voipfuture

WHITEPAPER

2017

Why Next-Generation Voice Services Need Next-Generation KPIs

There are quick and simple answers when it comes to measuring and monitoring the quality of IP-based voice services. Conventional wisdom suggests to use the KPIs defined in RFC 6076 to monitor signaling and the Mean Opinion Score (MOS) for the media quality. In reality the problem is more complex, particularly when trying to assess the media quality for groups of calls, e.g. for trunks or specific services like VoLTE and VoWifi.

Wendenstr. 4
20097 Hamburg
Germany

+49 40 688 90 01 0
info@voipfuture.com
www.voipfuture.com

The key performance indicators (KPI) defined in IETF's RFC 6076 provide a means to assess a network's SIP signaling performance, i.e. its general ability to set up and tear down calls, the ability of users' to register, the corresponding delays, call durations, etc. Many of the KPIs defined in the RFC adapt well-known metrics from traditional telecommunications to the SIP-based world of voice over IP (VoIP). For example, the Session Establishment Effectiveness Ratio (SEER) is a SIP-version of the old Network Efficiency Ratio (NER) defined by the ITU.

It is safe to say that the RFC 6076 KPIs define the de-facto standard in terms of control plane monitoring for VoIP services and Voipfuture Qrystal supports all of them.

4. SIP Performance Metrics ....
4.1. Registration Request Delay (RRD) ....
4.2. Ineffective Registration Attempts (IRAs) ....
4.3. Session Request Delay (SRD) ....
4.3.1. Successful Session Setup SRD ....
4.3.2. Failed Session Setup SRD ....
4.4. Session Disconnect Delay (SDD) ....
4.5. Session Duration Time (SDT) ....
4.5.1. Successful Session Duration SDT ....
4.5.2. Failed Session Completion SDT ....
4.6. Session Establishment Ratio (SER) ....
4.7. Session Establishment Effectiveness Ratio (SEER)
4.8. Ineffective Session Attempts (ISAs) ....
4.9. Session Completion Ratio (SCR) ....

Key performance Indicators (KPI) as defined in IETF's RFC 6076

What about voice quality?

The media plane quality somehow has to do with the MOS, but it cannot be easily applied to VoIP service monitoring. The MOS is originally an empirical measurement of subjective voice quality. The ITU-T Recommendation P.800 defines a procedure for how 'subjects', i.e. real people, should score the quality of a conversation or an audio stream on a scale of 1 ('Bad') to 5 ('Excellent'). The average of a series of tests with different subjects constitutes the MOS. The entire process is well-defined down to seemingly odd details, such as the decoration of the cabinets where the subjects should be seated.

Passive monitoring systems try to estimate this empirical MOS for actual phone calls. Unfortunately, the process of this estimation is not so well-defined. For example, P.800 suggests that listening tests should be done using audio samples with a duration between 4 and 15 seconds. How can this be transferred to the automatic analysis of calls which last minutes or even hours? The Voipfuture whitepaper 'MOS Calculation & Aggregation' describes how we address this question through our fixed time slicing technology.

The Voipfuture media KPI system

Based on the concept of time slicing, which creates atomic units of quality, this whitepaper presents a set of media plane KPIs that are provided by the Voipfuture Qrystal system. These KPIs are being used today by tier-1 communication service providers to monitor the media plane quality. They do not just cover the user experience expressed through the MOS, but also the networks' transport quality.

A recap: time slicing the media plane

Voipfuture's unique technology calculates highly accurate listening quality estimates for fixed 5 second time slices of every RTP stream. The MOS estimates are based on the E-Model defined in ITU-T G.107 using a number of input parameters, such as

- The codec
- Information about burst loss, i.e. consecutive packet loss
- Information about the critical loss density, i.e. the number of packets received in between loss events
- Information about individual packet interarrival times

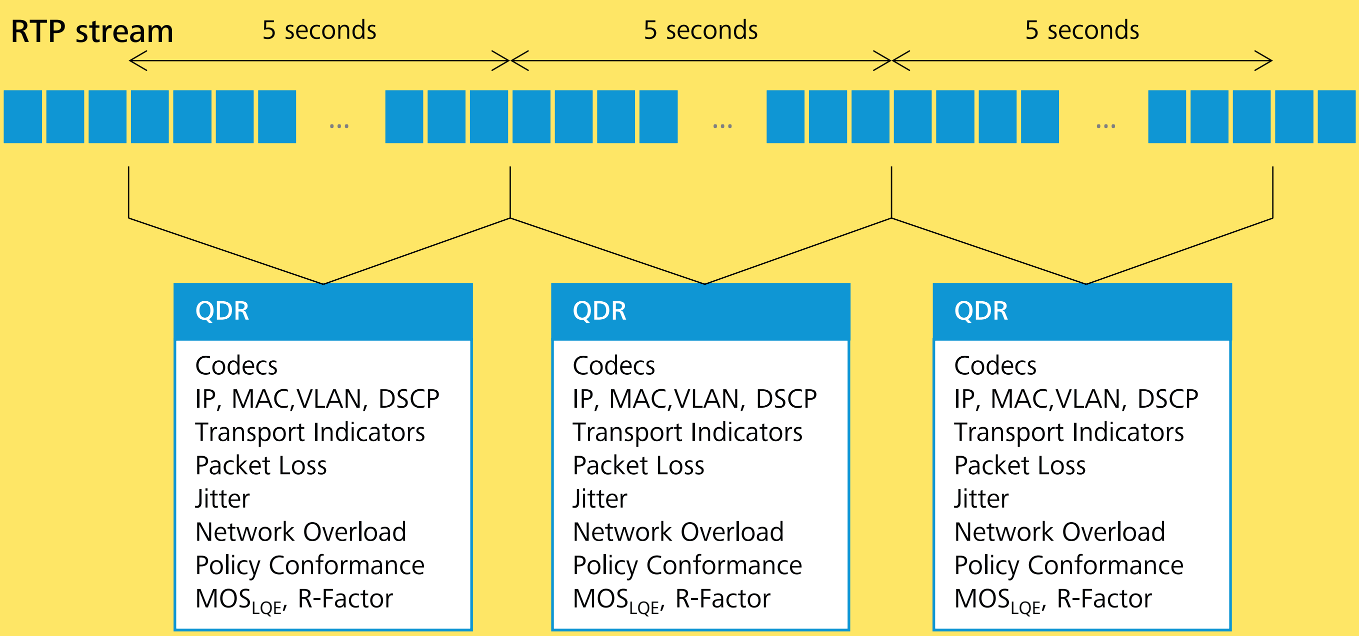
All quality data of a time slice is summarized in a quality data record (QDR). On top of the MOS each QDRs contains information about the transport quality, policy conformance and standard conformance as well as automatic root cause indicators.

This makes QDRs atomic units of quality summarizing all relevant characteristics of each RTP stream segment.

The fixed time slicing approach has been recommended by the TM Forum to account for the varying nature of voice quality in IP networks. Simple averaging per call leads to a significant loss of information.

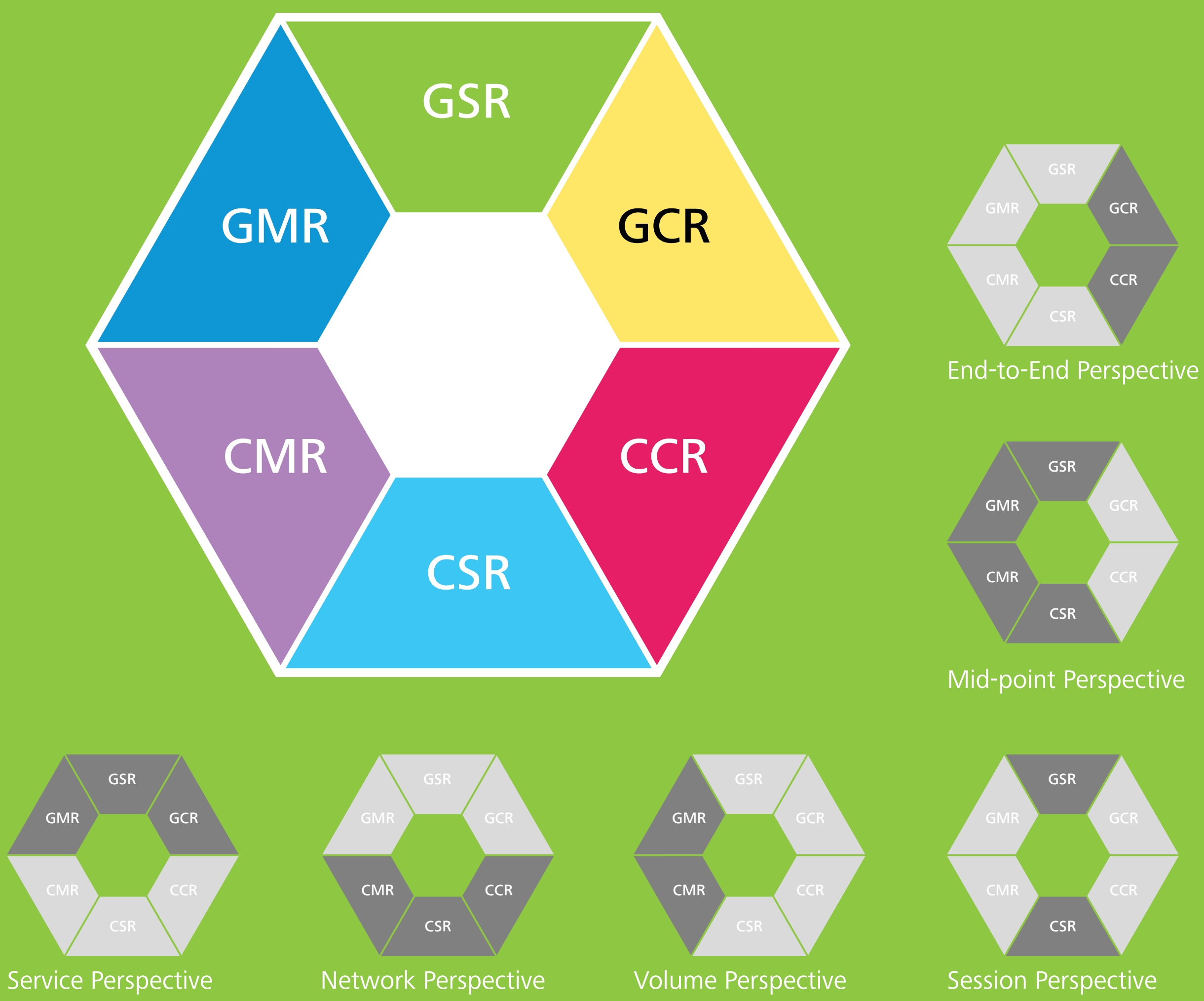
For example, the average quality of a perfect 5-minute call that suffers from severe impairments in the last 10 seconds will be close to perfect.

Nevertheless, the parties will hang up and complain about bad user experience. Averaged metric data does not help to troubleshoot or even just confirm any problems. Obviously, statistics based on such inaccurate data are hardly useful.



Atomic units of RTP quality offer different perspectives on the media plane:

- End-to-end vs. mid-point
- Service vs. network
- Volume vs. session



The KPI system offers different perspectives on the media quality of a VoIP service. For example, Qrystal offers two types of media plane KPIs defined by their scope: mid-point KPIs and end-to-end KPIs. The former provide a view on the quality as measured at a specific point in the network, e.g. at an interconnection. The latter describe the quality of entire calls using the measurements closest to the calling parties, thereby giving insights into the actual user experience.

Another perspective is given by the definition of 'media quality'. Since each time slice contains information about the MOS as well as technical parameters, such as packet loss and jitter, one can independently assess the estimated user experience and the network's transport quality. This is useful as technical issues may not always have an impact on the user experience, but frequently serve as an early warning of upcoming issues.

Finally, time slices can be aggregated to different levels, e.g. to fixed time units such as minutes, to the level of an RTP stream or to the level of a call. This dimension provides insights into the distribution of quality across a group of streams and allows to compare the overall traffic quality versus degradation of individual calls.

Together the dimensions of scope, quality definition and aggregation level offer different views on the service quality:

- End-to-end Perspective: Focused view on the quality on a call-level, i.e. as perceived by the endpoints of communication

- Mid-point Perspective: Focused view on the quality at specific points in the network, e.g. at network boundaries

- Service Perspective: Focused view on the media quality of the service in terms of the customer experience

- Network Perspective: Focus on the technical quality of service or network transport performance

- Volume Perspective: Focus on quality per minute (relying on time slices) indicating overall quality of the observed traffic

- Session Perspective: Focus on quality for entire streams indicating distribution of quality across call sessions.

The Voipfuture media KPI system provides exactly these views on the media plane thereby complementing standard control plane KPIs.

GMR Good Minute Ratio	CMR Critical Minute Ratio
GSR Good Stream Ratio	CSR Critical Stream Ratio
GCR Good Call Ratio	CCR Critical Call Ratio

Understanding service quality

Insights and use cases	mid-point	end-to-end
control plane	Network-centric view on service availability & SIP signaling performance Monitoring of availability & signaling performance of interconnection partners	User-centric view on service availability & SIP signaling performance Monitoring service availability and signaling performance from a customer perspective
media plane	Network-centric view on voice quality & network performance Monitoring voice quality at interconnections & quality of network segments	User-centric view on voice quality & network performance Monitoring of in-call user experience & overall network performance

Mid-point KPIs

Mid-point metrics are used to measure the quality directly at network boundaries, i.e. typically close to Session Border Controllers. This is done to monitor the traffic sent by an interconnection partner or when reporting on the quality of one's own traffic for an SLA. The following defines the four mid-point KPIs of the media KPI system, which can be calculated for any group of RTP streams as defined by time, probing point, origination and destination, direction, trunk association etc.

Good Minute Ratio (GMR)

This KPI is used to measure the overall quality of the measured voice traffic as perceived by the users. The GMR (Good Minute Ratio) reflects the proportion of 'good' RTP stream minutes over all minutes. A 'good minute' is a set of time slices representing one minute, where all time slices have a MOS above 4.0. The good minute ratio is calculated as follows:

GMR = # good minutes / total number of minutes

The GMR is calculated for narrowband and wideband MOS scales.

Good Stream Ratio (GSR)

This metric is used to determine the fraction of RTP streams that are of good quality as perceived by the user. A 'good stream' is a RTP stream, where all measured time slices have a MOS above 4.0. The good stream ratio is calculated as follows:

GSR = # good streams / total number of streams

The GSR is calculated for narrowband and wideband MOS scales.

The GMR and GSR are utilized to measure the quality at specific points in the network.

Both KPIs are based on the MOS and thus make a statement about the estimated user experience. Qrystal offers similar KPIs to assess the network performance based on RTP stream measurements.

Critical Minute Ratio (CMR)

This metric is utilized to measure the overall technical quality of the measured RTP traffic. A 'critical minute' is a set of time slices representing one minute, where all time slices have critical impairments. The following issues are regarded 'critical':

- Packet burst losses and
- Severe jitter that effectively leads to gaps in the audio playout.

The critical minute ratio is calculated as follows:

CMR = # critical minutes / total number of minutes

The best value for this KPI is 0%, meaning that little to no minutes with critical transport impairments have been monitored.

Critical Stream Ratio (CSR)

This metric is used to determine the fraction of RTP streams that are impacted by critical impairments. A ‚critical stream' is an RTP stream, where at least one measured time slice has a critical impairment. The critical stream ratio is calculated as follows:

CSR = # critical streams / total number of streams

The best value for this KPI is 0%, which means that little or no streams with critical transport impairments have been monitored.

End-to-end KPIs

As stated before GMR, GSR, CMR and CSR have their use cases in assessing the quality at network and domain boundaries. Other use cases require a call-based, i.e. end-to-end, view, which better reflects the actual user experience. End-to-end KPIs for the media plane are based on quality-enriched call detail records (qCDR), which contain call-related data on the signaling as well as on the media quality.

Qrystal Connect creates such quality-enriched call detail records summarizing the time slice data, e.g. by storing the minimum, average and maximum MOS value as well as transport quality metrics for each call direction. In total each qCDR has more than 200 fields.

A single RTP stream going from one call party to another may be measured a number of times at different monitoring points. Which measurements should be included in the qCDR, i.e. which is the most relevant in terms of the user experience?

Since the quality of a media stream can only degrade as it flows through the network, the measurement closest to the media's playout point is the most relevant. For the media direction from A-party to B-party, the measurement closest to the B-party best reflects the user experience and vice versa.

Selecting the most relevant measurements is crucial, but does not yet define the quality of call. Listening quality MOS according to ITU-T Recommendation P.800 is only defined for audio sequences with a duration between 4 and 15 seconds. Typical phone calls are much longer and time slicing will generate a lot of quality data that needs to be aggregated per call direction. This can be done in a number of ways, but we propose to use the GMR and CMR as described above.

Further aggregation of quality to the level of a complete call needs to consider the quality for both media directions. Possible metrics include the minimum quality or the average quality of the media directions. However, looking at groups of calls it becomes obvious that the conventional approach of repeatedly averaging MOS (or any other quality metric for that matter) is very problematic. Taken to the extreme one would average MOS of all time slices for each direction, calculate the average MOS of the two directions and then average over all MOS values per call. This approach will conceal any issues revealed by the measurements.

We therefore extend the concepts introduced for the mid-point KPIs to the call-level and define two KPIs based on the definition of 'good' and 'critical' calls. Note that these KPIs are supported from Qrystal 5.3 onwards.

Good Call Ratio (GCR)

This metric is used to determine the fraction of calls that are of good quality as perceived by the users. A ‚good call' is a call where the stream received by the A-party and the stream received by the B-party are good streams as defined above. (Both streams as measured closest to the respective listening parties.) The good call ratio is calculated as follows:

GCR = # good calls / total number of calls

The GCR is calculated for narrowband and wideband MOS scales.

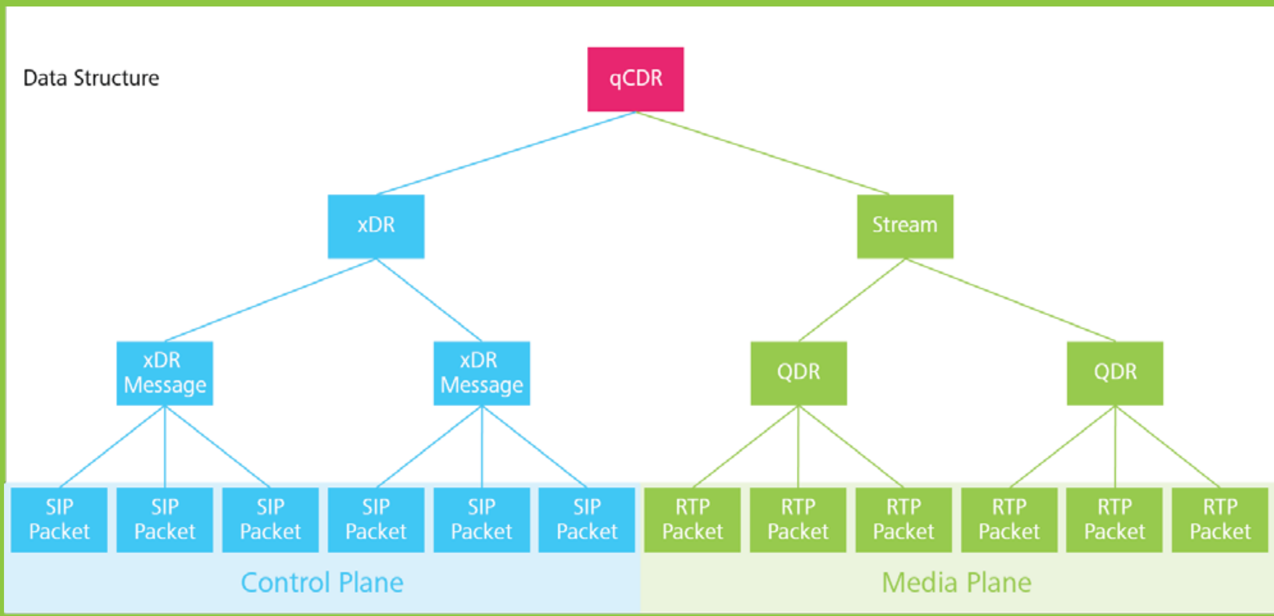
Critical Call Ratio (CCR)

This metric is used to determine the fraction of calls that are impacted by critical impairments of the media streams. A ‚critical call' is a call, where at least one call direction has a critical impairment. The critical call ratio is calculated as follows:

CCR = # critical call / total number of calls

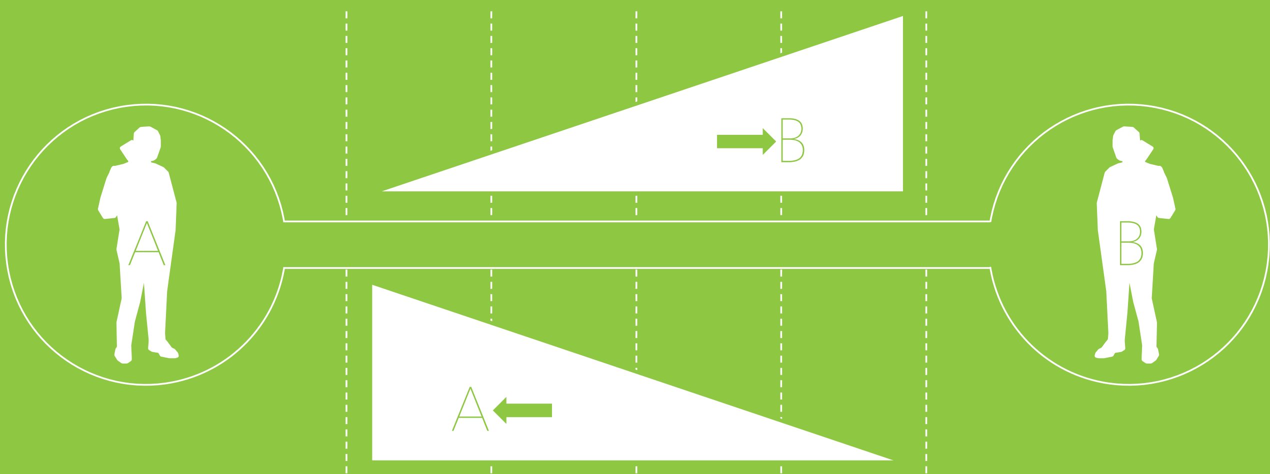
The best value for this KPI is 0%, which means that no calls with critical transport impairments have been monitored.

Call analysis: The hierarchy of collected data



Quality of Experience & the position of measuring points

The nearer to the listening party, the higher the relevance



Conclusion

This whitepaper has discussed the problem of effectively measuring the quality of VoIP services. The IETF RFC 6076 is a widely accepted standard for monitoring SIP signaling, i.e. the control plane. A similar standard for monitoring the media plane is currently not available. Most VoIP service providers therefore use some average MOS value per call, simple transport level metrics and maybe the average call duration to measure the user experience. Building statistics from these simple metrics leads to a significant loss of information and only allows to detect catastrophic events if any.

Voipfuture Qrystal offers a fixed time slicing technology, which provides MOS and other metrics for every five second slice. These atomic units of quality can be conveniently aggregated to create statistics for groups of calls and RTP streams without any loss of detail. Based on the time slice information we introduced six KPIs for monitoring voice quality, which provide different perspectives on the user experience and network performance. Four of these KPIs can be used to measure the media quality at a certain point in the network, e.g. at an interconnection. Two KPIs can be used to measure the end-to-end user experience for groups of calls. All KPIs can be calculated for different grouping criteria, in particular for trunks and routes.

About

Voipfuture is a premium voice quality monitoring vendor developing unique technology for assessing, aggregating, analyzing, and visualizing voice quality information. Voipfuture products offer a precise view on media and control plane to communication service providers, wholesalers and enterprises.

Since its launch, Voipfuture has been at the forefront of voice quality monitoring and continues to redefine Voice over IP by connecting their customers' view on service quality with high resolution user experience.