the B-party. The performance degradation is only visible in the sample statistics downstream from the impairment source, The measurement closest to an RTP receiver is most relevant to estimating the quality of experience. In figure 8, the sample statistics from measurement point 1 are most relevant for the assessing the user experience of the A-party and the statistics from measurement point 2 for the B-party.

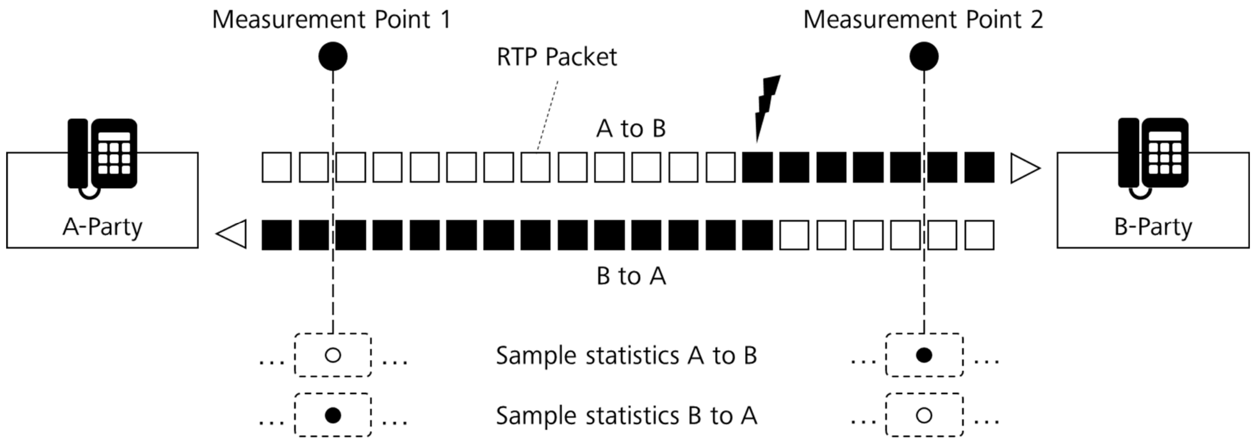

Figure 8: Impact of measurement points on sample statistics

Call aggregation requires higher-order composition, when aggregate metrics are derived from multiple RTP-VoIP flows and multiple measurement points. For example, the total number of packets lost in the context of a call is given by the sum of the series of RTP-VoIP-One-way-Packet-Loss-Count values (see clause 5.2.3) of all RTP-VoIP streams associated with the call measured closest to the receiving party.

6.3.3 Trunk aggregation

Trunk aggregation is the aggregation of timeslice metrics associated with a defined set of RTP-VoIP flows measured at one point in the network. Key performance indicators generated using this aggregation method typically serve to characterize the traffic between network elements or network segments.

Trunks are defined in different ways. Examples are:

- RTP-VoIP trunks, i.e. communication relations between trunk endpoints sending and receiving RTP-VoIP flows. Trunk endpoints are defined by tuples of IP addresses or network masks, possibly complemented by other information such as ports and VLANs.
- SIP-VoIP trunks, i.e. communications relations between endpoints sending and receiving SIP-based call signalling. Trunk endpoints are defined by tuples of IP addresses or network masks, possibly complemented by other information such as ports and VLANs. Timeslice metrics for a SIP-VoIP trunk relate to the RTP-VoIP flows associated with the calls in the context of the trunk.

Trunk statistics are generated using temporal aggregation of timeslice metrics relating to a trunk. The definition of a specific statistic needs to specify the time window for temporal aggregation and what this time window refers to. Temporal aggregation in the strict sense only includes timeslices fully within a statistic's time window. Alternative approaches include composing timeslices of RTP-VoIP flows starting within a statistic's time window or ending within the window. Furthermore, the definition of a trunk statistic needs to state the direction of RTP-VoIP flows relative to the trunk endpoints.

Trunk aggregation serves to create performance statistics for potentially large amounts of RTP-VoIP traffic, such as for RTP-VoIP flows observed at an interconnection between two carriers ending within a 15-minute time window.

6.4 Impact of timeslice duration on aggregation results

The choice of timeslice duration and metric has an impact on the resulting statistics. If one considers a statistic that determines the maximum RTP-VoIP-One-way-IPDV for the timeslices of an RTP flow, the resulting maximum is independent of the timeslice duration. This is not true for a statistic defined as the ratio of all timeslices for which RTP-VoIP-One-way-Burst-Loss-Event-1 is true (see clause 5.2.3) over all timeslices. The impact of different timeslice durations on the statistic value is illustrated in Figure 9.

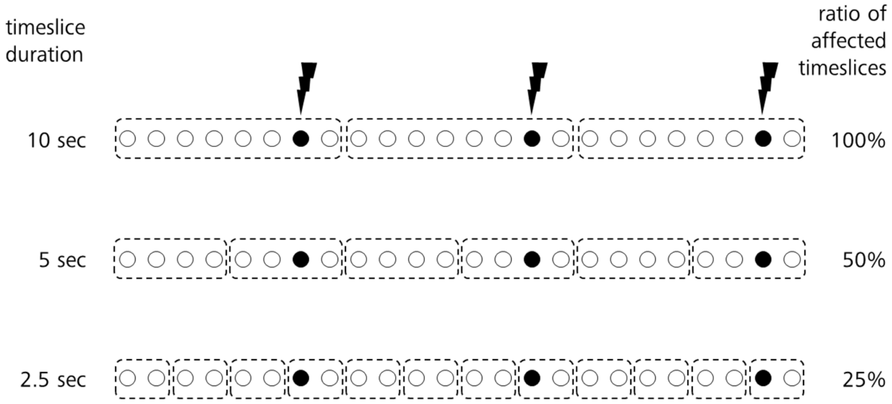

Figure 9: Impact of timeslice durations on aggregation results

Figure 9 shows the impact of different timeslice durations on timeslice statistics pertaining to an RTP-VoIP flow that is impacted by packet loss every ten seconds. The timeslice metric indicates 1 if packet loss occurred and 0 if not. A timeslicing method with 10 seconds slices will report 100 % of timeslices affected by packet loss, a timeslice duration of five seconds will yield a ratio of 50 % and a duration of 2,5 seconds a ratio of 25 %.

The ratio becomes smaller as the timeslice duration decreases. In the most extreme case, it corresponds to the packet loss ratio. This shows that timeslice statistics need to state the measurement method type as well as the timeslice duration for context.

Annex A:
Example timeslice KPIs for RTP-based speech transmission

A.1 Measurement system

The following applies the defined framework to describe a set of timeslice KPIs provided by a commercially available type D-5 timeslice measurement system.

The system has the following properties:

- It is a passive midpoint monitoring system with no information about the endpoints of RTP-VoIP communication. The hypothetical endpoint model assumes a static dejitter buffer of 40 ms.
- All RTP-VoIP flows observed at a measurement point are sliced simultaneously with a 5 seconds sample duration.
- The measurement system performs a carry-over across timeslices of information relating to multi-packet singleton metrics. Singleton metrics with carry-over information are attributed to the timeslice with the last RTP-VoIP packet contributing to a singleton.
- Speech quality in terms of MOS is estimated using the E-Model as described in Recommendation ITU-T G.107 [i.20] and Recommendation ITU-T G.107.2 [i.21].
- Call aggregation uses the last RTP-VoIP flows for each call direction measured closest to the respective RTP-VoIP receiver.
- Timeslice trunk statistics compose timeslices of RTP-VoIP flows ending within a statistic's time window.

Sources of systematic errors and uncertainties are incomplete samples (clause 4.4.1) and varying singleton counts (clause 4.4.2). Errors caused by sliced singletons (clause 4.4.3) and incomplete stream slicing (clause 4.4.4) are avoided through the carry-over mechanism and direct measurement of RTP-VoIP flows.

A.2 Basic definitions

The following defines the terms critical timeslice and good timeslice, which are used to define the timeslice KPIs.

Critical timeslice

A timeslice is considered critical if the timeslice statistics fulfil at least one of the following conditions:

- The timeslice statistic RTP-VoIP-One-way-Burst-Loss-Event-N is 1 for N=3 (see clause 5.2.3).
- The timeslice statistic RTP-VoIP-One-way-Loss-Gap-N is 1 for N=1 or N=2 (see clause 5.2.3).
- The timeslice statistic RTP-VoIP-One-way-IPDV-Max is greater than 40ms (see clause 5.3.3).

In essence, a timeslice is considered critical if it is subject to some form of substantial packet loss or jitter that likely leads to packet loss.

Good timeslice

A timeslice is considered *good* if the timeslice statistic RTP-VoIP-One-Way-MOS_{LQE} (see clause 5.5.4) is greater than 4.0. Note, that the speech bandwidth and its corresponding MOS scale need to be provided for proper context, e.g. by adding a subscript N (narrowband), W (wideband), S (super-wideband) or F (fullband) bandwidth denominator as suggested in Recommendation ITU-T P.800.1 [i.23].

A.3 Critical Minute Ratio

The Critical Minute Ratio (CMR) is used to measure the technical quality of RTP-VoIP traffic. A high CMR indicates problems with the network's transport performance, impacting many timeslices with packet loss or severe inter-packet delay variations.

The CMR is calculated using the following formula:

$$CMR\ [\%] = \frac{\#\, critical\ timeslices}{\# total\ timeslices} \text{x } 100$$

It is possible to compute the CMR for any collection of RTP-VoIP timeslices, since the KPI is defined directly on timeslice statistics. Typically, the CMR is calculated for RTP-VoIP flows, calls and trunks.

A.4 Critical Stream Ratio

A critical stream is an RTP-VoIP flow with at least one critical timeslice, i.e. *critical stream* is technically an RTP flow aggregate. The Critical Stream Ratio (CSR) is the ratio of the number of critical streams to the total number of streams in a set of RTP-VoIP flows. It complements the CMR by measuring the distribution of critical impairments over a set of RTP-VoIP flows.

The CSR is calculated as follows:

$$CSR\ [\%] = \frac{\#\, critical\ RTPVoIP\ streams}{\# total\ RTPVoIP\ streams} \text{x } 100$$

The CSR is measured for sets of RTP-VoIP flows, i.e. typically for trunks.

A.5 Critical Call Ratio

A critical call is a call which has at least one associated RTP-VoIP flow subject to call aggregation that is critical. This means that *critical call* is a call aggregate. The Critical Call Ratio (CCR) is the ratio of the number of critical calls to the total number of calls in a set of VoIP calls.

The CCR is calculated using the following formula:

$$CCR\ [\%] = \frac{\#\, critical\ calls}{\# total\ calls} \text{x } 100$$

The CCR is useful for assessing the technical quality of a set of calls.

A.6 Good Minute Ratio

The Good Minute Ratio (GMR) is a measure for the (estimated) speech quality of RTP-VoIP traffic. A high GMR is an indication for satisfactory user experience.

The GMR is calculated using the following formula:

$$GMR\ [\%] = \frac{\#\, good\ timeslices}{\# total\ timeslices} \text{x } 100$$

It is possible to compute the GMR for any collection of RTP-VoIP timeslices, since the KPI is defined directly on timeslice statistics. Typically, the GMR is calculated for RTP-VoIP flows, calls and trunks.

A.7 Good Stream Ratio

A good stream is an RTP-VoIP flow with only good timeslices, i.e. *good stream* is technically an RTP flow aggregate. The Good Stream Ratio (GSR) is the ratio of the number of good streams to the total number of streams in a set of RTP-VoIP flows. It complements the GMR by measuring the distribution of streams with satisfactory user experience over a set of RTP-VoIP flows.

The GSR is calculated as follows:

$$GSR\ [\%] = \frac{\#\ good\ RTPVoIP\ streams}{\#total\ RTPVoIP\ streams} \text{x}\ 100$$

The GSR is measured for sets of RTP-VoIP flows, i.e. typically for trunks.

A.8 Good Call Ratio

A good call is a call where all associated RTP-VoIP flows subject to call aggregation are good. This means that *good call* is a call aggregate. The Good Call Ratio (GCR) is the ratio of the number of good calls to the total number of calls in a set of VoIP calls.

The GCR is calculated using the following formula:

$$GCR\ [\%] = \frac{\#\ good\ calls}{\#total\ calls} \text{x}\ 100$$

The GCR is useful for expressing the user satisfaction with speech quality of a set of calls.

History

Document history		
V1.1.1	March 2021	Publication