END-TO-END ASSESSMENT OF MOBILE VIDEO SERVICES

ROHDE&SCHWARZ

Make ideas real

END-TO-END ASSESSMENT OF MOBILE VIDEO SERVICES

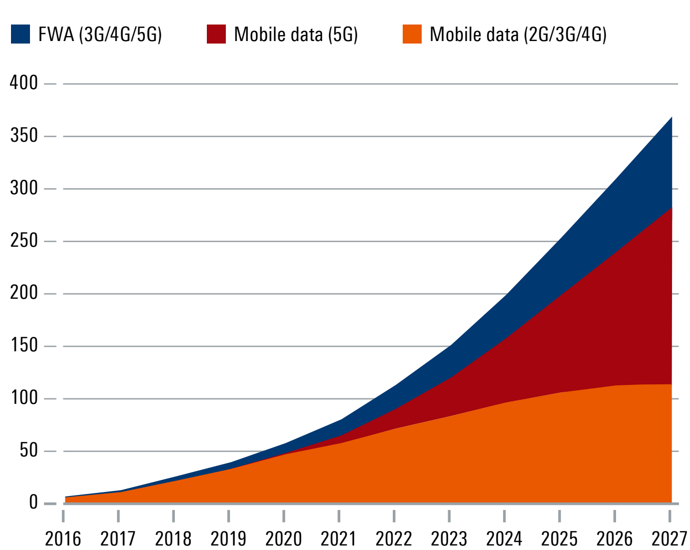
Videos are the most popular web content and have long formed the bulk of the data volume in mobile networks. Which is why not only video consumers but also video service providers and network operators have a vested interest in a high quality of experience. Mobile measuring systems assess this automatically and just as reliably as human viewers.

Nowadays everyone is talking about mobile video but it is not new. At the turn of the millennium and long before the first smartphone, videos could be accessed and streamed from media servers with QuickTime® or RealPlayer®. The limited network capacity meant that these UDP/RTP based video transmissions were highly compressed and successive in real time. Neither buffering under increased transmission speeds nor enhanced error correction were possible. These limitations applied despite modest image sizes of typically QCIF (144 × 176 pixel) or QVGA (240 × 320 pixel). Other mobile video technologies such as DVB-H were just as poor as these early streaming services.

There was no breakthrough in mobile video until HSPA transmission technology and VGA display sizes of 480 × 640 pixel and higher came along. This happened less than ten years ago. Since then, video use has grown exponentially and is now the dominant data type in networks (Fig. 1). One obvious reason is the increased transport capacity of mobile networks and less expensive data plans, but the widespread use of high-resolution, large-format smartphones to access practically all media plays an even larger role. As a result, video is increasingly the primary source of information for many people. YouTube is now the second most popular search engine, right behind Google. Today, resolutions over mobile networks to smartphones can exceed 1440 lines or even be UHD (4K video).

Fig. 1: Global mobile network data traffic (EB per month)

Video traffic is estimated to account for 69 % of all mobile data traffic, a share that is forecast to increase to 79 % in 2027 (Ericsson Mobility Report 2022).



Generally, mobile video services are not the primary services offered by network operators. They largely function independently of telecommunications norms and standards. Content, servers and applications are made available by independent service providers who just use mobile networks to transport data (OTT services). The data is normally encrypted and transported with proprietary protocols on the application layer. Video compression techniques are also service-specific. All information exchanged between an app and service is under the direction of the video service and subject to continuous optimization and adaptation. Providing accurate and detailed information about the many video services available on the market is nearly impossible. Instead, we will give a brief presentation of the principal techniques and explain the need for an insightful assessment of service quality.

Perceived video service quality can be roughly determined based on the following:

- ► Service availability
- ► The delay between the request and start of the video (time to first picture)
- ► The length of unwanted interruptions that occur (stalling)
- ► Image resolution and quality; how much is image quality affected by
  - – Compression loss (blurring from compression and/or reduced resolution, reduced frame rate), blocking artifacts
  - – Transmission errors (artifacts, corrupt images, brief stalling)

Desynchronization between audio and video is also possible.

Technical background for video transmission on mobile devices

Most requested videos are coder-compressed video files that are stored on a server waiting to be called up (video on demand as opposed to live video). Streaming is often used to describe transmission to a consumer device but streaming actually only means continuous transmission and real-time processing on a consumer device. Unlike the early days of mobile video mentioned above, data today is actually transmitted in larger sections and buffered.

An entire video can be downloaded as a file and viewed after being fully received. However, users do not want to wait. Progressive downloading helps solve this problem. The video starts as soon as the first section is available on the smartphone, while the rest of the video is downloaded from the server in the background. This strategy (with sufficient channel capacity) quickly provides the complete video on the device. The advantage is clear: once loaded, connection quality no longer matters and the video can play without interruption.

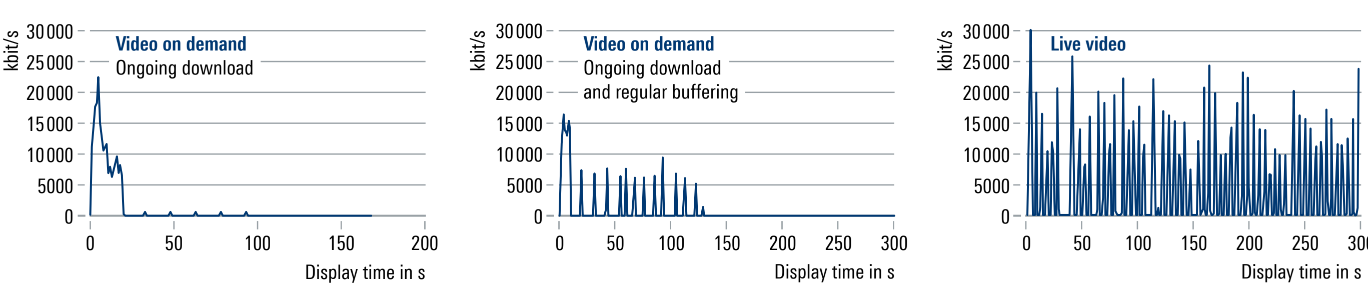
However, it is often the case that users do not watch videos to the end so a complete download would be a waste of transmission capacity. The solution is a compromise between the need to buffer video sections to ensure interruption-free playback and the need to be economical with the available transmission capacity.

First of all, a large initial section of the video file is saved. If it is apparent that the viewer wants to continue to watch the video, the next section is downloaded when a certain playback point is reached. The length of each loaded section ranges from a few seconds to minutes, depending on the philosophy of the video service. The trend is moving toward shorter sections and is therefore again approaching the streaming ideal. However, unlike real-time streaming, a large section of the video remains in the buffer so that long gaps in the connection can be bridged (Fig. 2).

From a technical viewpoint, video on demand is still a file download that does not require real-time transmission. Data transfer is mostly based on reliable TCP/HTTP protocols, which prevent the loss of data and are supported by all operating systems. This can be the classical TCP/HTTP transport or – increasingly – QUIC. Originally introduced by Google, it has been adopted by IETF and is now even part of HTTP/3. For performance reasons, QUIC is based on UDP instead of TCP and used on the transport layer, which can potentially lead to data losses. On the application layer, however, QUIC has implemented mechanisms that prevent losses.

Compared to video on demand, live video still plays a minor role in the network as far as volume is concerned, but places greater real-time demands on the transmission path. Typical applications include video telephony, images from surveillance cameras and video-assisted remote control systems. How narrowly the term real-time is to be interpreted in each case depends on the application. In the private domain, TV and live video are of primary importance in social media. In both cases, the real-time requirements are less strict and a time delay of a few seconds

Fig. 2: Three examples of data transmission measurements for video services (download strategies)



is accepted. For this reason, transmission can take place on the same technical basis as for video on demand. The only difference is that the storage and reload intervals are reduced to just a few seconds.

Using staggered, section-by-section transmission also makes it easy to adapt the bit rate to the transmission channel. Each video section can be delivered with the appropriate compression (e.g. in line with the DASH method) based on the current channel capacity. If things get tight on the transmission path, the video section is delivered with lower resolution or higher compression, which reduces the data rate. This does affect the image quality, but pauses caused by emptying the buffer memory would be even more annoying.

The video provider is responsible for defining whether the client on the smartphone or the server decides what channel information to use as the basis for selecting the appropriate compression level, defining the time constants that regulate this behavior, and all other details. The mobile network only provides the means of transmission; the video service reacts to the given situation, with the primary objective of avoiding image freezing while maintaining a high image quality (depending on channel capacity). The compression methods used are not lossless. A varying amount of detail will be lost depending on the coding scheme and compression level. In the best case, the effects remain below the perception threshold. If greater compression is necessary, blurring occurs, which becomes more evident in moving scenes. Even greater compression causes annoying artifacts such as pixelation blocks and absent color shading.

The data stream is only of limited help when assessing quality

The technical makeup of a video service is irrelevant – what counts is what the viewer sees, i.e. the quality of experience. The question is how to assess this using technical methods.

The size of a video file in relation to the playback time and the associated bit rate provide only limited information because the individual codecs function at different efficiency levels, i.e. transmit different image qualities at the same bit rate. Most codecs have multiple quality levels known as profiles. Profiles define the calculation effort that goes into compression. More complex compression results in a greater level of detail for the same data volume. Finally, the image content also affects data volume. Large-format images in a stationary scene can be encoded more effectively than small-format images with high motion, brightness and color dynamics.

The server and app react to changes in the network and image material by adjusting their settings in a feedback loop. An assessment tool simply based on data flow analysis and no knowledge of image and application metadata would fail to provide reliable quality information from the end customer perspective. And even if service-specific meta information were accessible, the change dynamics in this industry are so great that analysis tool manufacturers would hardly be able to develop their software fast enough to keep up (see page 7). Plus, the majority of today's video services already use encryption on the transport layer. Mere analysis of the received bitstream only delivers a small amount of the information necessary for quality assessment. As an alternative, the displayed image itself serves as the source for analysis. Everything that happens prior to display, such as compressing, transmitting and decoding the video and preparing it for display, is reflected in the image and included in the analysis. What's important is what the viewer ultimately sees. To analyze the screen content, it is necessary to access the image memory of mobile devices – a difficult but manageable challenge.

Ultimately, what counts is what the viewer sees

As already mentioned, the time from when a video is requested to when playback begins (time to first picture) is an essential parameter when assessing service quality. Due to data buffering, the display does not start when the first data package is received on the IP layer – it starts much later. This time delay can only be measured by looking at the screen or examining the image memory.

It is also not possible to accurately analyze purging of the received data stream from the buffer memory since you do not know how full the memory is or if warping measures are used. You also have to look at the display to see if the video freezes or stalls. The measured display time of each image is used as the basis.

Assessment of the actual image is also a challenge. It requires perceptual objective video quality models that take the peculiarities of human perception into consideration.

Perceptual objective video quality models

Perceptual objective video quality models evaluate frames in line with various criteria and analyze motion patterns over long image sequences, in the same way as a person reacts to static and dynamic aspects. The analysis is complex, but the result is simple. It is summarized as an overall value on a quality scale. The internationally recognized absolute quality scale describes the quality as a value between 1 (bad) and 5 (excellent) (Fig. 3). The mean opinion score (MOS) is the average of many individual assessments.

A simple example of perceptual objective analysis is stalling assessment. The more dynamic a scene is, the more annoying image freezing will be. In a scene with very little movement, stalling will result in the loss of just a small amount of information, and may not be perceived at all in the case of a static subject such as a landscape. With a sports broadcast, on the other hand, even brief interruptions will be perceived as extremely annoying. The perceptual objective measure for the motion aspect is referred to as jerkiness; it weights the display duration of an image with the movement in the video and returns a single value that represents the loss of information and the annoyance of waiting.

The environment in which a disruption occurs is also included in the assessement. Artifacts in the image foreground or in a moving object (attraction areas) result in a much more negative assessment than block formation in an extremely bright or extremely dark image area where such artifacts are less noticeable.

Video codecs also use perceptual objective strategies to optimize compression with video content characteristics, e.g. by encoding certain attractive image areas with a greater level of detail while permitting a greater loss of detail in unattractive areas.

Application fields for standardized video quality models

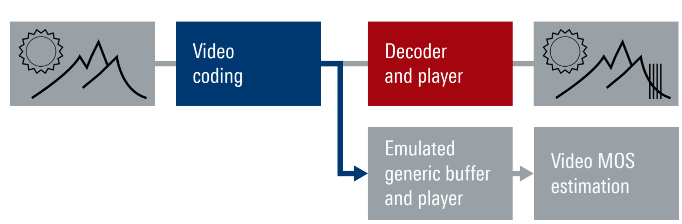
The widespread use of IPTV makes it necessary to measure video quality at various network distribution points. Many video quality estimators have recently been developed for this purpose. Although these estimators only analyze the video bitstream, they provide sufficiently precise results for these applications (Fig. 4). If the bitstream is not encrypted, content information (frame display duration, compression structures) as well as metadata (codec type and profile, packet size) can be used, and it may even be possible to decode the image. For encrypted bitstreams, the amount of information that can be evaluated is restricted. This depends on whether the encryption affects only the actual video data and on which protocol hierarchy level it is applied.

Bitstream based methods are intended for monitoring applications. Here, the video does not have to be known or be available in decoded form. Current methods are described in ITU P.1201, P.1202.1/.2 and P.1203.1-4.

Fig. 3: Commonly used international MOS ratings.

Rating	English	German	French	Spanish
5	excellent	ausgezeichnet	excellente	excelente
4	good	gut	bonne	buena
3	fair	ordentlich	assez bonne	regular
2	poor	dürftig	médiocre	mediocre
1	bad	schlecht	mauvaise	mala

Fig. 4: Bitstream based quality estimators use a small amount of metadata and heuristic methods to derive an MOS.



In with the new – the latest evolution of video standards

MPEG-4 (part 2), H.264 and H.265 are familiar standardized video codecs. For a long time, MPEG-4 (part 2) was the standard of choice for IPTV and DVD-Video. The next development step to H.264 (AVC) made HDTV practicable and is also used for the Blu-ray Disc format. The most recent standard codec is H.265 (HEVC), which is used by standards such as DVB-T2 and will establish itself as the codec for UHD1 transmissions (4K) because it delivers acceptable image quality even with an extremely high degree of compression.

There are also proprietary, mostly open (but not standardized) codecs such as Google VP9. VP9 is somewhere between H.264 and H.265 from a quality viewpoint and currently the only codec used by YouTube. The transition to AV1 (a VP9 based, open source video codec from the Alliance for Open Media) is now underway. AV1 appears as the new legacy codec in internet video streaming. It is open source, very efficient in compression and can be applied natively up to 4K and 8K video content. There is currently a trend among major internet players to move away from classic standardization work in ITU and MPEG. Instead, they are discussing and adopting coding and transmission standards within the framework of mergers and consortia. Since every service maintains its own technical ecosystem and does not need to ensure compatibility with others, codecs (just like communications between server and app) are usually changed without notice or disclosure.

YouTube is a perfect example. Less than five years ago, YouTube transmitted MPEG-4 coded videos in 3GP format via unencrypted TCP connections. Since then, videos have been encrypted, initially using TLS and later Google's own SPDY protocol. Videos were also recoded with H.264. Some time later, there was the transition to MPEG DASH to allow adaptive bit rates. Another step was to again recode the videos, this time with VP9, Google's own video codec, now with AV1. Already at the beginning of 2017, for Android smartphones, YouTube abandoned TCP in favor of UDP and the QUIC application protocol. This list of changes only relates to video transmission measures. With every new app version, YouTube also changes the way in which the buffer memory in the smartphone is managed, i.e. the rules that define how much and when data is buffered as well as the criteria according to which the bit rate is changed.

Other video services make similar adjustments. To compare the quality of different services without being influenced by service evolution, all criteria as well as measurement and assessment methods used have to be measurable for all services over a long period of time and include all components that play a role along the transmission path.

As mentioned above, image based models are advantageous for end-to-end tests, particularly in mobile communications. They are the most accurate in representing the user experience because they are based on human perception and can analyze the image. ITU J.341 and ITU J.343.1-6 provide HD-compatible measurement methods that are applicable to videos transported by unreliable protocols, which means they can rate the effect of erroneous video frames.

P.1204.3-5 is the latest family of standards, covering formats up to UHD-1 (4K video) transmitted using reliable transport protocols. The reference based P.1204.4 method was even approved for evaluating AV1 compressed videos. A basic distinction can be made between reference based and reference-free methods. Reference based methods (picture based, full reference methods, Fig. 5) have access to the source video and can calculate perception-relevant differences to the received video image by image and even pixel by pixel, and combine them to obtain a quality value. Such methods are described in ITU J.341, ITU J.343.5/.6 and most recently in P.1204.4. To use these methods, however, reference videos must be uploaded to the server of the video service to be tested. Quality measurements compare these videos with the same videos stored on the measuring instrument. This method is supported by services that permit private videos to be loaded and streamed (e.g. YouTube), but not usually by professional providers (e.g. Netflix). Reference based methods are also unsuitable for assessing live video because there is no previous playback source.

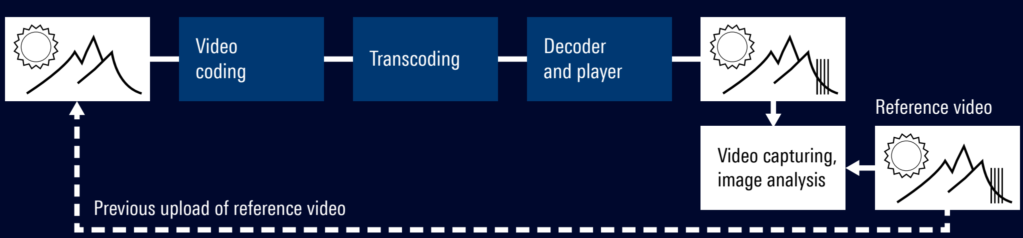
In contrast, reference-free methods (picture based, no reference methods) do not need any a priori knowledge of the source video. The received and decoded video is analyzed for typical disturbances (jerkiness, loss of detail, compression distortion, etc.) and this information is used to calculate the quality value. Standardized methods are described in ITU J.343.1/.2.

The advantage of these methods is their broad range of applications since they function irrespective of the transmission path. This is why ETSI TS 102250-2 recommends the use of J.343.1 for all types of mobile video streaming services.

Secure transmission methods used almost exclusively for mobile video streaming today do prevent bit errors that in the past resulted in severe artifacts and image errors and also reduce compression artifacts (loss of image details, blurred movements) and stalling, i.e. moving image freezing. But with the growing popularity of video telephony with its strict real-time requirements, non-secured (i.e. lossy) transmission is again becoming more prevalent on mobile devices. Many of the current measurement methods are prepared for this.

Fig. 5: Reference based methods compare the streamed video with the original stored on the measuring instrument.

Reference based video quality estimator



ITU J.343.1 structure and application in Rohde & Schwarz products

VMon is an in-house developed quality measurement method in line with ITU J.343.1 from SwissQual/Rohde & Schwarz. It was successfully tested and standardized in 2014 by the ITU and has since been implemented in Rohde & Schwarz mobile network testing Android based test applications. The method also uses meta information from the video stream. A jerkiness value is calculated from the movement and display duration of the individual frames, and a loss of detail is calculated from information indicating the complexity of the images. At the end, the video quality is assessed on an MOS scale from 1 (bad) to 5 (excellent).

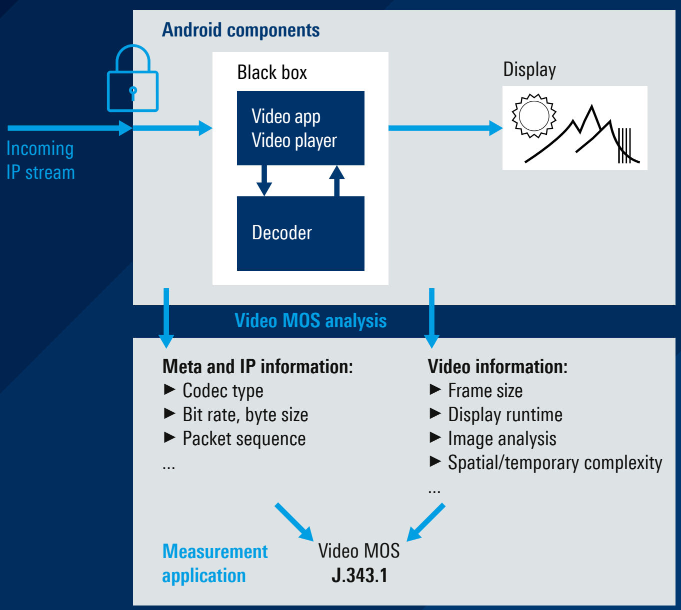
During development, a priority was to ensure that the measurement method could be used in real-time applications. The implementation analyzes only the current video frame in relation to a history comprising just a few images. Despite this constraint, the image assessment must be extremely quick so that it is completed before the next frame: with 25 frames per second, only 40 ms are available to analyze a 3 Mbyte image.

The method also obtains other information from the video signal. Stalling is detected, and the image size and video codec used are recognized (Fig. 6). The data of the deeper protocol layers is also recorded. This results not only in the cumulative quality value, but also yields information that can be used for troubleshooting and transmission path optimization. Video evaluation in line with ITU J.343.1 was adopted to assess 4K video and 60 fps in 2020 and verified for AV1 video codec evaluation in 2022.

Video quality assessment is the main task when determining a video service. The measurement applications support fully automatic control of YouTube, including YouTube live video as the most commonly used video service, as well as Facebook Watch. It is even possible to test almost any other video service in a semiautomatic measurement application. This allows you to quickly respond to new offerings as well as to assess and optimize regional video services.

The video test applications are supported by the QualiPoc Android product family. The family includes R&S®ROMES and QualiPoc Android handheld for network optimization tasks, QualiPoc Android remote control for autonomous network monitoring, and the R&S®FR4 Freerider walk test solution and Benchmarker 3/SmartBenchmarker systems.

Fig. 6: ITU J.343.1 video analysis is based on the images themselves as well as a small amount of metadata.



Quality of experience is more than just image quality

Image quality may be the most important criterion when evaluating a video service, but it is not the only one. Whether and how fast a service can be accessed as well as loading progress information are analyzed. To gain an overall picture, Rohde & Schwarz mobile network testing products include a test sequence that measures the video quality by simulating the actual usage behavior – from starting the video application on the smartphone and requesting a certain video, to analyzing the displayed images (Fig. 7). If waiting times play a role in the real world, the maximum waiting times of a hypothetical average user are used. If these times are exceeded, the test is regarded as failed in cases where the video never became visible, or as dropped if the video froze for a long period of time. Such abort criteria are indispensable for an automated test sequence.

The test sequence can be followed precisely in the test log on the smartphone (Fig. 8, left side). A successful test returns the overall quality (MOS) and other aspects such as jerkiness and freezing (stalling) (Fig. 8, right side). Many other measured values are collected in the background, including the image rate, image resolution, protocols used and the IP and trace log files. This means users have access to the measured video quality values and to all information required for optimizing the transmission path.

Summary

Videos have long accounted for the bulk of data transported in mobile networks, and forecasts predict continued dramatic growth. Network operators and video service providers therefore have a vested interest in keeping video consumers happy by offering high-quality services. Automatic test systems can be used to quickly and reliably determine the quality level. In the mobile sector, reference-free perceptual objective analysis methods are an effective alternative to video quality measurements. These methods deliver meaningful results with a computational effort that can even be achieved using smartphones and are therefore inexpensive and uncomplicated to use.

Although real-time applications such as video telephony do not play a major role at present, this will change in the foreseeable future. The upcoming, virtually latency-free 5G mobile standard will enable and facilitate high-quality applications in real time such as telemedicine video transmissions. Reliable, high path quality is essential. Rohde & Schwarz mobile network testing monitoring products are ready for current and future applications.

Dr. Jens Berger; Dr. Silvio Borer

Fig. 7: Video service measurement from start of the application to establishment of the connection

QualiPoc Android test sequence

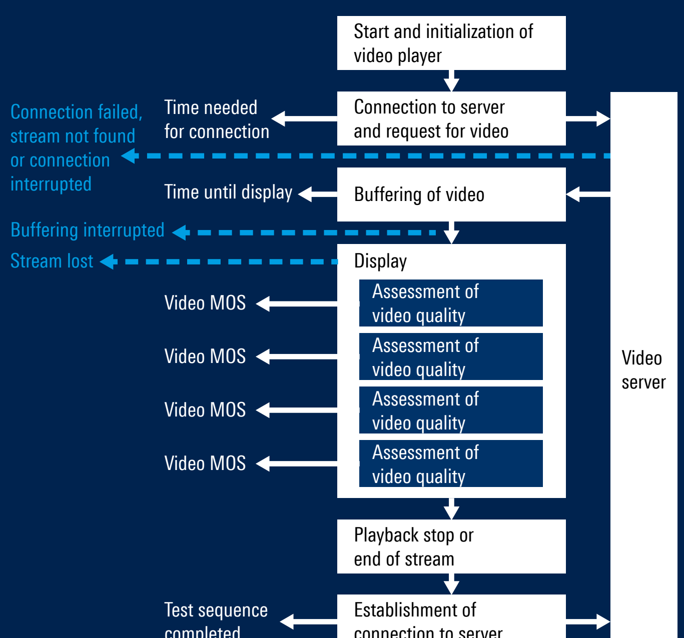


Fig. 8: Real-time analysis of a YouTube video with QualiPoc Android based on the flowchart in Fig. 7

REFERENCES

Quality evaluation of streaming video on mobile networks; Rohde & Schwarz white paper; search for 5215.4369.92 at www.rohde-schwarz.com.

ETSI Technical Report TR 102493 (v.1.3.1): Guidelines for the use of Video Quality Algorithms for Mobile Applications. ETSI, Sophia Antipolis, 2017.

ETSI Technical Specification TS 102250-2 (v.2.5.1): QoS aspects for popular services in mobile networks; Part 2: Definition of Quality of Service parameters and their computation. ETSI, Sophia Antipolis, 2016.

Recommendation ITU-T H.264: Advanced video coding for generic audiovisual services. ITU-T, Geneva, 2013.

Recommendation ITU-T H.265: High efficiency video coding. ITU-T, Geneva, 2016.

Recommendation ITU-T J.341: Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference. ITU-T, Geneva, 2016.

Recommendation ITU-T J.343.1: Hybrid perceptual bitstream models for objective video quality measurements. ITU-T, Geneva, 2014.

Recommendation ITU-T P.1201: Parametric non-intrusive assessment of audiovisual media streaming quality. ITU-T, Geneva, 2014.

Recommendation ITU-T P.1202.1: Parametric non-intrusive bitstream assessment of video media streaming quality. ITU-T, Geneva, 2014.

Recommendation ITU-T P.1203: Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport. ITU-T, Geneva, 2016.

Recommendation ITU-T P.1204: Recommendation ITU-T P.1204: Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport. ITU-T, Geneva, 2020.

Service at Rohde & Schwarz
You're in great hands

- ► Worldwide
- ► Local and personalized
- ► Customized and flexible
- ► Uncompromising quality
- ► Long-term dependability

Rohde & Schwarz

The Rohde & Schwarz technology group is among the trailblazers when it comes to paving the way for a safer and connected world with its leading solutions in test & measurement, technology systems and networks & cybersecurity. Founded more than 85 years ago, the group is a reliable partner for industry and government customers around the globe. The independent company is headquartered in Munich, Germany and has an extensive sales and service network with locations in more than 70 countries.

www.rohde-schwarz.com

Mobile network testing

The company's broad and diverse product portfolio for mobile network testing addresses every test scenario in the network lifecycle – from base station installation to network acceptance and network benchmarking, from optimization and troubleshooting to interference hunting and spectrum analysis, from IP application awareness to QoS and QoE of voice, data, video and app based services.

www.rohde-schwarz.com/mnt

Rohde & Schwarz training

www.training.rohde-schwarz.com

Rohde & Schwarz customer support

www.rohde-schwarz.com/support

PD 5215.5171.32 | Version 02.00 | February 2023 (sk)
End-to-end assessment of mobile video services
Data without tolerance limits is not binding | Subject to change

5215517132
5215.5171.32 02.00 PDP/PDW 1 en