Multimedia Streaming Analytics: Quo Vadis?

Cise Midoglu
cise@simula.no
SimulaMet
Oslo, Norway

Mariana Avelino
mariana.avelino@fokus.fraunhofer.de
Fraunhofer FOKUS
Berlin, Germany

Shri Hari Gopalakrishnan
shri.hari.gopalakrishnan@fokus.fraunhofer.de
Fraunhofer FOKUS
Berlin, Germany

Stefan Pham
stefan.pham@fokus.fraunhofer.de
Fraunhofer FOKUS
Berlin, Germany

Pål Halvorsen*
paalh@simula.no
SimulaMet
Oslo, Norway

ABSTRACT

In today's complex OTT multimedia streaming ecosystem, the task of ensuring the best streaming experience to end-users requires extensive monitoring, and such monitoring information is relevant to various stakeholders including content providers, CDN providers, network operators, device vendors, developers, and researchers. Streaming analytics solutions address this need by aggregating performance information across streaming sessions, to be presented in ways that help improve the end-to-end delivery. In this paper, we provide an analysis of the state of the art in commercial streaming analytics solutions. We consider five products as representatives, and identify potential improvements with respect to terminology, QoE representation, standardization and interoperability, and collaboration with academia and the developer community.

CCS CONCEPTS

• **Information systems** → **Multimedia streaming**; **Data analytics**; • **General and reference** → **Metrics**.

KEYWORDS

CMCD, CTA-2066, DASH, gap analysis, HLS, OTT, QoE, SAND, streaming analytics

ACM Reference Format:

Cise Midoglu, Mariana Avelino, Shri Hari Gopalakrishnan, Stefan Pham, and Pål Halvorsen. 2022. Multimedia Streaming Analytics: Quo Vadis?. In *Mile-High Video Conference (MHV '22), March 1–3, 2022, Denver, CO, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3510450.3517321

1 INTRODUCTION

Video streaming is one of the top traffic contributors in the Internet. It is expected that streaming traffic will grow 4-fold for video globally and 9-fold for mobile video between 2017 and 2022, with nearly 79% of the world's mobile data traffic coming from video by 2022 [1]. HTTP Adaptive Streaming (HAS) application variants such as Dynamic Adaptive Streaming over HTTP (DASH) and HTTP Live Streaming (HLS) are the dominant video delivery technologies used today, both for Video-on-Demand (VoD) and live streaming. Consequently, HAS is the biggest contributor to mobile video traffic.

With the rising prevalence of online video applications in recent years, corresponding video analytics practices and platforms have also grown in number and diversity. Over-the-top (OTT) video analytics platforms present data analytics related to OTT media streaming content and services, and are most commonly deployed as monitoring tools by broadcasters acting as Content Providers (CPs). Streaming analytics is crucial for numerous stakeholders and use cases. However, the development of online video analytics platforms is largely undertaken without the inclusion of the academic community, and research on the design, deployment, and evaluation of such platforms is scarce. To our knowledge, there is no taxonomy or comparative analysis of online video analytics platforms in academic literature.

In this paper, we provide an analysis of the state of the art in commercial streaming analytics solutions. We consider the top 5 products according to the Bitmovin Developer Report 2021 [12], excluding Google Analytics, as representatives. These products are offered by the companies Bitmovin [11], Conviva [19], MediaMelon [41], Mux [49], and Nice People At Work (NPAW) [52], and are listed in Table 1. Throughout this paper, we refer to the products with the name of their corresponding company, i.e., Bitmovin (B), Conviva (C), MediaMelon (Me), Mux (Mu), and NPAW (N). If multiple products are offered by the same company, we collectively refer to this group in the singular, using the abbreviations given above. All information presented in this paper is retrieved either from online materials that are publicly accessible at the time of writing or from explicit contact with company officials, and is valid as of 02.2022.

In Section 2, we introduce a taxonomy for the various functionalities of different products and propose a common terminology. In Section 3, we provide an overview of the practices surrounding the calculation and representation of Quality of Experience (QoE). In Section 4, we introduce a number of standards which are relevant for streaming analytics, and discuss interoperability. In Section 5, we provide a motivation for and an overview of collaboration opportunities, including a survey of academic publications referencing the products listed in Table 1. We conclude the paper in Section 6.

*Also affiliated with Oslo Metropolitan University, Norway

MHV '22, March 1–3, 2022, Denver, CO, USA

ACM ISBN 978-1-4503-9222-8/22/03.
https://doi.org/10.1145/3510450.3517321

Table 1: List of commercial products considered in this study: company name, the year the company was founded, and name of the product(s).

Company	Founded	Product(s)
Bitmovin	2013	Bitmovin Analytics
Conviva	2006	Conviva Stream Sensor
		Conviva Experience Insights
		Conviva Ad Insights
		Conviva Experience Benchmarks
		Conviva Precision
Mux	2015	Mux Data
NPAW	2008	Youbora Analytics
		Youbora Ads
		Youbora Users
		Youbora Infrastructure
MediaMelon	2012	MediaMelon SmartSight QoE
		MediaMelon SmartSight Ads
		MediaMelon SmartSight QBR

2 TAXONOMY AND TERMINOLOGY

In this section, we provide a general overview of product functionalities. We introduce a taxonomy and propose a common terminology for identifying different aspects of multimedia streaming analytics.

2.1 Collection of Metrics

Metrics, also called Key Performance Indicators (KPIs), refer to various measures of performance. Commercial products commonly employ an approach called Real User Monitoring (RUM), which refers to the strategy of measuring performance metrics from audience devices, rather than via synthetic monitoring solutions. We identify 4 main categories of streaming analytics metrics.

- **Audience:** These metrics are related to the audience (viewers) of a streaming service. They are generally collected from a service as a whole (in contrast to per-session metrics), and might be of more importance to actors such as service providers, who focus on avoiding churn and increasing monetization, and developers, who focus on service and platform stability and troubleshooting, rather than researchers, who focus on quality aspects.
- **Quality:** These metrics are related to the quality of streaming. They are generally collected per-session or per-subsession (in the form of "event"s with audio/video segment or chunk granularity).[1] We further divide this category of metrics into 3 subcategories.
  - **Startup:** These metrics are related to the time it takes the media asset to load and/or play.
  - **Buffer:** These metrics are related to buffering (also called "rebuffering" or "stalling"), as well as seeking.
  - **Quality:** These metrics are related to the adaptation for adaptive streaming (e.g., quality levels and quality switches), as well as picture quality for video (e.g., resolution and re-scaling).
- **Error**: These metrics are related to errors which might occur before or during a streaming session and interrupt playback. They are collected per-session and provide an overview, as well as troubleshooting opportunities, for various problems which prevent the execution of a smooth streaming session.[2]
- **Ad**: These metrics are related to the advertisement content that has been inserted into a stream (if applicable), which are typically collected per-session. Note that similar metrics might appear under both the "Quality" and the "Ad" categories, indicating that quality-related metrics are presented for the original content and ads separately. We do not consider ad metrics in the rest of this paper.

Table 1 in the supplementary material provides an overview of metrics collected by different products, grouped according to the above categories. We note that a very large number of metrics are provided, which are largely overlapping across products. However, there are different levels of detail, and different naming schemes across products, which, combined with the scarcity of public documentation, can make the meaning and use cases of certain metrics unclear at times.[3]

2.2 Collection of Metadata

Metadata refer to the types of background information which provide a context for the streaming session, and therefore insights into the interpretation of various performance metrics. We identify 6 main categories of metadata.

- **Identifiers:** These metadata are various identifiers for the streaming session.
- **Asset:** These metadata are related to the original media asset (audio and/or video). Examples include type, codec, and duration.
- **Client:** These metadata are related to the client device that the streaming session is running on. Examples include device model, platform, and Operating System (OS).
- **Application and transport:** These metadata are related to the application that is running on the client, as well as the underlying transport protocol. Examples include media player, web browser, mobile apk, and transport protocol.
- **Network:** These metadata are related to the network layer underlying the application and transport layers. Examples include Internet Service Provider (ISP), Mobile Network Operator (MNO), Autonomous System Number (ASN), Content Delivery Network (CDN), and IP address.
- **Spatio-temporal**: These metadata are related to location (country, city, etc.) and time (unix timestamp, time of day, day of week, month, year, etc.).

Table 2 in the supplementary material provides an overview of metadata collected by different products, grouped according to

[1] Raw quality metrics can be aggregated: (1) across all sessions of a streaming service, such as "Average Startup Time" which is the average of startup time across all sessions, or (2) across a single streaming session as a per-session metric, such as a per-session QoE score, which we elaborate upon in Section 3.

[2] Note that the other categories of metrics are considered for sessions which have no errors (e.g., a field such as "Startup Error" would be classified as an error metric, instead of a startup quality metric).

[3] Since our survey is based on public information, and customer data from commercial products are bound by various terms and conditions, we do not practically compare the calculation of metrics across providers (e.g., instrumenting a player with different analytics solutions, exposing it to a variety of controlled playback regimes, and comparing the data reported by the different products for the same metric). The results from such a study would not be possible to present in a non-anonymized fashion.

Table 2: Data export capabilities of different products (B: Bitmovin, Me: MediaMelon, Mu: Mux).

	Product				**Product**		
Cloud	**B**	**Me**	**Mu**	**Custom**	**B**	**Me**	**Mu**
AWS Kenesis			✓	API	✓	✓	✓
AWS S3	✓	✓	✓	**Formats**	**B**	**Me**	**Mu**
Azure Blob Storage	✓			Apache Parquet	✓	✓	
Azure DataLake		✓		CSV	✓	✓	✓
GCP Cloud Storage	✓		✓	JSON		✓	
GCP Pubsub			✓	Protobuf			✓

the above categories. Similarly to the metrics, we see that there is a slight difference in the focus placed on different metadata categories by different products, despite the large overlap of fields across products.

2.3 Data Export

The possibility of exporting raw data pertaining to a certain streaming service, such as the metrics and metadata described above, is of tremendous importance for academic or commercial in-house data analytics. Table 2 gives an overview of the data export capabilities of different products. Note that products differ in the means of data export they support, and not all of them provide such services.[4]

2.4 Dashboards and Visualization

All of the products we investigate are accompanied with a visualization functionality, in the form of an aggregate analytics dashboard. These can be accessed online with a valid account [13, 50] or via a demo request [21, 42, 53].

Individual playback sessions: Bitmovin, MediaMelon and Mux additionally provide details on individual streaming sessions in the form of an event timeline, updated in real-time ("Session details" for Bitmovin, "Microscope" for MediaMelon, and "Views" section for Mux). Details about individual sessions can be helpful to understand what happens during a playback, especially when a playback session has problems. Such interfaces are especially helpful for support teams trying to understand what a viewer actually experienced during playback, but developers will also be able to identify root causes from the timelines by recognizing similar playback experiences in ways that aren't easily expressed by metrics alone. A common workflow is to identify a population of viewers that are having poor experience, using the metrics, and then to look at the individual sessions that make up the aggregation, in order to better understand if and how there are commonalities across the playback sessions.

Live streaming: Bitmovin, MediaMelon, and Mux provide real-time dashboards (advertised latency for MediaMelon and Mux is around 30s), in addition to the VoD dashboard. These types of interfaces generally support fewer types of metadata, fewer metrics, and shorter retention for performance reasons, mostly due to operators wanting to focus only on the most important aspects that are likely to identify outages or issues in real-time.

[4]Conviva and NPAW have been removed from this analysis due to the lack of publicly available information.

Table 3: Integration capabilities of different products (B: Bitmovin, C: Conviva, Me: MediaMelon, Mu: Mux, N: NPAW).

	Product		
Integrations	**B**	**Me**	**Mu**
Google Data Studio	✓		
Grafana	✓ [8]	✓	✓
Tableau	✓ [9]	✓	
DataCoral (Cloudera)			✓
Datadog			✓
BigQuery	✓		✓

Big data and analytics integrations: In addition to visualization with their own analytics dashboard, some products also support integrations with external platforms for big data and analytics services. Table 3 gives an overview of these capabilities.[4]

2.5 Aggregate Views of Metrics

An aggregate view of a selected metric across different metadata dimensions is of crucial importance for data analysis and research. Consider the example: "the *median* of *total startup time* for all streaming sessions *in Norway*". Here, *median* serves as a statistical aggregator, *total startup time* is the metric of interest, and *in Norway* is a metadata dimension (namely, country) which is used as a breakdown (also called "filter"). Multiple filters, corresponding to different metadata dimensions, can also be used in cascade (e.g., "in Norway and using a Chrome browser"). Breakdowns and filters are the dimensions along which data can be aggregated before selected metrics are presented. These have many use cases ranging from performance benchmarking (across countries, networks, and devices, for instance) to troubleshooting and A/B testing. This functionality is provided in all dashboards.

2.6 Benchmarks

Two products allow their users to compare their service across the complete database of the company in an anonymized fashion, where explicit volunteer participation is required for the data from a particular service to be used in these "industry" comparisons. Bitmovin provides this information as "Industry Insights" [10], Conviva aggregates this information in their "State of Streaming" reports [22], and Mux exposes the average top line viewer score across all Mux customers. We refer to these comparisons as industry benchmarks.

2.7 Errors and Root Cause Analysis

Analytics solutions provide an abundance of information related to streaming sessions, but this information is only useful to operators and content/service providers as far as it can deliver actionable insights. One of the most important and challenging tasks is to quickly understand why, where and when errors happen and how to resolve them. This translates to the need for near real-time detection of issues in viewer experience, and the diagnosis of root causes. In this context, metrics and metadata cannot directly and definitively deliver answers, but instead serve as starting points for investigations in support of various troubleshooting efforts.

Streaming analytics products commonly market themselves as effective tools for root cause analysis [66]. Efforts in this direction range from help documents on metric and dashboard understanding [39, 51, 55] and error reporting [14, 57] to the use of Artificial Intelligence (AI) [20, 40, 46]. Video AI Alerts from Conviva [2] and Anomaly Detection from MediaMelon [40] are examples of the growing trend towards leveraging machine learning to enhance the speed and accuracy of root cause analysis.

3 QUALITY OF EXPERIENCE (QOE)

In this section, we focus on the representation of QoE.[5] We introduce different approaches to modeling QoE, and discuss these in relation to the practices of commercial products.

3.1 Modeling QoE

Composite/aggregate scores are commonly used provide an overall estimation of the QoE of individual streaming sessions. As opposed to the multitude of individual raw metrics, such scores function as a summary, and provide a means of quick comparison across sessions. There are numerous models proposed by various stakeholders, such as researchers, standardization bodies, and the industry, for the purpose of quantifying QoE as a single score.

Authors in [7] classify QoE models into 4 categories: *a*) signal-based models, *b*) parametric models, *c*) bitstream models, and *d*) hybrid models. *Signal-based* models, also known as pixel-based or media-layer models, utilize the decoded audio/video signal to estimate the video quality. Since such models do not use any codec specific information, they are widely used in codec comparison and optimization of unknown systems. Based on the amount of source (reference) information required, these can be further categorized as Full Reference (FR), Reduced Reference (RR), and No Reference (NR).[6] *Parametric* models use measured or expected packet/network related parameters to estimate quality. These can be further classified into packet-layer models and planning models. *Bitstream* models take into account the encoded bitstream and packet layer information. Features such as bitrate, frame rate, and Quantization Parameter (QP) are extracted and used as input. Such models are relatively computationally inexpensive and can be used for real-time QoE monitoring. *Hybrid* models combine two or more of the models mentioned above and hence can use much more information as input compared to any of the standalone models. Below, we list example models from different stakeholders.

Standardized models

- **P.1203:** The International Telecommunication Union (ITU) Recommendation P.1203 [54] defines a bitstream model for Mean Opinion Score (MOS) as a key perceptual metric. This models is used by authors in [60] to investigate the impact of individual objective metrics such as initial delay, stalling, and adaptation, on the output QoE score for streaming sessions. Model limitations include the lack of consideration of user-initiated state transitions such as pause, play, seek, end, and quality change, and the fact that the model is validated on relatively short viewing sessions ranging between 30*s* and 5*min* in duration.
- **P.1204:** ITU-T Recommendation P.1204 [58] is a NR bitstream model that predicts the subjective score of the video quality. The model has been trained on a large number of databases that includes videos with wide range of latest codecs and higher resolution up to 4K. However, it is not designed to model the impact of stalling and quality switches during the stream.

Academic models

- **FINEAS:** FINEAS [56] is a parametric statistical model that computes a QoE score on the basis of average quality levels and their standard deviations during the stream. The model also accounts for the impact of stall frequency and average stall duration. Specific weights are given to the impacts based on the studies [18, 25].
- **MPC:** Xiaoqi Yin et al. [69] proposed a parametric model to estimate the user precieved QoE based on selected bitrate, stalling and video quality switches during a stream. The score is then used to improve the client-side bitrate adaptation strategy using a Model Predictive Control (MPC) approach.

Industry models

- **VMAF:** Developed by Netflix, VMAF [37] is a FR model that fuses several temporal and spatial quality metrics to estimate a score on a linear scale of $1-100$. The model is trained over subjective test data from non-expert observers determining the visual degradation in encoded streams.
- **SSIMPLUS:** Developed by SSIMWAVE, SSIMPLUS [62] is a FR viewer metric using a linear scale of $0-100$ similar to VMAF, and is claimed to account for a variety of content complexity levels, frame rates, resolutions, dynamic range, and end-user devices.

Table 4 presents an overview of raw metrics and metadata required for the computation of four selected models.

3.2 Commercial Practices

Current commercial practices related to the calculation of aggregate session scores include the following.

- **Bitmovin:** Raw metrics displayed without any aggregate score.
- **Conviva:** Custom score called Streaming Performance Index (SPI) [20], calculated using Video Start Failure (VSF), Exits Before Video Start (EBVS), rebuffering, Video Playback Failures (VPF), Video Startup Time (VST), and picture quality.
- **MediaMelon:** Custom score called Q-Metric, calculated using Startup Failures, Startup Delays, Average Bitrate, Buffering Ratio, and Visual Quality (MOS).

[5] As a measure of the overall level of customer satisfaction with a video streaming service, QoE incorporates various components. These include perceptual quality (picture quality of the video presentation), streaming artifacts related to the playback (startup, buffering, errors) or the client (player software, UI elements), as well as the physical conditions in which a streaming session is conducted (distance to streaming device, hardware). For the sake of simplicity, we refer to any and all of these components as "QoE" in the following.

[6] The requirement for the reference information heavily affects the practical applicability of NR and RR models in streaming analytics products, as this information may or may not be available to the analytics integration in runtime, as well as the model requiring a lot more CPU power than most users have.

Table 4: List of raw metrics and metadata required for the computation of different models.

Raw Metrics	P.1203	P.1204	FINEAS	MPC
Startup delay	✓			✓
Highest quality			✓	
Used quality level	✓	✓	✓	✓
Num. quality switches	✓			✓
Stall duration	✓		✓	✓
Number of stalls	✓			✓
Rebuffering ratio			✓	
Average bitrate				✓
Metadata	**P.1203**	**P.1204**	**FINEAS**	**MPC**
Codec	✓	✓		
Distance from screen	✓	✓		
Client screen resolution	✓	✓		
Frame rate	✓	✓		
Duration	✓	✓		✓

- **Mux:** Custom score called Viewer Experience Score [48], composed of Playback Success Score (calculated using EBVS), Startup Time Score (calculated using startup time), Smoothness Score (calculated using rebuffer count and rebuffer percentage), and Video Quality Score (calculated using average and maximum upscale percentage).
- **NPAW:** Custom score called Happiness Score, public documentation unavailable.

The most notable observation is that the practices related to the calculation of a session score are not fully compatible across different products, as well as with standardized and academic approaches mentioned earlier. This is exacerbated by the fact that existing industry standards by ISO/IEC and CTA do not specify any composite session score, apart from the relatively well defined raw metrics and events.

As products already collect the majority of raw metrics required by various existing models (see Table 4 above and Table 1 in the supplementary material), it would technically be possible to compute an aggregate score as defined by one of these models, instead of presenting a custom score or none at all. However, there are a number of challenges associated with the integration of established models in a commercial context.

License/patent issues: Standardized models can be costly to use[7]. Academic models do not traditionally have any costs associated, but are sometimes patented (e.g., when they incorporate machine learning models trained on researchers' own datasets).

Implementation: The possibility and complexity of integrating new calculations into existing workflows depend on the software frameworks and development practices used by different products.

Model selection: Different models might emphasize different aspects of QoE, require different types of information, and ultimately be more suited to different target applications and audiences. For instance, implementing FR models such as VMAF might be unfeasible where reference information is not available to the analytics product. Similarly, while bitstream based models show comparatively higher correlation with subjective quality scores and are well-suited for real-time computation, they can only be used for specific codecs. As different models are influenced differently by raw metrics [59, 64], it is important to identify the relevant metrics based on streaming service type, asset length, content category, maximum screen resolution, and social/production value. It is also necessary to distinguish between VoD and live streaming.

Overall, there is no "silver bullet" model. However, it is possible to jointly establish an adaptive multi-model approach for the calculation of aggregate session scores, which is frequently adopted and interoperable across products, through well defined and realistic use cases, transparent methodology, open documentation and standardization.

4 STANDARDIZATION AND INTEROPERABILITY

In this section, we discuss the possibility of increasing interoperability across different analytics products through the use of standards.

- **SAND:** Server and Network Assisted DASH (SAND) is a standard specified in ISO/IEC 23009-5 [33] which defines how clients (e.g., video players), servers (e.g., CDNs), and networks (e.g., ISPs) should communicate with each other. Per SAND specifications, clients, servers and networks should exchange real-time status information, such as network metrics (e.g., bandwidth) and video player metrics (e.g., buffer level). In doing so, different network components can become aware of the current network conditions and adjust their behaviour to improve bandwidth utilization and streaming experience. SAND builds upon ISO/IEC 23009-5 Annex D for metric reporting.
- **CMCD:** Common Media Client Data (CMCD), or CTA-5004 is a specification published by the Consumer Technology Association (CTA) in the Web Application Video Ecosystem (WAVE) project [24]. The main idea behind CMCD is to enhance object requests from a video player to the CDN with information that can be useful in log analysis, quality of service monitoring and content delivery optimization.
- **CTA-2066:** Streaming Quality of Experience Events, Properties and Metrics (CTA-2066) [23] aims to standardize how streaming quality is measured and in this way makes comparisons of performance across media players and analytics solutions more objective. It prescribes standardized terminology and a minimum set of QoE events to be reported by media players, as well as a minimum set of QoE metrics to be computed by analytics solutions.

Interoperability across standards: CMCD metrics are object-based and meant for CDN optimizations. SAND (23009-1 Annex D) provides raw streaming metrics for time series data, which can only be partially mapped to CMCD. CTA-2066 provides a common terminology. Most of these metrics and events can be mapped to SAND. While the choice between SAND and CMCD metric reporting depends on the use case (end-user monitoring vs. CDN optimizations), we recommend aligning metric taxonomy with the definitions provided in CTA-2066. Table 1 in the supplementary material provides examples of relevant metrics defined by standards that correspond

[7] For a commercial license, respective rights holders of the standards ITU-T Rec. P.1203, ITU-T Rec. P.1203.1, ITU-T Rec. P.1203.2, and ITU-T Rec. P.1203.3 need to be contacted. See https://www.itu.int/en/ITU-T/ipr/Pages/default.aspx.

to product metrics, indicating a large gap in the current state of commercial practices.[8]

Interoperability across products: Interoperability also relates to the export of raw data (Section 2.3) in a cross-product context. If streaming analytics solutions use the same taxonomy to describe various metrics and metadata, the same raw dataset can be applicable for use by multiple products (e.g., visualization) through basic import/export operations. As long as the same taxonomy is used across different platforms, missing fields cease to be a problem.[9]

5 RESEARCH AND COLLABORATION

Finally, we focus on a number of possible areas for collaboration between the industry, academia, and the developer community. In Table 5, we present a list of academic publications referencing the product in Table 1. We use ACM Digital Library (ACM DL) [3] and IEEE Xplore [32] as literature search databases and use the company and product names listed in Table 1 as search strings. Overall, we see 3 academic publications using Bitmovin, and 22 publications using Conviva. Note here that almost all publications using commercial products are themselves authored by persons with company affiliation.

Commercial products provide tremendous opportunities to the academic community for research, with their numerous player integrations, extent of metrics and metadata, easy visualization functionalities, and ability to scale. However, as shown in Table 5, the academic popularity of the products in terms of independent, non-affiliated research is relatively low. Initiatives by the companies such as free product trials and scientific conference/workshop/challenge involvements could facilitate such interactions, along with the integration of established QoE models into product workflows, as mentioned in Section 3. Community resources such as reports, case studies, and white papers [15, 22, 43], alongside technical documentation, also play an essential role in facilitating research interactions.

Research topics which could benefit from industry collaboration include the evaluation of network operators and ISPs [45], development of adaptation algorithms [5, 35] and new QoE models [6], investigation of the influence of different transport protocols [45] and network generations [44] on streaming performance, media player benchmarks [45, 65], and the integration of AI into CDN selection and analytics pipelines [47].

6 CONCLUSION

In this paper, we provide an analysis of the state of the art in commercial streaming analytics solutions, using 5 products as representatives. We identify potential improvements with respect to terminology, QoE representation, standardization and interoperability, and collaboration with academia and the developer community. Our work is limited to information retrieved from online materials that are publicly accessible at the time of writing, or from explicit contact with company officials.

Throughout the paper, we advocate for increased communication and cooperation between various stakeholders involved in the end-to-end multimedia delivery pipeline. Given that monitoring information is crucial to many stakeholders, this exchange is very relevant, albeit challenging. In this respect, we plan to undertake a more detailed analysis of what we call "the incentive problem" as future work. The incentive problem refers to the inherent challenges related to the exchange of information between independent stakeholders who operate in a commercial and privacy-sensitive context. We also aim to present a proof-of-concept for an open-source lightweight community analytics solution which can serve as an interoperable core for collaborative analytics, with the goal of encouraging practices similar to those surrounding the development of Dash.js, and demonstrating the potential benefits of open resources and community contributions.

Table 5: List of academic publications using different products, sorted according to publication year. (*) Indicates author affiliation(s) from the corresponding company.

Publication	Year	Product
Chu et al. [31]	2002	Conviva*
Gummadi et al. [29]	2003	Conviva*
Stoica et al. [63]	2003	Conviva*
Chu et al. [17]	2004	Conviva*
Sripanidkulchai et al. [61]	2004	Conviva*
Greenberg et al. [28]	2005	Conviva*
Yan et al. [68]	2007	Conviva*
Zaharia et al. [71]	2008	Conviva*
Hindman et al. [30]	2009	Conviva*
Chowdhury et al. [16]	2011	Conviva*
Dobrian et al. [26]	2011	Conviva*
Jiang et al. [35]	2012	Conviva*
Zaharia et al. [70]	2012	Conviva*
Liu et al. [38]	2012	Conviva*
Balachandran et al. [6]	2013	Conviva*
Xin et al. [67]	2013	Conviva*
Ganjam et al. [27]	2015	Conviva*
Jiang et al. [34]	2016	Conviva*
Mukerjee et al. [47]	2017	Conviva*
Jiang et al. [36]	2017	Conviva*
Akhtar et al. [4]	2018	Conviva*
Akhtar et al. [5]	2018	Conviva*
Midoglu et al. [44]	2019	Bitmovin
Midoglu et al. [45]	2019	Bitmovin*
Taraghi et al. [65]	2020	Bitmovin*

ACKNOWLEDGMENTS

We would like to thank Phil Cluff and Steve Lyons from Mux, Pierre Hunter, Kumar Subramanian, and Samit Sanganee from MediaMelon, and Daniel Hoelbling-Inzko from Bitmovin for their valuable feedback and contributions to this work.

[8] Perhaps the most suitable standard for establishing interoperability across different products through common terminology and calculations, CTA-2066 is ironically not implemented by its own commercial contributors.

[9] However, interoperability is not only related to field naming, as data types and units need to be aligned as well. For instance, consider the seemingly small difference between `s` vs. `ms`, or `Int` vs. `Decimal`, and the complex repercussions of transferring a dataset of more than 50 fields, using one such schema, across multiple analytics platforms which might each have a slight mismatch for one or two different fields.

REFERENCES

[1] 2019. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper. https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.pdf

[2] 2022. Video AI Alerts. https://www.conviva.com/wp-content/themes/gj-boilerplate/img/ai-alerts/video-ai-alerts-whitepaper.pdf

[3] ACM. 2021. ACM Digital Library. Retrieved November 08, 2021 from https://dl.acm.org

[4] Zahaib Akhtar, Yun Seong Nam, Jessica Chen, Ramesh Govindan, Ethan Katz-Bassett, Sanjay Rao, Jibin Zhan, and Hui Zhang. 2018. Understanding Video Management Planes. In *Proceedings of the Internet Measurement Conference 2018* (Boston, MA, USA) *(IMC '18)*. Association for Computing Machinery, New York, NY, USA, 238–251. https://doi.org/10.1145/3278532.3278554

[5] Zahaib Akhtar, Yun Seong Nam, Ramesh Govindan, Sanjay Rao, Jessica Chen, Ethan Katz-Bassett, Bruno Ribeiro, Jibin Zhan, and Hui Zhang. 2018. Oboe: Auto-Tuning Video ABR Algorithms to Network Conditions. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication* (Budapest, Hungary) *(SIGCOMM '18)*. Association for Computing Machinery, New York, NY, USA, 44–58. https://doi.org/10.1145/3230543.3230558

[6] Athula Balachandran, Vyas Sekar, Aditya Akella, Srinivasan Seshan, Ion Stoica, and Hui Zhang. 2013. Developing a Predictive Model of Quality of Experience for Internet Video. *SIGCOMM Comput. Commun. Rev.* 43, 4 (aug 2013), 339–350. https://doi.org/10.1145/2534169.2486025

[7] Nabajeet Barman and Maria G. Martini. 2019. QoE Modeling for HTTP Adaptive Video Streaming–A Survey and Open Challenges. *IEEE Access* 7 (2019), 30831–30859. https://doi.org/10.1109/ACCESS.2019.2901778

[8] Bitmovin. 2018. Integrating Bitmovin Analytics with Grafana.

[9] Bitmovin. 2020. Integrate Analytics Data into Tableau.

[10] Bitmovin. 2020. What are Industry Insights? Retrieved January 09, 2022 from https://bitmovin.com/docs/analytics/faqs/what-are-industry-insights/

[11] Bitmovin. 2021. Adaptive streaming video infrastructure encoding and HTML5 player - Bitmovin. Retrieved December 13, 2021 from https://bitmovin.com/

[12] Bitmovin. 2021. Bitmovin Video Developer Report 2021. Retrieved December 13, 2021 from https://go.bitmovin.com/video-developer-report

[13] Bitmovin. 2021. Create Bitmovin Account. Retrieved December 15, 2021 from https://bitmovin.com/dashboard/signup

[14] Bitmovin. 2022. API Error Codes. Retrieved February 14, 2022 from https://bitmovin.com/docs/analytics/api-reference/help#api-error-codes/

[15] Bitmovin. 2022. Digital resources for online video - Bitmovin. Retrieved January 09, 2022 from https://bitmovin.com/resources/

[16] Mosharaf Chowdhury, Matei Zaharia, Justin Ma, Michael I. Jordan, and Ion Stoica. 2011. Managing Data Transfers in Computer Clusters with Orchestra. In *Proceedings of the ACM SIGCOMM 2011 Conference* (Toronto, Ontario, Canada) *(SIGCOMM '11)*. Association for Computing Machinery, New York, NY, USA, 98–109. https://doi.org/10.1145/2018436.2018448

[17] Yang-hua Chu, Aditya Ganjam, T. S. Eugene Ng, Sanjay G. Rao, Kunwadee Sripanidkulchai, Jibin Zhan, and Hui Zhang. 2004. Early Experience with an Internet Broadcast System Based on Overlay Multicast. In *Proceedings of the Annual Conference on USENIX Annual Technical Conference* (Boston, MA) *(ATEC '04)*. USENIX Association, USA, 12.

[18] Maxim Claeys, Steven Latré, Jeroen Famaey, Tingyao Wu, Werner Van Leekwijck, and Filip De Turck. 2014. Design and optimisation of a FAQ-learning-based HTTP adaptive streaming client. *Connection Science* 26 (01 2014), 25–43. https://doi.org/10.1080/09540091.2014.885273

[19] Conviva. 2021. Elevate Streaming with Conviva | 1 in Streaming Analytics. Retrieved December 13, 2021 from https://www.conviva.com/

[20] Conviva. 2021. OTT 101: Your Guide to Streaming Metrics that Matter|Conviva. Retrieved December 15, 2021 from https://www.conviva.com/ott-101-your-guide-to-streaming-metrics-that-matter/

[21] Conviva. 2021. Request Demo. Retrieved December 15, 2021 from https://www.conviva.com/request-demo/

[22] Conviva. 2022. State of Streaming. Retrieved January 09, 2022 from https://www.conviva.com/state-of-streaming/

[23] CTA. 2020. *Streaming Quality of Experience Events, Properties and Metrics (CTA-2066)*. Standard. Consumer Technology Association (CTA), Virginia, US. https://shop.cta.tech/products/streaming-quality-of-experience-events-properties-and-metrics

[24] CTA. 2020. *Web Application Video Ecosystem - Common Media Client Data - CTA-5004*. Standard. Consumer Technology Association (CTA), Virginia, US. https://cdn.cta.tech/cta/media/media/resources/standards/pdfs/cta-5004-final.pdf

[25] Johan De Vriendt, Danny De Vleeschauwer, and David Robinson. 2013. Model for estimating QoE of video delivered using HTTP adaptive streaming. In *2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)*. 1288–1293.

[26] Florin Dobrian, Vyas Sekar, Asad Awan, Ion Stoica, Dilip Joseph, Aditya Ganjam, Jibin Zhan, and Hui Zhang. 2011. Understanding the Impact of Video Quality on User Engagement. *SIGCOMM Comput. Commun. Rev.* 41, 4 (aug 2011), 362–373. https://doi.org/10.1145/2043164.2018478

[27] Aditya Ganjam, Junchen Jiang, Xi Liu, Vyas Sekar, Faisal Siddiqi, Ion Stoica, Jibin Zhan, and Hui Zhang. 2015. C3: Internet-Scale Control Plane for Video Quality Optimization. In *Proceedings of the 12th USENIX Conference on Networked Systems Design and Implementation* (Oakland, CA) *(NSDI'15)*. USENIX Association, USA, 131–144.

[28] Albert Greenberg, Gisli Hjalmtysson, David A. Maltz, Andy Myers, Jennifer Rexford, Geoffrey Xie, Hong Yan, Jibin Zhan, and Hui Zhang. 2005. A Clean Slate 4D Approach to Network Control and Management. *SIGCOMM Comput. Commun. Rev.* 35, 5 (oct 2005), 41–54. https://doi.org/10.1145/1096536.1096541

[29] K. Gummadi, R. Gummadi, S. Gribble, S. Ratnasamy, S. Shenker, and I. Stoica. 2003. The Impact of DHT Routing Geometry on Resilience and Proximity. In *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications* (Karlsruhe, Germany) *(SIGCOMM '03)*. Association for Computing Machinery, New York, NY, USA, 381–394. https://doi.org/10.1145/863955.863998

[30] Benjamin Hindman, Andy Konwinski, Matei Zaharia, and Ion Stoica. 2009. A Common Substrate for Cluster Computing. In *Proceedings of the 2009 Conference on Hot Topics in Cloud Computing* (San Diego, California) *(HotCloud'09)*. USENIX Association, USA, Article 19.

[31] Yang hua Chu, S.G. Rao, S. Seshan, and Hui Zhang. 2002. A case for end system multicast. *IEEE Journal on Selected Areas in Communications* 20, 8 (2002), 1456–1471. https://doi.org/10.1109/JSAC.2002.803066

[32] IEEE. 2021. IEEE Xplore. Retrieved November 08, 2021 from https://ieeexplore.ieee.org/Xplore/home.jsp

[33] ISO. 2017. *Information technology – Dynamic adaptive streaming over HTTP (DASH) – Part 5: Server and network assisted DASH (SAND)*. Standard. International Organization for Standardization (ISO), Geneva, CH. https://www.itu.int/rec/T-REC-P.1203

[34] Junchen Jiang, Vyas Sekar, Henry Milner, Davis Shepherd, Ion Stoica, and Hui Zhang. 2016. CFA: A Practical Prediction System for Video QoE Optimization. In *Proceedings of the 13th Usenix Conference on Networked Systems Design and Implementation* (Santa Clara, CA) *(NSDI'16)*. USENIX Association, USA, 137–150.

[35] Junchen Jiang, Vyas Sekar, and Hui Zhang. 2012. Improving Fairness, Efficiency, and Stability in HTTP-Based Adaptive Video Streaming with FESTIVE *(CoNEXT '12)*. Association for Computing Machinery, New York, NY, USA, 97–108. https://doi.org/10.1145/2413176.2413189

[36] Junchen Jiang, Shijie Sun, Vyas Sekar, and Hui Zhang. 2017. Pytheas: Enabling Data-Driven Quality of Experience Optimization Using Group-Based Exploration-Exploitation. In *Proceedings of the 14th USENIX Conference on Networked Systems Design and Implementation* (Boston, MA, USA) *(NSDI'17)*. USENIX Association, USA, 393–406.

[37] Zhi Li, Anne Aaron, Ioannis Katsavounidis, Anush Moorthy, and Megha Manohara. 2016. Toward A Practical Perceptual Video Quality Metric. Retrieved December 13, 2021 from https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652

[38] Xi Liu, Florin Dobrian, Henry Milner, Junchen Jiang, Vyas Sekar, Ion Stoica, and Hui Zhang. 2012. A Case for a Coordinated Internet Video Control Plane. *SIGCOMM Comput. Commun. Rev.* 42, 4 (aug 2012), 359–370. https://doi.org/10.1145/2377677.2377752

[39] Steven Lyons. 2020. Respond to and Resolve Incidents with the Real-Time Dashboard. Retrieved February 14, 2022 from https://mux.com/blog/respond-to-and-resolve-incidents-with-the-real-time-dashboard/

[40] MediaMelon. 2021. Automated Anamoly Detection for Rapid Resolution | OTT Video Analytics. Retrieved December 15, 2021 from https://mediamelon.com/automated-anomaly-detection-for-rapid-resolution/

[41] MediaMelon. 2021. OTT Video Intelligence & Optimization Platform | MediaMelon. Retrieved December 13, 2021 from https://mediamelon.com/

[42] MediaMelon. 2021. Request Demo. Retrieved December 15, 2021 from https://info.mediamelon.com/demonstration-rq-mediamelon

[43] MediaMelon. 2022. Resources - White Papers - Case Studies - Datasheets - MediaMelon Inc. Retrieved January 09, 2022 from https://mediamelon.com/resources/

[44] Cise Midoglu, Özgü Alay, and Carsten Griwodz. 2019. Evaluation Framework for Real-Time Adaptive 360-Degree Video Streaming over 5G Networks. In *Proceedings of the 2019 on Wireless of the Students, by the Students, and for the Students Workshop* (Los Cabos, Mexico) *(S3'19)*. Association for Computing Machinery, New York, NY, USA, 6–8. https://doi.org/10.1145/3349621.3355726

[45] Cise Midoglu, Anatoliy Zabrovskiy, Ozgu Alay, Daniel Hoelbling-Inzko, Carsten Griwodz, and Christian Timmerer. 2019. Docker-Based Evaluation Framework for Video Streaming QoE in Broadband Networks. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France) *(MM '19)*. Association for Computing Machinery, New York, NY, USA, 2288–2291. https://doi.org/10.1145/3343031.3350538

[46] John Moulding. 2018. Conviva uses AI-based learning to automate root cause analysis, removing performance headaches for streamers. Retrieved February 14, 2022 from https://www.v-net.tv/2018/01/18/conviva-uses-ai-based-learning-to-automate-root-cause-analysis-removing-performance-headaches-for-

streamers/
[47] Matthew K. Mukerjee, Ilker Nadi Bozkurt, Devdeep Ray, Bruce M. Maggs, Srinivasan Seshan, and Hui Zhang. 2017. Redesigning CDN-Broker Interactions for Improved Content Delivery. In *Proceedings of the 13th International Conference on Emerging Networking EXperiments and Technologies* (Incheon, Republic of Korea) *(CoNEXT '17)*. Association for Computing Machinery, New York, NY, USA, 68–80. https://doi.org/10.1145/3143361.3143366
[48] Mux. 2018. Changes to the Mux Viewer Experience Score. Retrieved January 09, 2022 from https://mux.com/blog/changes-to-the-mux-viewer-experience-score/
[49] MUX. 2021. The API to Video | Mux. Retrieved December 13, 2021 from https://mux.com/
[50] Mux. 2021. Create Mux Account. Retrieved December 15, 2021 from https://dashboard.mux.com/login
[51] Mux. 2022. Understand metric definitions. Retrieved February 14, 2022 from https://docs.mux.com/guides/data/understand-metric-definitions/
[52] NPAW. 2021. NPAW – The global video intelligence company - Video analytics and more. Retrieved December 13, 2021 from https://npaw.com/
[53] NPAW. 2021. Request Demo. Retrieved December 15, 2021 from https://npaw.com/request-demo/
[54] Telecommunication Standardization Sector of ITU. 2017. *Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport.* Standard. International Telecommunication Union (ITU), Geneva, CH. https://www.itu.int/rec/T-REC-P.1203
[55] Bonnie Pecevich. 2019. Get to the root cause (even) faster with new real-time metric breakdowns. Retrieved February 14, 2022 from https://mux.com/blog/get-to-the-root-cause-even-faster-with-new-real-time-metric-breakdowns/
[56] Stefano Petrangeli, Jeroen Famaey, Maxim Claeys, Steven Latré, and Filip De Turck. 2015. QoE-Driven Rate Adaptation Heuristic for Fair Adaptive Video Streaming. *ACM Trans. Multimedia Comput. Commun. Appl.* 12, 2, Article 28 (oct 2015), 24 pages. https://doi.org/10.1145/2818361
[57] Christoph Prager. 2020. Metrics That Matter: Top-Down Error Reporting. Retrieved February 14, 2022 from https://bitmovin.com/top-down-error-reporting/
[58] Rakesh Rao Ramachandra Rao, Steve Göring, Peter List, Werner Robitza, Bernhard Feiten, Ulf Wüstenhagen, and Alexander Raake. 2020. Bitstream-Based Model Standard for 4K/UHD: ITU-T P.1204.3 — Model Details, Evaluation, Analysis and Open Source Implementation. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. 1–6. https://doi.org/10.1109/QoMEX48832.2020.9123110
[59] Anika Seufert, Florian Wamser, David Yarish, Hunter Macdonald, and Tobias Hoßfeld. 2021. QoE Models in the Wild: Comparing Video QoE Models Using a Crowdsourced Data Set. In *2021 13th International Conference on Quality of Multimedia Experience (QoMEX)*. 55–60. https://doi.org/10.1109/QoMEX51781.2021.9465422
[60] Michael Seufert, Nikolas Wehner, and Pedro Casas. 2018. Studying the Impact of HAS QoE Factors on the Standardized QoE Model P.1203. *IEEE 38th International Conference on Distributed Computing Systems (ICDCS)* (2018).
[61] Kunwadee Sripanidkulchai, Aditya Ganjam, Bruce Maggs, and Hui Zhang. 2004. The Feasibility of Supporting Large-Scale Live Streaming Applications with Dynamic Application End-Points. In *Proceedings of the 2004 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications* (Portland, Oregon, USA) *(SIGCOMM '04)*. Association for Computing Machinery, New York, NY, USA, 107–120. https://doi.org/10.1145/1015467.1015480
[62] SSIMWAVE. 2021. SSIMPLUS - SSIMWAVE. Retrieved December 13, 2021 from https://www.ssimwave.com/technology/
[63] Ion Stoica, Robert Morris, David Liben-Nowell, David R. Karger, M. Frans Kaashoek, Frank Dabek, and Hari Balakrishnan. 2003. Chord: A Scalable Peer-to-Peer Lookup Protocol for Internet Applications. *IEEE/ACM Trans. Netw.* 11, 1 (feb 2003), 17–32. https://doi.org/10.1109/TNET.2002.808407
[64] Babak Taraghi, Minh Nguyen, Hadi Amirpour, and Christian Timmerer. 2021. Intense: In-Depth Studies on Stall Events and Quality Switches and Their Impact on the Quality of Experience in HTTP Adaptive Streaming. *IEEE Access* 9 (2021), 118087–118098. https://doi.org/10.1109/ACCESS.2021.3107619
[65] Babak Taraghi, Anatoliy Zabrovskiy, Christian Timmerer, and Hermann Hellwagner. 2020. CAdViSE: Cloud-Based Adaptive Video Streaming Evaluation Framework for the Automated Testing of Media Players. In *Proceedings of the 11th ACM Multimedia Systems Conference* (Istanbul, Turkey) *(MMSys '20)*. Association for Computing Machinery, New York, NY, USA, 349–352. https://doi.org/10.1145/3339825.3393581
[66] Telestream. 2021. Telestream and NPAW offer advanced integration of video network and client analytics. Retrieved February 14, 2022 from https://www.telestream.net/company/press/2021-06-29-NPAW.htm/
[67] Reynold S. Xin, Josh Rosen, Matei Zaharia, Michael J. Franklin, Scott Shenker, and Ion Stoica. 2013. Shark: SQL and Rich Analytics at Scale. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data* (New York, New York, USA) *(SIGMOD '13)*. Association for Computing Machinery, New York, NY, USA, 13–24. https://doi.org/10.1145/2463676.2465288
[68] Hong Yan, David A. Maltz, T. S. Eugene Ng, Hemant Gogineni, Hui Zhang, and Zheng Cai. 2007. Tesseract: A 4D Network Control Plane. In *Proceedings of the 4th USENIX Conference on Networked Systems Design amp; Implementation* (Cambridge, MA) *(NSDI'07)*. USENIX Association, USA, 27.
[69] Xiaoqi Yin, Abhishek Jindal, Vyas Sekar, and Bruno Sinopoli. 2015. A Control-Theoretic Approach for Dynamic Adaptive Video Streaming over HTTP. *SIGCOMM Comput. Commun. Rev.* 45, 4 (aug 2015), 325–338. https://doi.org/10.1145/2829988.2787486
[70] Matei Zaharia, Mosharaf Chowdhury, Tathagata Das, Ankur Dave, Justin Ma, Murphy McCauley, Michael J. Franklin, Scott Shenker, and Ion Stoica. 2012. Resilient Distributed Datasets: A Fault-Tolerant Abstraction for in-Memory Cluster Computing. In *Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation* (San Jose, CA) *(NSDI'12)*. USENIX Association, USA, 2.
[71] Matei Zaharia, Andy Konwinski, Anthony D. Joseph, Randy Katz, and Ion Stoica. 2008. Improving MapReduce Performance in Heterogeneous Environments. In *Proceedings of the 8th USENIX Conference on Operating Systems Design and Implementation* (San Diego, California) *(OSDI'08)*. USENIX Association, USA, 29–42.