



WHITEPAPER

Quality of Service / Quality of Experience: Overview and Best Practices

Reference to the Media over IP Value Chain



Arbeitsgruppe Media over IP
Version 1.0, Januar 2021

Inhaltsverzeichnis

Einführung	3
IP Streaming	4
Methoden zur Beurteilung der Videoqualität	6
Objektive Qualitätsbewertungsmethoden	6
Messung der Streaming Video Qualität	7
Adaptive Streaming Metriken und Modelle	7
MPEG DASH Annex D	8
SAND	8
CTA-2066 - Streaming Quality of Experience Events, Properties and Metrics	8
CTA-5004 - Common Media Client Data (CMCD)	8
Streaming Video Alliance - Key Network Delivery Metrics	8
ITU-T Rec. P.1203, P.1204	8
Streaming Metriken (KPIs)	9
Parameter für die Zuverlässigkeit der Stream Wiedergabe	9
Parameter für die Videoqualität (KQIs)	10
Anwendungsszenarien	11
Encoding Analyse und Optimierung	11
Infrastruktur und Netzwerk Analyse	13
Netzwerk Optimierung	12
SAND SRA	12
CMCD	13
KPI / KQI Analyse im Player	13
QoE Messung auf Basis integrierter KQIs unter Betrachtung der auf Player-Seite vorliegenden Streaming-Session	13
Zusammenfassung	14
Abkürzungsverzeichnis	15
Appendix	15
Impressum	18

Einführung

Die Arbeitsgruppe Media over IP beschäftigt sich im Kern mit den Möglichkeiten und Herausforderungen neuer Technologien für das Fernsehen. Eine zentrale Technologie hierbei ist (Unicast) Videostreaming, also die paketweise, individuelle Übertragung von linearen und nichtlinearen Videoinhalten über das Internet. Dienste wie Netflix und YouTube verwenden diese Technologie, um Ihre Nutzer direkt und individuell mit Inhalten zu beliefern. Die Vorteile des Videostreamings werden auch für TV immer relevanter: Inhalte können zu beliebiger Zeit und auf einer Vielzahl internetfähiger Geräte konsumiert werden und es ist keine spezielle TV-Netzwerkinfrastruktur (Satellit, Terrestrik, Kabel) mehr nötig. Um eine breite Nutzerakzeptanz für die neuen Übertragungswege zu erzielen ist es unerlässlich, dass Stabilität und Qualität auf gleichem Niveau sind wie beim traditionellen Rundfunk (Broadcasting).

Um über Stabilität und Qualität wiederum verlässliche Aussagen treffen zu können, beschäftigt sich dieses Whitepaper mit der Qualitätsmessung von Videostreaming:

- Zunächst werden Begrifflichkeiten definiert und anhand der Streaming-Kette schematisch beschrieben, welche Komponenten Einfluss auf die Qualität haben ([Abschnitt 2](#))
- In [Abschnitt 3](#) gibt es eine Einführung in die relevanten Messmethoden
- [Abschnitt 4](#) gibt Einblick in die Anwendung der Messmethoden und definiert anwendbare KPIs und Richtwerte
- In [Abschnitt 5](#) wird anhand von Fallbeispielen die praktische Anwendung der Messmethoden und KPIs erläutert.

Ziel dieses Whitepapers ist es, den Mitgliedern der DTVP einen praxisorientierten Einstieg in die relativ neue, aber hochrelevante Welt der Qualitätsmessung bei Videostreaming zu geben. Insbesondere soll gezeigt werden, dass es am Ende auf die sogenannte Quality of Experience (QoE) ankommt, also die Qualität dessen, was die Nutzer individuell auf ihren Endgeräten erleben. Die ITU-T definiert QoE als “The degree of delight or annoyance of the user of an application or service.” (ITU-T Rec. P.10/G.100). Die sogenannte Quality of Service, also das messbare Zusammenspiel der einzelnen Komponenten, ist eine wichtige, aber nicht die einzige Grundlage der QoE. Die ITU-T definiert QoS entsprechend als “The totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service.” (ITU-T Rec. P.10/G.100). Neben der QoS bzw. dem System beeinflussen auch Nutzer- und kontextspezifische Faktoren die QoE (Le Callet et al., 2012; Möller & Raake, 2014).

Dieses Whitepaper entstand in Zusammenarbeit führender deutscher Forschungseinrichtungen und Videostreaming Plattformen und hat einen klaren Fokus auf dem Anwendungsszenario TV-Streaming.

IP Streaming

IP-basierte Medienauslieferung kann als Multicast (z.B. IPTV) oder als Unicast erfolgen. Wir konzentrieren uns in diesem Dokument auf adaptives Unicast-Streaming basierend auf dem HTTP-Protokoll (HyperText Transfer Protocol), welches auch als ABR (Adaptive Bitrate Streaming) oder HTTP-based Adaptive Streaming (HAS) bezeichnet wird und die Basistechnologie für OTT-Streaming darstellt. Dazu gehören die zurzeit marktrelevanten Streaming-Technologien DASH (Dynamic Adaptive Streaming over HTTP) und Apple's HLS (HTTP Live Streaming) sowie CMAF (Common Media Application Format), welches einen ersten Schritt zur Vereinheitlichung von DASH und HLS spezifiziert hat. Es gibt zwei grundlegende Content-Formate, die über ABR-Streaming übertragen werden und unterschiedliche Anforderungen an die Servicequalität stellen: Lineares TV und Video-on-Demand (VoD). Lineares TV repräsentiert hierbei die höchsten und komplexesten Anforderungen an die End-zu-Ende Servicequalität.

OTT-Streaming wird im Allgemeinen auch als „Best Effort“ Streaming bezeichnet, weil der Serviceanbieter seinen Dienst über das öffentliche Internet anbietet (z.B. YouTube) und somit die Servicequalität beim Kunden nur teilweise beeinflussen kann. Ist ein Netzwerkanbieter selbst auch Anbieter eines Streamingdienstes in seinem Netzwerk spricht man von „Managed OTT“-Streaming, da er die Qualität des Dienstes weitestgehend kontrollieren kann.

Videostreaming ist ein komplexer Prozess mit zahlreichen Einflussfaktoren entlang der Streamingkette. Diese Kette besteht aus den in Abbildung 1 dargestellten Komponenten Transcoding/Kompression, Paketierung, CDN und Endgerät inklusive Player, wobei nur das optimale Zusammenspiel aller Komponenten eine gute Servicequalität gewährleistet.

Während des Transcoding/Kompressionsschritts werden entsprechend einer vorab festgelegten Konfiguration („Bitrate-Ladder“) verschiedenen Versionen der Originalvideos erzeugt und in kurze Segmente unterteilt. Diese Versionen (Repräsentationen) unterscheiden sich entsprechend der Bitrate-Ladder in Auflösung und Komprimierung gegenüber dem Original. Der ABR-Algorithmus (Adaptive Bitrate) im Player wählt während des Streamings dynamisch eine der momentan verfügbaren Netzwerkkapazität angemessene Repräsentation aus, um die bestmögliche Videoqualität bei verfügbarer Bandbreite zu erhalten.

Fehler in den Eingangssignalen (Ingest) werden direkt weitergereicht, weshalb eine Qualitätskontrolle dieser erfolgen sollte. Nur ein gutes nachfolgendes Encoding/Transcoding ermöglicht auch eine gute Servicequalität beim Endkunden. So kann z.B. eine fehlerhafte oder zu starke Kompression des Videosignals zu sichtbaren Artefakten im encodierten Video führen. Standard-Kodiervorfahren im Markt sind H.264 (MPEG-4 AVC), H.265 (MPEG-H Part 2 / HEVC), VP9 und AV1.

In der Paketierung erfolgt das Verpacken der einzelnen Videoelemente in Zielformate wie fMP4 (fragmented MP4) oder MPEG-TS (MPEG Transport Streams) für ein DASH oder HLS-Streaming, wobei fMP4 für DASH und MPEG-TS für HLS verwendet wird. Für CMAF wird fMP4 benutzt. Hier findet auch die DRM-Verschlüsselung (DRM: Digital Rights Management) statt. Eine der größten Herausforderungen bzgl. der QoE ist hier die Ende-zu-Ende

Verzögerungszeit (Latency) für das lineare TV-Format, welche durch die Segmentierung des Videostreams entsteht. In den Standardformaten für HLS und DASH beträgt diese Verzögerung je nach eingestellter Segmentlänge zwischen 15 und 45 Sekunden im Vergleich zum Eingangssignal. Mit dem neueren CMAF-basierten Verfahren kann diese auf 3-6 Sekunden gesenkt und damit vergleichbare Werte wie beim klassischen Rundfunk erzielt werden.

Das CDN (Content Delivery Network) verteilt die angeforderten Videostreams im Fileformat zuverlässig an die Vielzahl der Endgeräte. Es repliziert den Inhalt vom Origin-Server zu einer Vielzahl von im Netzwerk verteilten Cache-Servern. Um eine hohe QoE beim Endkunden zu erreichen, sollten die Cache-Server möglichst „nahe“ beim Kunden lokalisiert sein, d.h. im Edge-Netzwerk. Einen möglichen Vorteil für die QoE liefert ein CDN, welches Cache-Server innerhalb des Netzwerkes des vom Endnutzer verwendeten Zugangsnetz betreibt.

Die Charakteristik des Zugangsnetzwerkes beeinflusst die Servicequalität beim Kunden entscheidend. Hier können Festnetz (Kabel, DSL), WLAN oder Mobilfunk (3G, 4G 5G) unterschieden werden.

Das vom Kunden genutzte Endgerät spielt ebenfalls eine bedeutende Rolle in der Streamingkette für die QoE. Hier ist zu unterscheiden, ob der Kunde ein eigenes Endgerät (z. B. Smartphone) oder ein vom Anbieter zur Verfügung gestelltes Gerät (Streamingbox) nutzt. Bei letzterem hat der Serviceanbieter eine bessere Kontrolle über die Servicequalität. Die Eigenschaften der Endgeräte (Größe des Bildschirms, Auflösung, CPU-Leistung) sowie die Performance des verwendeten Videoplayers sind entscheidend für die Darstellung und Qualität der Streams.

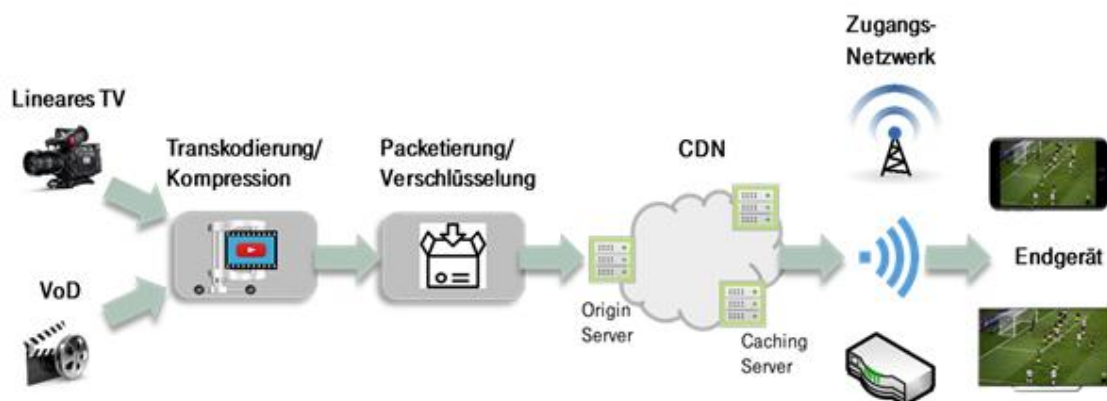


Abbildung 1: Streaming-Kette

Methoden zur Beurteilung der Videoqualität

Für die Bewertung der Qualität von Videostreaming sind zwei Perspektiven zielführend: Die QoS-Perspektive spiegelt das Erfassen von Netz- oder Service-Parametern wider, gemäß der in Abschnitt 1 angegebenen Definition. Diese Messergebnisse sind i.A. mit den technischen Stellschrauben des Systems verbunden, wie ausgewähltes Encoding und Repräsentation, Netzeinstellungen, der Konfiguration von CDNs und den im Player eingesetzten Algorithmen und Einstellungen.

Die QoE-Perspektive adressiert die tatsächlich vom Nutzer wahrgenommene Qualität im Sinne des "Qualitätserlebens". Diese lässt sich entweder im Sinne von konkreten, von den Nutzern vorgenommenen Bewertungen ausdrücken, wie beispielsweise auf der weit verbreiteten 5-Punkte Mean Opinion Score Skala (MOS¹), oder in Messwerten in Bezug auf das Nutzerverhalten, wie die Watch Time (Dauer der Betrachtung eines Videos, oder Abweichung vom Mittelwert der Watch Time unter bestimmten Netzbedingungen) oder Aktionen wie das Stoppen und Neustarten eines Streams.

Zur Erhebung entsprechender Messwerte zur QoE aus Benutzersicht gibt es zwei grundsätzliche Methoden: (1) Verfahren zur Bewertung oder zum Vermessen des Verhaltens von tatsächlichen Nutzern (im Labor, mittels Crowd-Sourcing oder im Feld), oder (2) instrumentelle Verfahren, die darauf abzielen, von realen Nutzern erhobene Messwerte entsprechend (1) automatisiert, "objektiv" vorherzusagen, d.h. auf Basis von Netz- oder Stream-basierten Messungen zu schätzen. Diese "objektiven" Modelle und Verfahren können auf Basis von Daten wie Medien-Bitströmen, Video und Audiosignalen und/oder Metadaten (Netz-, Medieninformationen wie ausgespielte Segmente, Stallings, Nutzerdaten usw.) operieren. Entsprechend dem Fokus des vorliegenden Whitepapers auf technische Messungen sind die folgenden Betrachtungen auf instrumentelle, "objektive" Verfahren (2) beschränkt.

Objektive Qualitätsbewertungsmethoden

Objektive Qualitäts-Bewertungsmethoden stellen einen Zusammenhang zwischen gemessenen QoS Streaming Parametern (KPI's) und der wahrgenommenen QoE der Nutzer durch subjektive Tests her. Sie werden auch als QoE-Vorhersagemodelle bezeichnet. Davon existieren drei grundlegende Kategorien, Bitstream-basierte QoE-Modelle, Pixel-basierte QoE-Modelle und Metadaten-basierte QoE-Modelle.

- Metadaten-basierte QoE-Modelle nutzen Informationen aus dem Metadaten-Layer wie Bildauflösung, Bildwechselfrequenz und Bitrate. Sie können auch als leichtgewichtige Varianten von Bitstrommodellen aufgefasst werden, die nur den Metadatenanteil analysieren. Ein Beispiel dafür ist ITU-T Rec. P.1203.1, "Mode 0".
- Bitstream-basierte QoE-Modelle analysieren den IP-Video Stream ohne Decoding und benötigen keinen Zugriff auf den originalen Bitstream des Quellsignals. Beispiele sind ITU-T Rec. P.1203.1 (Modes 1 und 3) und Rec. P.1204.3.

¹ Bei der sogenannten MOS-Skala handelt es sich eigentlich um eine 5-Punkte Absolute Category Rating Skala gemäß z.B. ITU-T Rec. P.800 oder P.910, mit Darstellung eines Mittelwerts über die Bewertungen ("Mean Opinion Score").

- Pixel-basierte QoE-Modelle (z.B. VMAF, Video Multimethod Assessment Fusion, s. Abschnitt 4) analysieren die dekodierten Frames des Videos. Hier können unterschiedliche Varianten unterschieden werden:
 - Full Reference (FR), bei denen der Originalinhalt für den Vergleich mit dem empfangenen, wahrgenommenen Bild verglichen werden kann, wie z.B. ITU-T Rec. P.1204.4 oder VMAF.
 - Reduced Reference (RR), bei denen eine "reduzierte" Darstellung von Referenz und zu bewertender Sequenz verwendet werden, wie z.B. ITU-T Rec. P.1204.4.
 - No Reference (NR), bei denen die Beurteilung ohne Zugriff auf den Referenzinhalt erfolgt. Derzeit ist kein NR-Modell bekannt, das eine ausreichend gute Vorhersagegenauigkeit im Vergleich zu tatsächlichen Nutzerbewertungen liefert.
- Für Live Streaming sind Pixel-basierte RR- oder FR-Modelle wegen ihrer Komplexität und der Verwendung von Referenzinformationen nicht ohne Weiteres anwendbar. Auch bei Video-on-Demand können solche Modelle im Allgemeinen wegen ihrer Komplexität nur unter Einsatz einer Vorausberechnung der Qualität auf Server-Seite verwendet werden, ggf. mit entsprechender Übertragung dieser Informationen als Metainformation.
- Hybride Modelle basieren auf einer Auswertung von Pixel-Informationen und zusätzlich Bitstrom- oder Metadaten-Informationen, wie z.B. bei ITU-T Rec. P.1204.5.

Messung der Streaming Video Qualität

Adaptive Streaming Metriken und Modelle

Moderne OTT-Videoplayer empfangen die wiederzugebenden Videosegmente per HTTP. Eine Manifest-Datei beschreibt die auf dem Streamingserver verfügbaren Repräsentationen. Der ABR-Algorithmus (Adaptive Bitrate) des Videoplayers misst permanent verschiedene Metriken und passt daraufhin sein weiteres Vorgehen an. Kann zum Beispiel bei schlechtem Datendurchsatz der Videobuffer nicht schnell genug gefüllt werden, so dass es zum "Stillstand" oder "Ruckeln" des Videos kommen würde (Stalling), ruft der Player eine Repräsentation mit niedriger Bitrate auf. Die Metriken, die dem Player zur Abspielzeit zur Verfügung stehen, können auch an einen externen Dienst gesendet werden ("Reporting"), um retrospektiv Analysen durchzuführen und das Gesamtsystem zu optimieren. Zusätzlich können dem Player auch aggregierte Information (z.B. Ende-zu-Ende Latenz), die nicht direkt im Endgerät zur Verfügung stehen, übermittelt werden, um die Ende-zu-Ende Qualität in Echtzeit zu optimieren.

Streamingmetriken können prinzipiell in allen marktüblichen Streamingumgebungen und Playern erhoben werden. Die grundlegenden HTTP-basierten ABR-Übertragungsmechanismen sind überall gleich, egal ob es sich um Live-TV oder VoD, oder die Wiedergabe in einem nativen Player oder in einem Webbrowser handelt. Die folgenden Abschnitte geben einen Überblick über verschiedene Industriestandards mit dem Ziel, die darin definierten Streaming Metriken zu harmonisieren und ihren Austausch zu vereinfachen. Dabei werden zunächst unterschiedliche Standards und Messprotokolle aufgelistet, die eine Erfassung oder Kommunikation von QoS- und QoE-spezifischer

Information adressieren. Im weiteren Verlauf dieses Abschnitts werden spezifische Key Performance Indicators (KPIs) als QoS-bezogene Messindikatoren tabellarisch aufgeführt und kurz erläutert. Im letzten Teil dieses Abschnitts werden sogenannte Key Quality Indicators (KQIs) gelistet und erläutert. Sie ermöglichen es, u.a. auf Basis zugrunde liegender KPIs, Aussagen zur vom Kunden wahrgenommenen QoE zu liefern.

MPEG DASH Annex D

Im Annex D des MPEG-DASH ([ISO/IEC 23009-1](#)) Standards werden Metriken für das adaptive Streaming-Format MPEG-DASH definiert. Dazu gehören folgende Roh-Metriken: TcpList (Metriken zu Transaktionen auf TCP-Ebene), HttpList (Metriken zu Transaktionen auf HTTP-Ebene), RepSwitchList (Metriken zu Qualitätswechsel), BufferLevel (die Größe des Buffers zu einem bestimmten Zeitpunkt) und PlayList (z.B. Play/Pause-Events). Die Metriken sind auch auf andere Streaming-Formate (z.B. HLS) anwendbar.

SAND

SAND (Server and Network-assisted DASH), in [ISO/IEC 23009-5](#) spezifiziert, bietet ein standardisiertes Nachrichten-Format und -protokoll für die Übertragung von Streaming-Metriken an. SAND-Nachrichten werden als XML-Dokumente verpackt und per HTTP oder WebSockets verschickt. SAND referenziert MPEG DASH Annex D für den Inhalt der SAND-Nachrichten.

CTA-2066 - Streaming Quality of Experience Events, Properties and Metrics

[CTA-2066](#) standardisiert Player-Events und -Eigenschaften, QoE-Metriken und die dazugehörige Terminologie. Die aggregierten Metriken werden in vier Kategorien unterteilt: Availability, Startup Time, Continuity, Video & Audio Quality.

CTA-5004 - Common Media Client Data (CMCD)

[CMCD](#) (noch nicht veröffentlicht, [Demo-Link](#)) standardisiert den Informations-Austausch zwischen Player und CDN. Die Informationen dienen CDN-seitig der Log-Analyse, QoS-Monitoring und Optimierung des Datentransfers.

Streaming Video Alliance - Key Network Delivery Metrics

Die Streaming Video Alliance definiert in diesem [Dokument](#) Metriken, die die Effektivität der Netzwerk-Übertragung (Streaming) beschreiben. Dafür wurden nur die gängigsten und universell messbaren Metriken im Standard aufgenommen.

ITU-T Rec. P.1203, P.1204

[ITU-T Rec. P.1203](#) besteht aus mehreren Teilkomponenten zum Monitoring der QoE von Adaptive Streaming Sessions von 1 bis 5 min Dauer. Der Teilstandard P.1203.1 spezifiziert dabei vier unterschiedlich komplexe Qualitätsmodelle, die jeweils auf Basis von Bitstrominformationen eine Pro-Sekunde Schätzung der Videoqualität auf einer 5-Punkte "MOS-Skala" liefern: Das einfachste sogenannte Mode 0 Modell verwendet dabei

Eingangsinformationen ähnlich zu CTA-2066 - wie etwa den verwendeten Codec - sowie die für die ausgespielten Segmente gemessene Auflösung, Bit- und Frameraten. "Mode 1" liefert genauere Schätzungen, unter Verwendung von Frame-Typen und Frame-Größen. "Mode 2" und "Mode 3" basieren auf einem vollständigen Zugriff auf den encodierten Videobitsstrom, mit unterschiedlicher Menge an ausgewerteter Datenmenge.

P.1203.2 beschreibt eine entsprechende Audioqualitätskomponente, die ebenfalls Pro-Sekunde MOS-Schätzungen liefert, anwendbar für unterschiedliche Audiocodexs.

Zusammen mit Informationen zur initialen Ladeverzögerung und den Zeiten von Stalling bzw. Rebuffering (vergl. CTA-2066 KPIs) werden die Audio- und Videoqualitätswerte im P.1203.3 Modul zu einem Session-QoE-Wert integriert. Die "Ground-Truth" zur subjektiven Gesamt-QoE ist in umfassenden Laborstudien mit Probanden erhoben worden, um die Modelle zu trainieren und validieren, s. Raake et al. (2017), Robitza et al. (2018). Für exemplarische Anwendungen von P.1203 s. Robitza et al. (2018), Robitza et al. (2020) und Schwarzmann et al. (2020).

Für die Vorhersage der Videoqualität mit Auflösungen bis zu 4K / UHD wurde Ende 2019 die neue Standard-Serie ITU-T Rec. P.1204 verabschiedet. Diese umfasst bisher drei Arten von Videoqualitätsmodellen auf Basis unterschiedlicher Eingangsinformationen. Sie liefern einen geschätzten Videoqualitätswert für Abschnitte bis zu 10 Sekunden Dauer sowie pro-Sekunde-Schätzungen, für die Videocodexs H.264, HEVC/H.265 und VP9 mit unterschiedlichen Presets: (i) Bitstrom-basiertes Modell P.1204.3 (s. Ramachandra Rao, 2020), (ii) Pixel- und Referenzbasiertes (RR/FR) Modell P.1204.4, (iii) hybrides, Metadaten- und Pixel-basiertes "No-Reference" Modell P.1204.5.

Streaming Metriken (KPIs)

Die folgenden KPIs sind eine Auswahl von Metriken aus den o.g. Standards. Sie können in Parameter für die Zuverlässigkeit der Stream Wiedergabe und in Parameter für die Videoqualität (KQIs) unterteilt werden. Im Folgenden werden die wichtigsten aufgeführt.

Parameter für die Zuverlässigkeit der Stream Wiedergabe

Video Start-up Time (sec) *	Zeit zwischen dem Auslösen einer Videowiedergabe und dem ersten Abspielen auf dem Bildschirm. Alternative Begriffe: Initial Loading Time, Initial Loading Delay.
Playback Stalls *	Zeitpunkt und Zeitraum des Buffer-bedingten Stoppens beim Abspielen eines Videostreams. Alternativer Begriff: Stalling.
Video Frame Drops	Anzahl der vom Player weggelassenen Video Frames aufgrund von Performance-Problemen
Video Start Failure (yes/no) *	Das erste Video-Segment wird nicht innerhalb von 10 sek nach Initialisierung ausgeliefert
Video Bitrate (bits/sek) *	Zeitpunkt, Zeitraum und Größe der abgespielten Bitrate (Bitratenverlauf über der Zeit)
Video Resolution	Verlauf der Auflösung des Videostreams (z.B. 4096x2160, 1920x1080)

Video Frame Rate	Bildwechselfrequenz (z.B. 25 fps)
Quality Switches *	Zeitlicher Verlauf der jeweiligen Darstellung (Codec, Bitrate, Auflösung, Framerate) entsprechend der verwendeten Streaming Bitrate Ladder.
Re-Buffer Ratio (%)	Verhältnis zwischen der gesamten Wiedergabezeit des Videostreams und der gesamten Wiedergabezeit plus der Zeit für das Rebuffering (Stalls, Stalling).
Average Bitrate (bits/sek)	Mittlere Bitrate des Players für die Videowiedergabe

* in [SVA](#) definiert

Parameter für die Videoqualität (KQIs)

PSNR (Peak Signal-to-Noise-Ratio)	PSNR beschreibt beschreibt den sogenannten Mean Squared Error (MSE), also die mittlere quadratische Abweichung zwischen zwei Frames. Der PSNR Wert wird in Dezibel angegeben, dabei gilt: je höher der Dezibel Wert desto kleiner ist die Abweichung zwischen dem komprimierten Bild und dem Original.
VMAF (Video Multimethod Assessment Fusion)	VMAF ist eine von Netflix entwickelte Videoqualitätsmetrik zum Erfassen von subjektiv wahrgenommener Videoqualität. VMAF basiert auf Verfahren des maschinellen Lernens. Der wesentliche Unterschied zwischen VMAF und anderen Qualitätsmetriken wie PSNR besteht darin, dass das zugrundeliegende Modell auf Basis echter Zuschauer-Bewertungen trainiert wurde. Hierfür wurden in Labortests Werte für verschiedene Kategorien von Filmen aus dem Netflix-Katalog erhoben.
SSIM (Structural Similarity Index)	SSIM ist ein objektives Vorhersagemodell für die wahrgenommene Videoqualität. SSIM mißt die Ähnlichkeit zwischen zwei Bildern. Als Referenz dient ein initiales unkomprimiertes Bild (Video).
Audio- und Videoqualität nach ITU-T Rec. P.1203 und P.1204	Die Standards ITU-T Rec. P.1203 und P.1204 beinhalten Modellkomponenten zur Audio- und Videoqualitätsschätzung mit Ausgabe auf einer 5-Punkte Skala [1,5]. Rec. P.1203.1 beschreibt vier unterschiedliche "No Reference" Algorithmen zur Metadaten- bzw. Bitstrom-basierten Schätzung der Videoqualität mit einer Pro-Sekunde-Auflösung, die dann für längere Zeiträume integriert werden können. Die Modes 0 bis 3 unterscheiden sich in der Komplexität der betrachteten

	<p>Eingangsdaten, siehe Abschnitt 4, "ITU-T Rec. P.1203, P.1204". Rec. P.1203.2 spezifiziert ein Metadaten-basiertes Audioqualitätsmodell, das Pro-Sekunde-Schätzungen für unterschiedliche Audio-Codecs liefert.</p> <p>Die Standardfamilie Rec. P.1204 spezifiziert drei unterschiedliche, sehr genaue Videoqualitätsmodelle für eine Pro-Sekunde-Vorhersage für Auflösungen bis 4K/UHD, Frameraten bis 60 fps und die Videocodecs H.264, HEVC/H.265 und VP9. Mehr Details s. Abschnitt 4, "ITU-T Rec. P.1203, P.1204".</p>
--	---

Anwendungsszenarien

Encoding Analyse und Optimierung

Modelle wie VMAF oder die P.1204-Serie (s. Abschnitt 4) ermöglichen prinzipiell die Qualitätsanalyse von bereits encodiertem Videomaterial. Sie können somit für die Berechnung einer sogenannten "Bitrate Ladder" verwendet werden, im Sinne der für eine Sequenz oder einzelne Szene besten Repräsentationen in Bezug auf Auflösung, Frame- und Bitrate. Entsprechend werden derartige Maße beispielsweise von Netflix für ihren "Dynamic Optimizer" als szenenspezifisches "Per-Scene" Encoding verwendet, d.h. unter Berücksichtigung von Szenen-Wechseln (Katsavounidis et al., 2018). Auch Artefaktanalysen und die generelle Optimierung von Encodereinstellungen können mit Hilfe von entsprechend genauen Qualitätsmodellen vorgenommen werden.

Infrastruktur und Netzwerk Analyse

Großen Einfluß auf QoS und QoE haben die Service-Infrastruktur und die verschiedenen Wege des OTT-Streams von der Origin über das CDN ins Edge-Netzwerk und schließlich zum Benutzer. Ausfälle von Teilen der Infrastruktur führen zu Einschränkungen in der Servicequalität, die verschiedenen Komponenten in der Streaming-Kette haben jeweils unterschiedliche Auswirkungen.

Encoder im Live-Betrieb übertragen unvorhersehbaren Content und können daher auch zeitweise unterschiedlich rechenintensive Szenen zur Verarbeitung bekommen. Das kann zu CPU-Engpässen führen und wirkt sich potentiell auf die resultierende Videoqualität aus. Die Antwortzeiten kommunizierender Komponenten können ebenfalls variieren, je nach Auslastung des Systems. Vor allem zu Peakzeiten kann dies zu Einschränkungen führen und andere Komponenten beeinflussen. Die Netzwerkbandbreite innerhalb des eigenen Netzwerks und zu IPSs kann durch die enorme Datenmenge von Video Content ebenfalls zu Engpässen sowie Unterbrechungen des Service führen. Die Auslieferung der Streams vom CDN über HTTP erzeugt Statuscodes. Durch Monitoring und Auswertung der Response Codes können hier direkt Rückschlüsse auf die QoS und indirekt Rückschlüsse auf negative Einflüsse der QoE beim User gegeben werden. Alle genannten Beispiele lassen sich mit

geeigneten QoS Metriken messen und geben einen Einblick auf Einflussfaktoren auf die Gesamtqualität, die beim Benutzer zu einem gegebenen Zeitpunkt möglich ist. Abseits der QoS-Metriken der Infrastruktur und des Netzwerks können durch Auswertung der Nutzungsdaten auch Rückschlüsse auf die tatsächliche QoE gegeben werden, wie sie von Kunden bei der Benutzung wahrgenommen wurde. Durch das Kombinieren von maximal zur Verfügung stehender Qualität und tatsächlich vom Player angefragter Qualität kann man feststellen, wie viele Nutzer gerade mit verminderter Qualität streamen. Hier sehen wir vor allem deutliche Unterschiede zwischen Managed OTT und OTT über das freie Internet (vgl. Abb. 2 OTT und Abb. 3 Managed OTT). Eine maximale Qualität erhalten zwischen 65% und 92% der Benutzer bei Unmanaged OTT Streaming. Die Rate hängt zum einen von der Tageszeit ab und zum anderen vom ISP (Internet Service Provider) der Nutzer. Managed OTT besticht dagegen durch hohe Prozentsätze von ca. 98,5%-99,5% an Nutzern mit Wiedergabe in maximaler Qualität, da kaum externe Einflussfaktoren den Service beeinträchtigen. Desweiteren kann zwischen Nutzung über mobile Netzwerke und festen Internetverbindungen unterschieden werden.

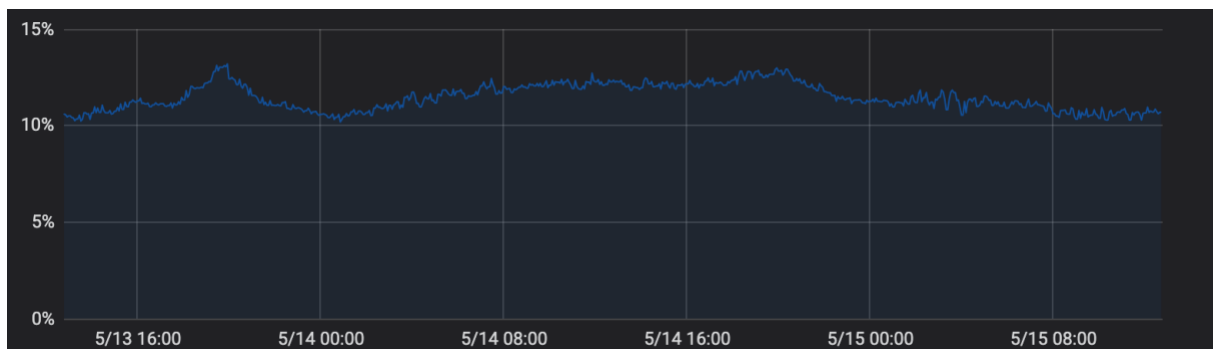


Abb. 2: Beispiel für prozentuale Verteilung un-managed OTT: Nutzer nicht auf maximaler Qualität, zw. 10% - 13%.

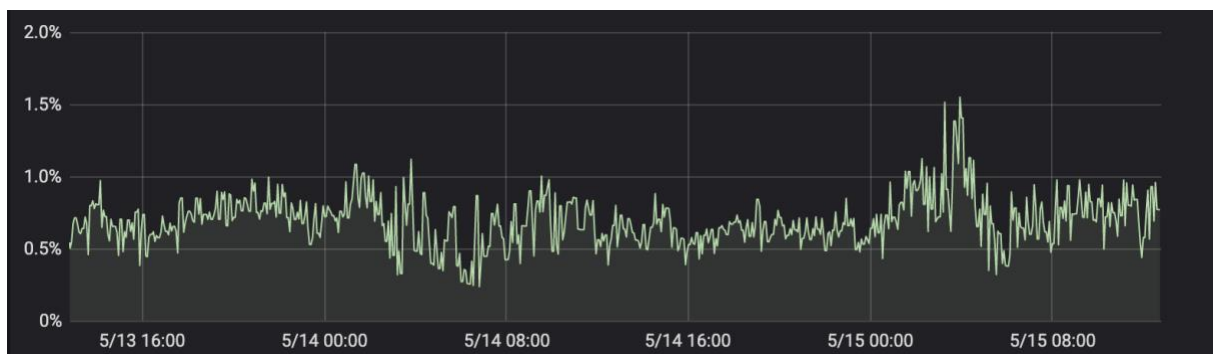


Abb. 3: Beispiel für prozentuale Verteilung managed OTT: Nutzer nicht auf maximaler Qualität, zw. 0,5% - 1,5%.

Netzwerk Optimierung

Zur Verbesserung der möglichen Beeinträchtigung der QoE durch Netzwerkeffekte gibt es inzwischen auch wirksame Technologien. Im Folgenden werden die aktuell vielversprechendsten Ansätze beschrieben.

SAND SRA

Die in SAND (DASH Part 5, Server and network assisted DASH, [ISO/IEC 23009-5](#)) definierte Funktion Shared Resource Allocation (SRA) ermöglicht Servern die Zuweisung von

Bandbreite an Streaming-Clients. Zu diesem Zweck sammelt der Server Streaming-Metriken der Clients. Die Serverkomponente sendet dann Bitraten-Zuweisungen an die Clients. Diese Bitrate muss im Client gehalten werden, z. B. durch Festlegen einer maximalen Bitrate im Player oder durch Traffic Shaping. Insgesamt ermöglicht dies Fairness und/oder Priorisierung der Clients, die an denselben Netzwerk Engpass (WiFi AP) angeschlossen sind, z. B. um die Anzahl der Clients zu maximieren, die einen Videostream flüssig wiedergeben können.

CMCD

CMCD definiert Metriken, die von einem Streaming-Client erfasst und mit jeder Segment- oder Manifestanfrage an das CDN gesendet werden. Dabei hilft CMCD dem CDN-Anbieter, die Gesamtleistung des CDN zu optimieren. Die in der Spezifikation definierten Metriken können für die "Protokollanalyse, QoS-Überwachung und Übertragungsoptimierung" nützlich sein. Außerdem können CDNs "ihren Datenverkehr kontrollieren und Performanceprobleme mit Player-Softwareversionen oder bestimmten Geräten korrelieren". Insgesamt kann ein CDN-Anbieter die CDN-Leistung und somit auch die QoE des Benutzers verbessern.

KPI / KQI Analyse im Player

[Streaming-Metriken](#), die im Video-Player erfasst werden, spiegeln wider was am Ende der Streaming-Kette bzw. vor Kunde geschieht. Insbesondere bei der Analyse von Streaming-Problemen sind diese Metriken wertvoll. Die Streaming-Metriken werden üblicherweise an einen Server geschickt (oft in Echtzeit), der die Daten vorhält und für eine detaillierte Analyse aufbereitet und visualisiert.

Es existieren verschiedene kommerzielle Lösungen im Markt, die ein Gesamtpaket für Streaming Analytics anbieten. Ein interoperabler Ansatz basiert auf dem [SAND-Standard \(Demo\)](#).

Auch für eine Messungen zur Audio- und Videoqualität gemäß ITU-T Rec. P.1203 und P.1204.3 stehen für Forschungszwecke [Implementierungen](#) zur Verfügung, die komplementär zu KPIs auch Informationen zur ausgespielten Qualität liefern.

QoE Messung auf Basis integrierter KQIs unter Betrachtung der auf Player-Seite vorliegenden Streaming-Session

Wenn auf Server- und/oder Client-Seite entsprechende KPIs und damit KQIs gemessen werden können - wie die Audio- und Videoqualität über die Zeit sowie qualitätsrelevante Ereignisse wie das "Stalling" eines Videostreams (s. Abschnitt 4) - können diese mittels eines Integrationsmodells wie ITU-T Rec. P.1203.3 zu einem QoE-Schätzwert integriert werden. Dieser ist ein Maß für die vom Endkunden bei typischen kurzen Streaming-Sessions von einigen Minuten Dauer wahrgenommene QoE (s. Abschnitt 4). Auch andere Integrationsmodelle für größere Zeiträume sind denkbar, oder Modelle für die Vorhersage von Abbrüchen einer Streaming-Session durch den Kunden.

Zusammenfassung

Wie eingangs erwähnt ist die systematische Erfassung der Nutzererfahrung (QoE) ein essentielles Anliegen der Betreiber von TV/Video Streaming Diensten. Die genauere Analyse des Themas und der in diesem Whitepaper gegebenen Überblick zeigen die Komplexität und den Facettenreichtum der heutigen Möglichkeiten.

Es lässt sich schwer eine einheitliche Empfehlung für die Mitglieder der Deutschen TV Plattform hinsichtlich der zu verwendenden Metriken und Technologien geben. Die richtige Wahl hängt stark von Zielsetzung, Qualitätsanspruch, technischen Möglichkeiten und Rahmenbedingungen ab.

Dieses White-Papier soll daher als gut fundierter Einstieg und als Überblick über die aktuell relevantesten Ansätze dienen. Bei der Erarbeitung des Papiers wurde klar, dass eine enge Verzahnung von praxisorientierter Wissenschaft und zukunftsorientierten Diensteanbietern ein sehr effizienter Weg ist, die Materie zu durchdringen und im Praxiseinsatz schnell Ergebnisse zu erzielen.

Die Autoren des Dokuments stehen gerne zum vertieften Austausch und Besprechen von Best Practices zur Verfügung.

Abkürzungsverzeichnis

ABR = Adaptive Bitrate Streaming
CDN = Content Delivery Network
CMAF = Common Media Application Format
CMCD = Common Media Client Data
CTA = Consumer Technology Association
DASH = Dynamic Adaptive Streaming over HTTP)
DRM = Digital Rights Management
FR = Full Referenz
ISP = Internet Service Provider
ITU = International Telecommunication Union
HAS = HTTP-based Adaptive Streaming
HLS = HTTP Live Streaming
KPI = Key Performance Indicator
KQI = Key Quality Indicator
MOS = Mean Opinion Score
NR = No Reference
PSNR = Peak Signal-to-Noise Ratio
QoE = Quality of Experience
QoS = Quality of Service
OTT = Over The Top
RR = Reduced Reference
SAND = Server and Network-assisted DASH
SRA = Shared Resource Allocation
SSIM = Structural Similiarity Index
VMAF = Video Multimethod Assessment Fusion

Appendix

- Komplementäre Quelle ITU Definitionen und regulatorische Betrachtungen:
[ITU-D “Quality of service regulation manual” \(2017\)](#), Geneva, Switzerland: ITU Telecommunication Development Bureau
- ITU-T Rec. P.10/G.100 (2017). Vocabulary for performance, quality of service and quality of experience. Geneva, Switzerland: International Telecommunication Union.
- P. Le Callet, S. Möller, and A. Perkis, eds (2012). Qualinet white paper on definitions of quality of experience. European network on quality of experience in multimedia systems and services (COST Action IC 1003), 3(2012).
- S. Möller, & A. Raake (Eds.) (2014). Quality of experience: advanced concepts, applications and methods. Springer.
- ITU-T Rec. P.800 (1996). Methods for subjective determination of transmission quality. Geneva, Switzerland: International Telecommunication Union.
- ITU-T Rec. P.910 (2008). Subjective video quality assessment methods for multimedia applications. Geneva, Switzerland: International Telecommunication Union.
- A. Raake, M. Garcia, W. Robitza, P. List, S. Göring and B. Feiten (2017). "A bitstream-based, scalable video-quality model for HTTP adaptive streaming: ITU-T P.1203.1," Ninth International Conference on Quality of Multimedia Experience (QoMEX), Erfurt, 2017, pp. 1-6, DOI:[10.1109/QoMEX.2017.7965631](https://doi.org/10.1109/QoMEX.2017.7965631)
- W. Robitza, S. Göring, A. Raake, D. Lindegren, G. Heikkilä, J. Gustafsson, P. List, B. Feiten, U. Wüstenhagen, M.-N. Garcia, K. Yamagishi, and S. Broom (2018). HTTP adaptive streaming QoE estimation with ITU-T Rec. P. 1203: Open databases and software. In Proceedings of the 9th ACM Multimedia Systems Conference (MMSys '18). Association for Computing Machinery, New York, NY, USA, 466–471.
DOI:<https://doi.org/10.1145/3204949.3208124>
- W. Robitza, D. G. Kittur, A. M. Dethof, S. Göring, B. Feiten and A. Raake (2018). Measuring YouTube QoE with ITU-T P.1203 Under Constrained Bandwidth Conditions. Tenth International Conference on Quality of Multimedia Experience (QoMEX), Cagliari, 2018, pp. 1-6, DOI:[10.1109/QoMEX.2018.8463363](https://doi.org/10.1109/QoMEX.2018.8463363)
- R. R. Ramachandra Rao, S. Göring, W. Robitza, A. Raake, B. Feiten, P. List, and U. Wüstenhagen (2020), “Bitstream-based Model Standard for 4K/UHD: ITU-T P.1204.3 -- Model Details, Evaluation, Analysis and Open Source Implementation”, Twelfth International Conference on Quality of Multimedia Experience (QoMEX), Athlone, pp. 1-6
- W. Robitza, A. Dethof, S. Göring, A. Raake, A. Beyer, T. Polzehl (2020). “Are you still watching? Streaming Video Quality and Engagement Assessment in the Crowd”, Twelfth International Conference on Quality of Multimedia Experience (QoMEX), Athlone, pp. 1-6
- S. Schwarzmann, N. Hainke, T. Zinner, C. Sieber, W. Robitza, and A. Raake (2020). Comparing fixed and variable segment durations for adaptive Video-Streaming: a holistic analysis. In Proceedings of the 11th ACM Multimedia Systems Conference (MMSys '20). Association for Computing Machinery, New York, NY, USA, 38–53.
DOI:<https://doi.org/10.1145/3339825.3391858>
- ITU-T Rec. P.1203 (2016). Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport. Geneva, Switzerland: International Telecommunication Union.

ITU-T Rec. P.1203.2 (2017). Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport – Audio quality estimation module. Geneva, Switzerland: International Telecommunication Union.

ITU-T Rec. P.1203.3 (2020). Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport - Quality integration module. Geneva, Switzerland: International Telecommunication Union.

ITU-T Rec. P.1204 (2020). Video quality assessment of streaming services over reliable transport for resolutions up to 4K. Geneva, Switzerland: International Telecommunication Union.

ITU-T Rec. P.1204.3 (2020). Video quality assessment of streaming services over reliable transport for resolutions up to 4K with access to full bitstream information. Geneva, Switzerland: International Telecommunication Union.

ITU-T Rec. P.1204.4 (2020). Video quality assessment of streaming services over reliable transport for resolutions up to 4K with access to full and reduced reference pixel information. Geneva, Switzerland: International Telecommunication Union.

ITU-T Rec. P.1204.5 (2020). Video quality assessment of streaming services over reliable transport for resolutions up to 4K with access to transport and received pixel information. Geneva, Switzerland: International Telecommunication Union.

Impressum

Herausgeber/Publisher:
Deutsche TV-Plattform e.V.
www.tv-plattform.de
Vereinsregister-Nr. 73VR9797

Leiter der AG Media over IP:
Dr. Niklas Brambring, Zattoo
Leiter der Task Force Delivery:
Peter Pogrzeba, Deutsche Telekom AG

Kontakt:
Deutsche TV-Plattform
Lyoner Str. 9, c/o ZVEI
60528 Frankfurt am Main
mail@tv-plattform.de
Tel.: 0049-69-6302-311
Fax: 0049-69-6302-361

Redaktion:
Task Force Delivery der Arbeitsgruppe
Media over IP der Deutschen
TV-Plattform

Autorenteam:
Alexander Raake / TU Ilmenau
Peter Pogrzeba / Deutsche Telekom AG
Nimesh Karia / Sky Deutschland
Stefan Arbanowski / Fraunhofer Fokus
Stefan Kaiser / Zattoo
Stefan Lietsch / Zattoo
Stefan Pham / Fraunhofer Fokus

Über die Deutsche TV-Plattform
Die Deutsche TV-Plattform ist ein
Zusammenschluss von über 50
Mitgliedern, darunter private und
öffentlich-rechtliche Sender, Streaming-
Anbieter, Gerätehersteller,
Internetunternehmen,
Infrastrukturbetreiber, Service- und
Technik-Provider, Forschungsinstitute
und Universitäten, Bundes- und
Landesbehörden sowie andere, mit den
digitalen Medien befasste Unternehmen,
Verbände und Institutionen. Ziel des
eingetragenen Vereins ist seit seiner
Gründung 1990 die Einführung digitaler
Technologien auf Grundlage offener
Standards.

Haftung:
Die Informationen in diesem Bericht
wurden genau und gründlich recherchiert
und im besten Wissen zusammengestellt
unter Beachtung des neutralen Ansatzes
der Arbeitsgruppe Media over IP / Task
Force Delivery der Deutschen TV-
Plattform. Alle Informationen spiegeln
den aktuellen Stand zum Zeitpunkt des
Redaktionsschlusses wider. Allerdings
können die Mitglieder der Arbeitsgruppe
und die Deutsche TV-Plattform nicht für
die Aktualität, Korrektheit, Vollständigkeit
und / oder Qualität der hier
veröffentlichten Informationen
garantieren. Deshalb sind
Haftungsansprüche gegen die Deutsche
TV-Plattform e.V. als Herausgeber
wegen materiellen und immateriellen
Schäden, die durch die Verwendung
dieser Publikation bzw. der
dargebotenen Inhalte bzw. durch die
Nutzung fehlerhafter oder unvollständiger
Informationen verursacht werden,
grundsätzlich ausgeschlossen.