

Article

Multi-Layer QoE Learning System Implemented by Fiberhost

Paweł Andruloniw ^{1,2} , Karol Kowalik ¹, Bartosz Partyka ¹ and Piotr Zwierzykowski ^{2,*} ¹ Fiberhost S.A., Wysogotowo, ul. Wierzbowa 84, 62-081 Przeźmierowo, Poland² Institute of Computer and Communication Networks, Faculty of Computing and Telecommunications, Poznań University of Technology, Pl. Marii Skłodowskiej-Curie 5, 60-965 Poznań, Poland

* Correspondence: piotr.zwierzykowski@put.poznan.pl

Abstract: Modern telecommunications networks, despite their ever increasing capacity, mostly attributed to optical fiber technologies, still fail to provide ideal channels for transmitting information. Disruptions in ensuring data throughput or the continuous flow of data required by applications remain as major unresolved problems. Most network mechanisms, protocols and applications feature adaptations that allow them to change the parameters of the transmission channel and try to minimize the negative impact of the network on the perceived quality, for example by temporarily changing the modulation scheme, or coding scheme, or by re-transmitting lost packets, or buffering to compensate for the interruptions in transmission. To respond appropriately, network operators are interested in knowing how well these adaptations are performing in order to assess the ultimate quality of their networks from the user perspective, i.e., Quality of Experience (QoE). Due to the huge amount of data associated with the collection of various parameters of the telecommunications network, machine learning methods are often needed to discover the relationships between various parameters and to identify the root cause of the observed network quality. In this paper, we present a Multi-layer QoE learning system implemented by Fiberhost for QoE analysis with a multi-layer approach based on machine learning tools.

Keywords: multi-layer QoE; learning system; fiberhost

Citation: Andruloniw, P.; Kowalik, K.; Partyka, B.; Zwierzykowski, P. Multi-Layer QoE Learning System Implemented by Fiberhost. *Appl. Sci.* **2023**, *13*, 2300. <https://doi.org/10.3390/app13042300>

Academic Editor: Mirosław Klinkowski

Received: 8 December 2022

Revised: 2 February 2023

Accepted: 8 February 2023

Published: 10 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Telecommunications operators are generally well versed in responding to alerts from network devices that suggest cable cuts, power outages, or major service disruptions. However, the identification of service degradation based on real-time analysis of the customer data and experience is a much more complex task. Due to the large amount of data obtained from monitoring systems, it is not possible to analyze the impact of each change of parameters on the global condition of the operator's infrastructure, and only the direct causes of service degradation are analyzed. In cases where there is no obvious cause for degradation, it is necessary to analyze logs, fragments (chunks) of network traffic and the configuration of devices or services, which significantly extends the time of problem resolution. In general, network fault detection and management is very challenging due to the existence of a variety of devices and their configuration and different operators in the end-to-end path. Moreover, various nodes or links within the end-to-end path may belong to different network owners, including the end customers. This results in time consuming efforts to detect and remedy any deterioration in service quality, or a malfunction of equipment or structure in good time, and the limited ability of the operator to prevent them.

1.1. Commercial Solutions

There exist commercial network management platforms that, in addition to fulfilling their primary objectives, are capable of delivering quality of service (QoS)-related statistics of the networks, such as Dynatrace [1] dedicated to virtual environments, or those provided

by hardware vendors, such as Huawei's NCE [2] or Nokia's NSP [3] provided as the extension of vendor-specific technology solutions. From the perspective of the TV service, Witbe Workbench [4] and Enensys StreamProbe [5] allow the behavior of clients on the set-top box (STB) to be simulated, as well as the quality of experience (QoE) associated with the video stream displayed on the STB to be evaluated. As a result, they allow numerous maintenance and operational tasks to be optimized, which in turn can improve operational efficiency and reduce field work. It should also be noted that commercial solutions often focus on delivering the concept of Intent-Based Networking [6,7]. Theoretically, this should bridge the gap between the network and business goals, using artificial intelligence and machine learning. Such an Intent-Based Network should enhance preventive maintenance and increase the reliability of the network. Existing commercial solutions, however, are only the first steps to achieve the described goals and do not focus on QoE estimates, which was our primary goal.

1.2. Related Works

A number of studies introduce various QoE models that can be utilized by network operators in practice. Predictive models based on automatic modelling, decision trees and linear regression have been used to estimate the sustainable QoE level based on the survey referenced to contextual-relational, legal-regulatory, subjective-user, technological-process, content-formatted and performative factors [8]. The Q-meter system has the advantage of detecting customers' complaints using deep learning-based models. The sentiment analysis can be applied to determine the cause of decreased quality level and detect service failures within cellular networks. This real-time system uses complaints posted on online social networks for degradation detection [9]. The full-reference method, alongside supervised and unsupervised algorithms, has been used to estimate the QoE level for multicast-based video environments. The QoS parameters include delay, jitter, and packet loss, as well as APSNR, SSIM, and VQM video parameters and have been used to achieve the required QoE level and adapt the video signal to increase the perceived quality. From the network operator's perspective, estimations can be performed for a single or group of viewers due to tv channel clustering [10]. WebQoE based on decision trees, kNN, and SVM have been introduced to evaluate the QoE level based not only on video-related parameters i.e., initial delay, stalls, or quality switches, but also on user behavior and social context. Moreover, the authors introduced the QoE influence factors (IFs) classification to system, human, context and service metrics [11]. The data-driven approach for video quality estimation has been proposed to assess QoE in a mobile environment. 89 network features related to a mobile network, video characteristics and playback were gathered by mobile applications. In addition, the MOS distortion grades on a 1–5 scale were obtained. Over 80,000 samples were used to train and evaluate the deep neural network in a batch manner [12]. D-DASH framework that combines deep learning and reinforcement has been proposed to assess the QoE level for video streams based on the HTTP DASH (Dynamic Adaptive Streaming over HTTP) standard. Authors employed feed-forward and recurrent deep neural networks to utilize features related to the video quality of DASH segments and playback stalls [13]. DeepQoE is an end-to-end framework that combines word embedding and a 3D convolutional neural network to extract generalized features and fed neural networks for representation learning. Regression and classification are used to retrieve the final assessment. This novel approach has the advantage of calculating a QoE score in different environments [14]. The idea of unsupervised deep learning techniques makes it possible to perform quality assessment in an online mode. The no-reference features related to the pixel and bitstream layer were employed to train the restricted Boltzmann machine model in an offline mode, which then can be sent to the client side to assess the continuous approach. The model was evaluated with the LIMP Video Quality Database [15]. The content-aware and distortion features were used to feed the GRU neural network to assess the video sequence quality. Distortions that occurred during video acquisition and those related to compression or packet transmission were detected, and as a result the

video quality was assessed by a single neural network. Additionally, the authors proposed a pooling mechanism to detect distorted frames that significantly influenced the perceived video quality [16]. The HA-SVQA algorithm has been proposed to speed up the process of subjective video quality assessment. The solution based on active learning eliminates less valuable video sequences, considering the image quality features, encoding, and statistical parameters. Afterwards only meaningful sequences can be processed by an objective quality assessment, which can be useful from the point of view of TV operators that have to deal with thousands of sequences for non-linear functions of the TV service [17]. The RBM deep learning method has been applied to assess the video quality of HTTP DASH streams in a TV environment. The authors used the bitstream, motion and video parameters and distortions to assess the no-reference video quality for digital video broadcasting channels. This framework might meet TV operators' needs within the OTT stream quality assessment, but might be difficult to be applied for hardware-sealed STBs used by customers to receive TV streams [18].

1.3. Overview

Commercial solutions focus on a specific technology and do not allow a multi-layer analysis to be performed, where each implementation layer is not based on a single vendor, and some layers are even specific to the operator's internal processes. Therefore, we found it necessary to extend Fiberhost's existing system with dedicated modules that would enable the use of machine learning methods to create multi-layer QoE learning specifically tailored to the needs of the operator. In addition, our aim was to compare the QoE from various angles, such as QoS metrics, geographic location, technology, human interactions, etc. On account of such a wide scope, Fiberhost realized that there were no ready-to-use third party systems to be implemented and decided to develop this system in-house.

This article is divided into seven parts. In Section 2, the operator network architecture used for data collection is described. Section 3 presents the impact of the infrastructure and the different service layers of the carrier network on QoE. Section 3.1 outlines the problems in the optical access network infrastructure and their impact on QoE. Section 3.2 discusses potential problems that may arise at the ISP level. Sections 3.3 and 3.4 present the sources of QoE degradation at the TV service layer and the telephony service layer, respectively. Section 4 presents a proposal for an aggregate metric that Fiberhost uses to assess QoE. This section also presents its rationale and the assessment of the network quality measured by the proposed metric. Section 5 presents the proposed implementation of unsupervised machine learning methods for assessing QoE in a carrier network, while Section 6 proposes the use of supervised learning in the estimation of the parameters used to determine QoE in a carrier network. The paper concludes with a brief summary.

2. Data Collection Infrastructure

At Fiberhost—a Fiber-to-the-Home (FTTH) operator from Poland—a new architecture of Multi-layer QoE learning system is being implemented and developed in order to simplify the problem of QoE analysis. The system is based on two measurement techniques for data collection: Active probing (for example RFC 2544 [19] or ITU-T Y.1564 [20]) and Passive polling (implemented by observing counters of Simple Network Management Protocol (SNMP) [21] or other performance indicators available for example in the performance logs).

The system architecture is presented in Figure 1. The left part of the figure shows the access networks and the data collection process organized in the form of querying access devices (CMTS—Cable Modem Termination System or OLT—Optical Line Termination), using the SNMP protocol or end devices (CMs—Cable Modems, or ONTs - Optical Network Terminals), using SNMP or TR-069 [22] protocols. In the left section, one can also identify the Probes which are performing active service probing on the selected areas of the network. Except for the probes, the SNMP and TR-069 are not intrusive and implement the use of the passive probing technique. Only the Probe devices which are connected to Customer

Promises Equipment's (CPEs), such as ONTs or CMs, implement active probing at the user layer, thus they try to simulate the behavior of a typical end user. The probes are not on site, but are in the technical cabinets closest to the end customer's location, and the entire probe population is in turn managed by the probe management and data collection system [23].

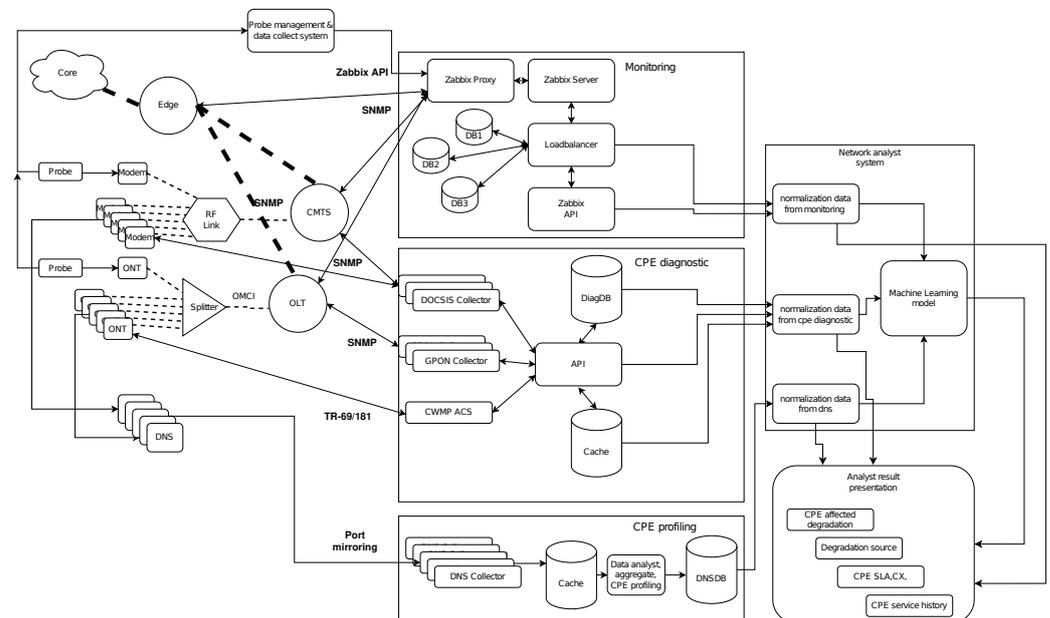


Figure 1. Data Collection Architecture.

The central part of Figure 1 is dedicated to the monitoring servers and databases that store all control information. The central part consists of three main functional blocks:

- Service and device Monitoring—which is the part responsible for monitoring of the health of all network devices, except CPEs and services that are configured on these devices, as well as performance Key Performance Indicators (KPIs) such as temperature, CPU load, memory utilization, etc.
- CPE diagnostics—which is the part responsible for monitoring of CPEs as well as allowing simple diagnostics to be conducted, such as ping, etc.
- CPE profiling—which is still under construction, and which will enable user behavior profiling, though this will require the end customer's consent.

The functional block shown in the right part of Figure 1: “Network analyst system” and “Analyst result presentation” will support various data models to be used by machine learning and presentation mechanisms [24].

The right part of Figure 1 is still in progress. It includes data filtering and normalization, machine learning models, analysis and the presentation of the results. Therefore, this part covers most of the mechanisms described further on in this article.

Our implementation of the multi-layer QoE learning is built on the basis of an open source solution: Zabbix [25], but was expanded to include a number of proprietary modules. A choice of a commercial solution supporting machine learning would be impractical on account of the need for necessary and time-consuming changes, preparation of new procedures and implementation of the new tool throughout the organization. Zabbix itself does not have machine learning mechanisms implemented, but a simple predictive model that allows one to analyze parameter values based on historical trends. However, we found that Zabbix was relatively easy to extend. Therefore, we used it as the basis for our own multi-layer QoE learning system.

3. Multi-Layer QoE Metric Nature

The essence of the Internet is that it is used not only for websites, but also for access to video or voice content, or a mix of various types of data in conjunction with the implemen-

tation of interactivity, which also imposes certain requirements on the quality of service. Therefore, it is recommended to use compound QoS metrics to evaluate not just a single parameter, but many of them, which in turn reflect the observed quality of service, such as the parameter called: Speech and Multimedia Transmission Quality (STQ) [26]. STQ contains recommended practices for testing mobile networks, but can also be used by ISPs operating over wired networks. It is based on a multi-tiered approach, whereby the total test score is calculated from the weighted scores of the underlying tiers. The STQ does not have fixed weights but recommends weights for each indicator, and the examples of the weighting factors, limits, and thresholds are given in Annex A to [26].

3.1. Infrastructure Layer

The Infrastructure layer is the base layer to all the other layers, since it defines the Network Availability that the following services rely on.

For example, cable cuts and power outages are the most common causes of service disruptions. This can be seen in Figure 2, where we classified the major service outages in a typical year at Fiberhost. It is observable that the cable cuts and power outages contribute to more than 94% of all service outages. Therefore, when building telecommunications networks, we should strive for the maximum dispersion and diversification of the resources. Additionally, it is important to increase the existing network capacity and density, so that efficient resources take over the function of corrupted resources when needed.

Machine learning methods are not required to analyze this type of data, however they can be useful in performing the root cause analysis to identify the source of a problem and reduce the number of alarms the operator needs to analyze.

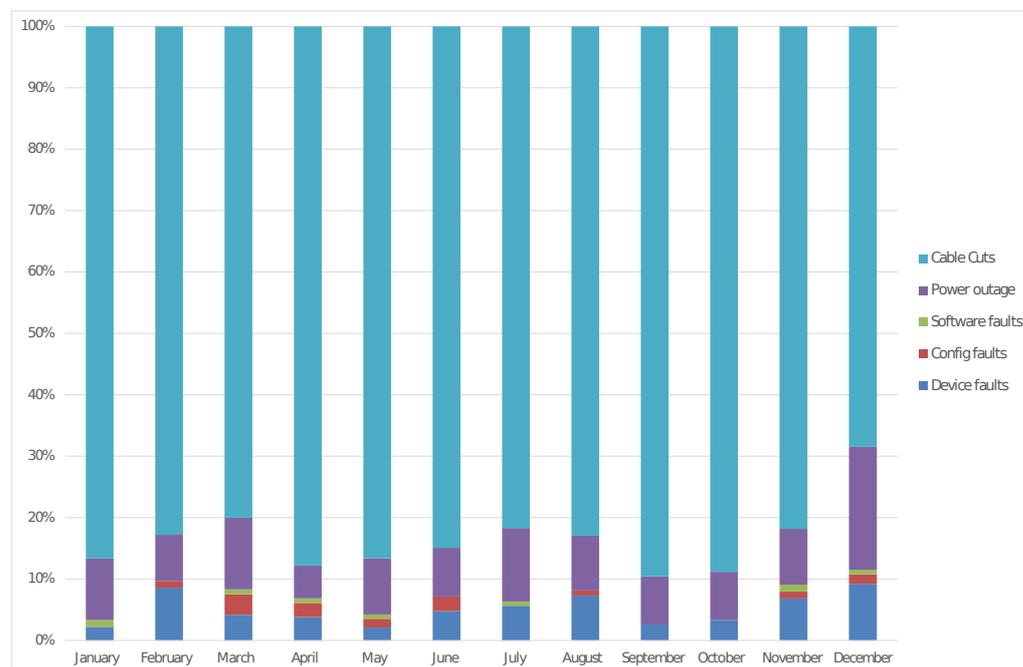


Figure 2. Reasons for service disruptions.

The signal level parameters belong to the infrastructure layer as well. These can be used to check if the optical budget of the optical link is met. The distribution of the signal levels received at the 1310 nm and 1490 nm wavelengths in the Fiberhost GPON network is shown in Figure 3. Even though the distribution in Figure 3 shows that some ONTs are below the optical budget, it does not necessarily mean that the service will be interrupted. In addition, we found that the levels of signals received on the 1310 nm and 1490 nm wavelengths in the GPON optical links do not correlate with bit errors, as it is shown in Figure 4. The bit errors include both bit-interleaved parity (BIP) errors, which provide

measurement of the bit errors for the GPON frame header (PCBd), and HEC errors, which indicate the bit errors on the GEM header (the header of the payload). However, as shown in the Pearson’s correlations diagram in Figure 4, the correlation between bit errors does not exist. This is because dirty connectors, broken fibers, etc, may be present even if there is not a significant signal attenuation.

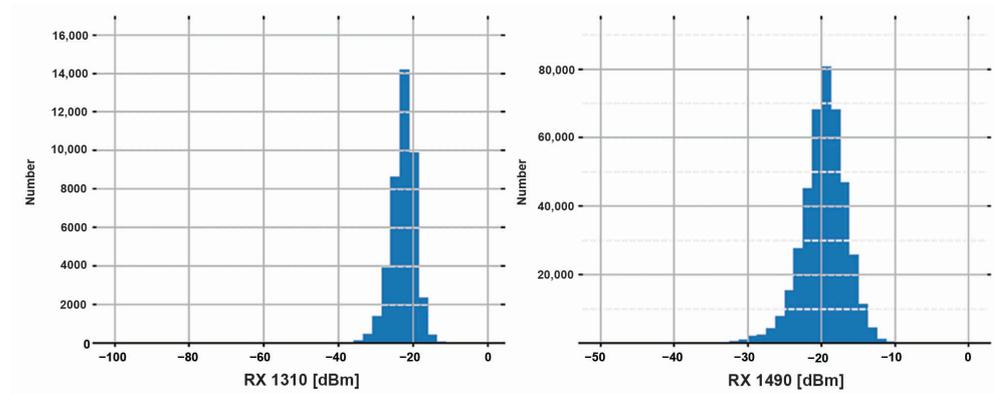


Figure 3. Distribution of GPON signal levels.

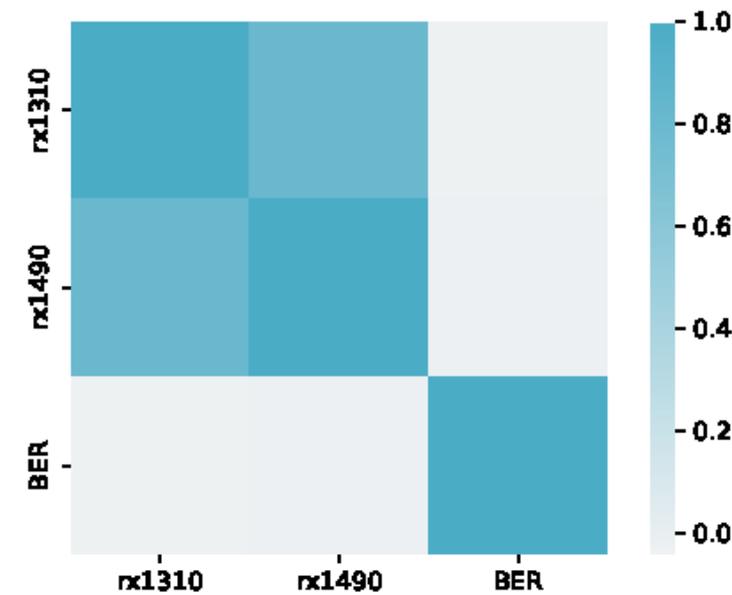


Figure 4. Correlation between GPON signal levels and BER.

A simple analysis of the signal parameters can result in many trips of technicians to address too-low parameter values. However, a more complex multi-layer analysis shows that there is no simple correlation of signal parameters with BER. So, this is an example that demonstrates the benefits of a multi-layer analysis, which allows the operator to avoid unnecessary trips and in this way improves the operational efficiency.

3.2. Internet Service Layer

The Internet service layer is the most complex because many services can be offered there, including over-the-top (OTT) services. Therefore, the simple and most cited measure of bandwidth, commonly referred to as the Internet speed, is not a good representation of QoE levels. Latency or jitter may seem much more important, as their high values can be much more annoying to users. In fact, the Internet service itself consists of streams of data coming from many geographically and logically dispersed parts of the Internet. Therefore, in our multi-layered QoE assessment, we not only used throughput, latency and jitter

measurements, but also probed end-user observable KPIs, such as those shown in Table 1. These parameters also reflect the complex nature of the telecommunications services as defined by ETSI [27]. Thus, polling also performs a multi-level test scenario, where the overall test score is calculated based on the weighted scores from all the test types.

Table 1. Service Probing.

Test Type	Bad Limit	Good Limit	Weight	Group
Video Test Success Ratio	80%	100%	58%	Video
Average Video-MOS	3	4.5	16.50%	
10% Percentile Video-MOS	2	4	16.50%	
Average Time to 1st Picture	7 s	2 s	4.50%	
Time to Fires Picture > 10 s Ratio	5%	0%	4.50%	
Website download test success ratio	80%	100%	66.67%	WWW
Average Website transfer Time	15 s	3 s	28.57%	
Website transfer Time > 15 s	5%	0%	4.76%	
Social Media Success Ratio	80%	100%	66.67%	Social media
Average Social Media Transfer	15 s	3 s	28.57%	
Social Media Transfer > 15 s	5%	0%	4.76%	
Transfer Success Ratio DL	80%	100%	11%	Service
Download file avg speed	1 Mbit/s	100 Mbit/s	14%	
Download file speed 10% percentile	1 Mbit/s	40 Mbit/s	18%	
Download file speed 90% percentile	10 Mbit/s	240 Mbit/s	7%	
Transfer Success Ratio UL	80%	100%	11%	
Upload file avg speed	0.5 Mbit/s	50 Mbit/s	14%	
Upload file speed 10% percentile	0.5 Mbit/s	30 Mbit/s	18%	
Upload file speed 90% percentile	5 Mbit/s	100 Mbit/s	7%	

The approach we took here was based on predefined weights set for all the parameters involved. These weights are currently set as fixed values based on the knowledge of network administrators, so these calculations were more like an expert system. However, in future work, we plan to use the weights obtained from machine learning models, while these weights will be trained offline. Offline training typically provides greater accuracy as compared to online training and, additionally, offline training has the advantage of using multiple data sources over a sufficiently long period of time in order to produce an accurate model.

However, since the network environment is constantly changing, following the changes in the network topology, ever-increasing network bandwidth (which operators have to deal with every year), software and firmware updates, etc., weights need to be regularly trained to update the model to the current environment and to improve its accuracy.

3.3. Tv Service Layer

A typical Internet Protocol TeleVision (IPTV) is based on multicast transmission [28,29]. Its main advantages are the reduction of bandwidth consumption in IP networks and the reduction of the number of servers on the head end side for streaming content. In IPTV, all that is needed to receive video content is a CAM (Conditional Access Module) module to decrypt the stream. If a TV operator wants to provide a recording service, then it usually supplies an additional device, i.e., a set-top box with a disk.

However, multicast technology can be an impediment and brought some limitations when operators wanted to implement additional features such as catchup, timeshift, start-over, and cloud recording. The content provided by these features is usually recorded on a storage array, and when the viewer chooses what they wants to watch, then they receive an individual video stream, which very often will not overlap with the content streamed to other viewers.

The answer to these limitations seems to be adaptive streaming technology based on HTTP Unicast transmission [30]. It is more complicated in terms of providing TV services, while it allows for the mentioned additional functions to be added. Set-top boxes (STBs) that handle such transmission could have their size reduced, even to the size of a USB stick. If needed, the connection between STB and ONT could be wireless, which previously was a significant limitation.

Video services have a perceptual nature. Therefore, the QoE metrics that can reflect the human perception of video can be more reliable than those related to QoS. Adaptive streaming is widely used in television services to provide linear content and additional features, including video-on-demand, startover, timeshift, catchup, or cloud recording. Combined with Internet services, over-the-top (OTT) STBs can serve the interactive program guide or movie recommendations, as well as external OTT services such as Disney+, HBO Max, Netflix, and others. Modern television CPEs can act as a hybrid entertainment center. From the network operator's perspective, it is thus necessary to provide the highest possible QoE level. In adaptive streaming, the most harmful events related to the drop of QoE are rebufferings and quality switches on account of their unexpectable nature, while the initial delay is considered less destructive [31]. Therefore, the stall ratio, stall frequency, and the number of bitrate changes, along with their frequency and bitrate decrease concerning the highest possible value, can be crucial KPIs, regardless of the protocol used.

Typically, video content is encoded at various bitstream speeds. The stream bandwidth can be affected by the codec used, resolution, frame rate and other parameters related to the video. Commonly, streams with lower bitrates are established for mobile devices, since they have a smaller screen size and may struggle with less efficient network access. On the other hand, streams with the highest bitrates are targeted at larger screens, such as those of TV sets. If the network is not efficient enough, it is also possible to play intermediate streams that provide video content of average quality. Continuous stream adaptation allows for video stream playback in degraded network environments. The linear TV service tends to target consumers who play their content on TV sets. In that case, set-top boxes are expected to strive to display the highest quality available. In pure OTT applications, it is usually possible to change the bitrate manually, while in a TV service such a feature seems not to be relevant. From the perspective of a TV operator, any playback of a stream with the quality other than the maximum should be treated as video degradation within a given set-top box. This is particularly true when TV services are provided along with Internet access by the same operator, because it is directly associated with the operator's managed network degradation.

Figure 5 demonstrates the percentage of intermediate stream playbacks compared to playbacks of the highest quality streams over 30 days of observation in the Fiberhost TV system. We found that for a given linear TV, the playback of non-maximum streams was more or less constant, constituting about 35% of all playbacks. A significant change occurs when the content delivery network (CDN) servers are close to saturation or there is an increased level of degradation, e.g., due to a system failure or CPU overload of the server streaming Unicast content. When the decoder begins playback, its goal is to play the highest quality stream, but it can accomplish this only when the video buffer reaches a certain level. In the meantime, the decoder will start playback from the lowest video stream, which in this case is 800 kbps. Optimally, the streams with bitrates other than the maximum and other than 800 kbps would not appear in the described chart. In that case, customers could experience stalls that would impact QoE much more strongly than while playing the quality slightly worse than the maximum. Increasing the percentage of the intermediate streams significantly affects the QoE level of individual customers, and is also an indication of the health of the TV service as a whole. On the operator's side, improvements could be made in the area of streaming servers and the access network used for content distribution. On the customer/CPE site, improvements could be made in the area of the Wi-Fi connectivity to ensure the highest possible signal quality and strength.

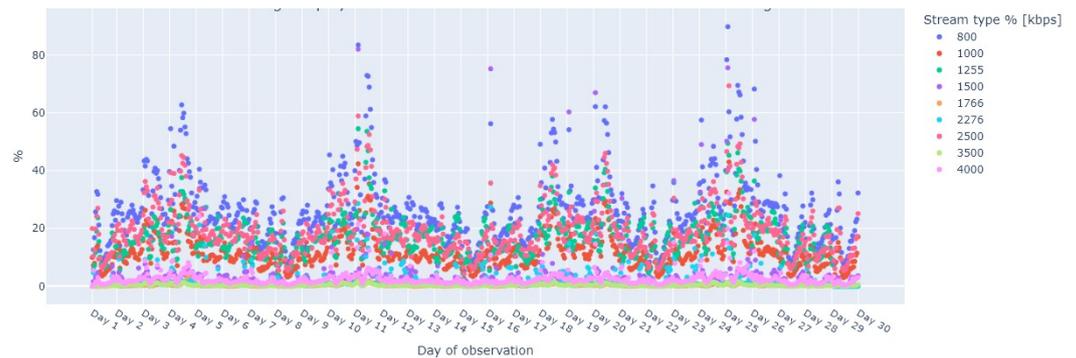


Figure 5. Percentage of played intermediate streams.

Rebufferings are related to video player buffer saturation. During a stall event, the black screen or loading animation is displayed on the video player. When the playback begins, the CPE represented by the OTT decoder burst downloads video chunks to feed the buffer. After the specified threshold is reached, the STB plays the video content. When a network impairment occurs, the STB does not download new video chunks, so the buffer gets empty, and as a result the stall occurs. It is these occurrences of stalls that were investigated during the 30 days of observation period in Fiberhost TV system.

The highest number of errors occurred when two CDN servers malfunctioned and the decoders needed a certain amount of time to download more video segments from other available streaming servers. Table 2 indicates the number of stalls from the hour in which the described event occurred, along with the number of all decoders connected to each aggregation router. The names of the routers have been anonymized at the operator's request. Across the 41 aggregation devices distributed throughout the country, the results show that the percentage number of stalls with respect to all connected devices in a given location oscillated between 20 and 33%, which is a similar value for all aggregation routers. The errors appeared in the entire TV infrastructure, which means that the set-top boxes were not oriented to specific streaming servers as regards the location. With this approach, it can be expected that customers throughout the area will report an outage through a support line. However, from the QoE and CX (Customer eXperience—as described in Section 4) assessment perspective, the percentage of decoders experiencing a problem may not be a suitable metric due to the fact that for the location covered by the Aggregation-12 device, the possible number of customers who would be willing to report a problem is 1357, whereas for Aggregation-17, the number is only 65. The corresponding figures for both locations are 28.88% and 28.26% of devices with problems. Nevertheless, in the former case, the operator must expect that the queue of customers who would wish to report a problem to the inbound hotline will be much larger than in the latter case. This is significant information for the AX parameter (Analog eXperience—as described in Section 4), which would affect the level of QoE associated with the helpline calls.

Table 2. Stalls per aggregation router.

Aggregation Router	STBs with Stalls	All Devices	%
Aggregation-1	26	80	32.50
Aggregation-2	53	170	31.18
Aggregation-3	317	1018	31.14
Aggregation-4	413	1337	30.89
Aggregation-5	158	519	30.44
Aggregation-6	216	711	30.38
Aggregation-7	635	2124	29.90
Aggregation-8	629	2132	29.50
Aggregation-9	399	1363	29.27
Aggregation-10	221	762	29.00
Aggregation-11	756	2613	28.93
Aggregation-12	1357	4699	28.88
Aggregation-13	276	957	28.84
Aggregation-14	150	525	28.57
Aggregation-15	165	580	28.45
Aggregation-16	260	915	28.42
Aggregation-17	65	230	28.26
Aggregation-18	132	472	27.97
Aggregation-19	308	1112	27.70
Aggregation-20	543	1973	27.52
Aggregation-21	321	1167	27.51
Aggregation-22	603	2196	27.46
Aggregation-23	117	428	27.34
Aggregation-24	263	966	27.23
Aggregation-25	177	651	27.19
Aggregation-26	112	412	27.18
Aggregation-27	110	405	27.16
Aggregation-28	136	502	27.09
Aggregation-29	190	715	26.57
Aggregation-30	279	1054	26.47
Aggregation-31	150	567	26.46
Aggregation-32	408	1545	26.41
Aggregation-33	274	1039	26.37
Aggregation-34	194	742	26.15
Aggregation-35	318	1226	25.94
Aggregation-36	188	737	25.51
Aggregation-37	121	478	25.31
Aggregation-38	7	28	25.00
Aggregation-39	61	250	24.40
Aggregation-40	131	555	23.60
Aggregation-41	130	629	20.67

Figure 6 presents the geographic distribution of stalls in a percentage manner. Figure 6a presents the moment when two CDN servers experienced an outage. Figure 6b shows the total number of stalls from another day between 6 and 7 pm, which is considered the beginning of prime time. In this case, the number of errors was small and the stalls occurred in isolated instances, mainly due to wireless network problems at the STB connection site. Because of the percentage distribution, it is difficult in this case, from the operator's perspective, to plan the resources related to the incoming line for the area or the number of technicians who could help customers solve their problems. These are the primary factors that can affect AX.

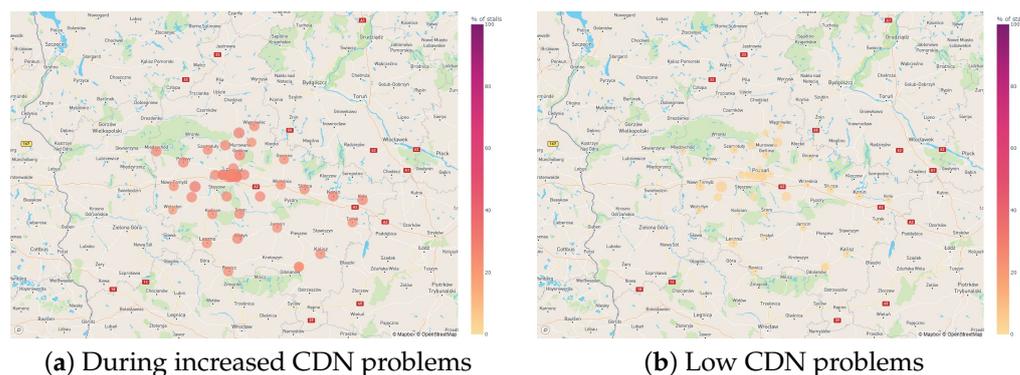


Figure 6. Percentage of stalls.

The initial delay is proportional to the time required to fill the player buffer above the predefined threshold to start playback. The buffer must be filled after every channel change or usage of additional features such as timeshift, catchup, startover, and others, and finally, after the STB start-up. TV services do not inject ads at the beginning of playback of linear content or VOD content that is part of the service. Because of this, and because it is less important than stalls and changes in the quality of streams, the initial delay will not be considered in this article.

Stalls and bitrate changes of video streams are the most important factors affecting the perception of video quality. User engagement during declining service quality is also significant, and can be expressed by phone calls to a helpline. Given the number of variables that can affect changes in the bitrate, it seems impossible that particular users will not experience bitrate drops in a linear TV system. It is important to catch sufficient data at the local level with respect to individual user and at the global level, i.e., in a given area. Stalls or bitrate changes, which relate to video quality, should be correlated with call center notifications, i.e., the user engagement parameters. Such correlation should occur in real time. Using regression, classification or reinforcement learning methods, it would be possible to generalize the QoE level to the entire user population, due to the fact that it is difficult to obtain an engagement rating from each customer of a TV service.

There are a number of works that address the problems of the QoE assessment. A good reference can be the ITU-T recommendation P.1203 [32], which evaluates QoE in adaptive streaming based on stalls, bitrate of audio and video streams, and data associated with the codec used. Other algorithms that analyze quality or QoE include Video ATLAS [33], KSQI [34], or VMAF [35]. There are also algorithms that provide infrastructure improvement or adaptations based on the obtained QoE score, including SDNDASH [36], QDASH [37] or described by Jingteng [38]. In our research, we aimed to acquire user engagement and assess video quality per user account on the system but also to assess the overall system delivering TV content per location. Our work is based on a continuous dataset that comes from linear TV content, rather than the finite number of video files.

3.4. Telephony Service Layer

Telecommunications operators need to constantly monitor the voice quality to detect things such as cracklings, distortion, echo, dropped calls, and unidirectional sound. For this purpose, Mean Opinion Score (MOS) is a parameter commonly used by telecommunications operators as it represents the subjective audio quality parameter calculated after compression, decompression or transmission. As a consequence, we used it in our multi-tier QoE learning system to evaluate the VoIP quality. Since we used the G.711 codec [39], the expected MOS value for the codec was around 4.3. However, since Fiberhost uses optical fiber as the transmission medium, the experienced delays are usually very low. Therefore, most of the recorded MOS values were around 4.5, as shown in Figure 7.

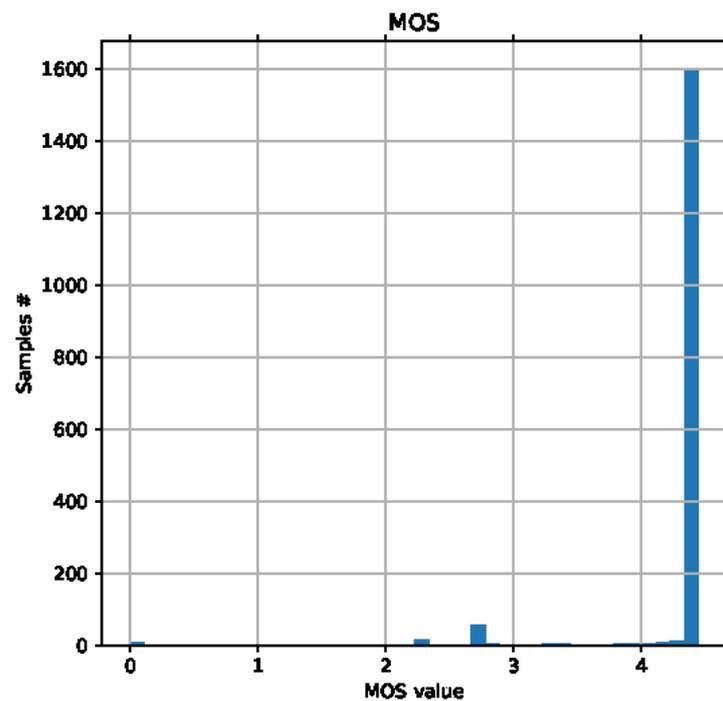


Figure 7. MOS Distribution.

4. Compound QoE Metric Used in Multi-Layer QoE assessment

In this section, we present Fiberhost's efforts to implement a system that monitors not only the QoE related to offered services, but also all interactions with the services and the personnel of the operator. This idea is already described in great detail in [23]. However, for reference, we also provide brief information here.

The QoE metric used in Fiberhost is the end product of the multi-level QoE learning system described in the previous sections. We named the metric CX (Customer eXperience) and defined it as the sum of Digital eXperience (DX) and Analog eXperience (AX). DX is a component calculated from metrics collected from the network, so it is similar in nature to compound QoE metrics such as NPS [40]. Analog eXperience (AX) is designed to capture all human-operator interactions, not only with technical staff, but also with sales, finances, etc. An example of the distribution of CX values in one of the operational geographical areas is shown in the Figure 8. The current CX implementation uses predefined weights set for all parameters, as in the approach described in Section 3.2. However, in future work, we plan to use weights obtained from machine learning models, and these weights will be trained offline, for better accuracy and adaptability to the short and long-lasting changes occurring in the network.

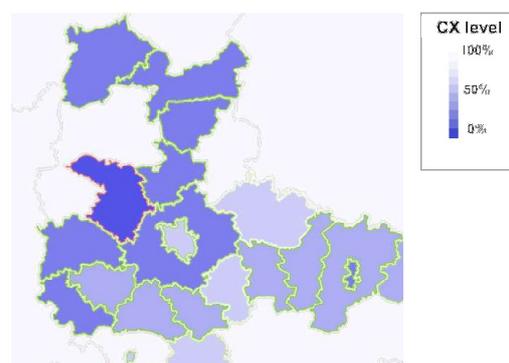


Figure 8. Example CX aggregation as calculated for a specific geographical area.

The implementation of CX creates the foundation that can then be used by machine learning (ML) methods to identify more complex network dependencies and optimize service performance. However, ML-based optimization is planned for future work, while currently the practical use of CX is to improve the level of CX in all underperforming areas.

5. Implementation of Unsupervised Learning

In TV services, the CX is reflected by a multi-path approach, shown in Figure 9. The core of the AX component is the incoming call center line and interaction between customer service team and the client that usually makes a complaint or has a number of questions with regard to the service. The DX part is implemented through two parallel paths. One is based on subjective opinions about the evaluation of the TV service obtained through a questionnaire directed to a customer calling the inbound hotline, or a single-page application that runs on the STB. The customer is asked six questions contained in Table 3 rated on a scale of 1 to 5, where 5 is the most positive feeling and 1 is the most negative feeling.

Table 3. QoE survey.

Question
How do you rate the quality of video today?
How do you perceive the level of image quality on the service?
How much do image stalls affect your satisfaction with the service?
How much do video quality changes affect your satisfaction?
How much does the delay in starting playback affect your satisfaction?
How would you rate the overall performance of the service?

In the second DX track, a subjective QoE rating is calculated based on the data collected from the STBs. The logs cover the level of video streams played, time spent on the selected stream, frequency of video quality changes, and the metrics related to stalls, i.e., the frequency and their duration per decoder. The analysis also includes additional errors specific to the overall system implementation which could affect user perception. It includes data related to API errors or system errors that generate messages displayed during playback. After appropriate filtering and analysis of the features, a QoE evaluation is performed on the basis of unsupervised learning models based on continuous clustering, as described in [24]. The SCI, SCTI, STCSI and VSBCT parameters are calculated and mapped to scores from 1 to 5, where 5 is the highest level of QoE. The QoE value is estimated due to the lack of actual evaluation obtained from the customer at this stage. The QoE per set-top box estimation should be aggregated into QoE per account ratings due to the fact that one customer may have multiple STB devices as part of a multiroom service. Aggregation will also allow the estimated rating to be mapped to the rating obtained from the survey in the parallel DX path. Combining the actual rating from the customer and the estimated one will allow the model to be improved for estimation to better represent the QoE within that particular service. Unfortunately, there is no way to collect the survey-based data from every customer, and for this reason either an unsupervised learning model can be better suited or a supervised learning model can be created after collecting as many responses as possible.

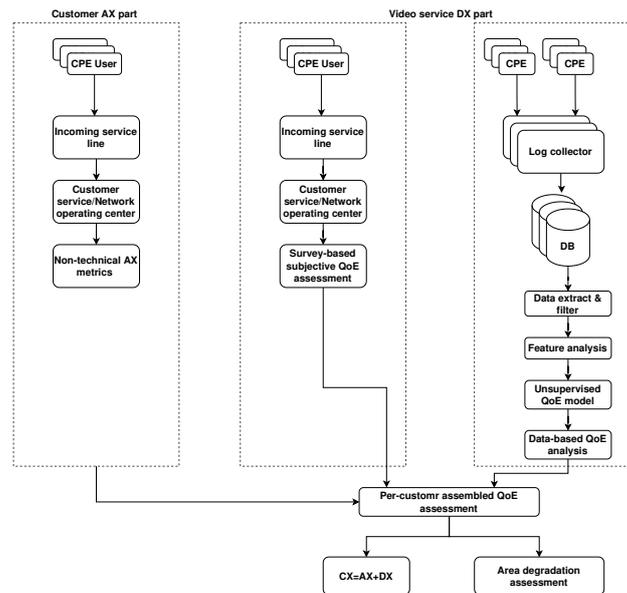


Figure 9. Multi-path QoE assessment.

5.1. Outline of Unsupervised Part of System

In a distributed TV system, the main problem in video assessment is the lack of labels showing whether the evaluation is intended to use supervised machine learning techniques. Specifically, the evaluation will be relying on data collected from set-top boxes through telemetry protocols such as MQTT. The type of data could provide the details associated with the video content that is currently being played. In the TV system environment, it relates to the TV broadcast channel, video-on-demand content, or other services available from the decoder. We decided that it was possible to collect metrics related to the level of the video stream being played, stalls, channel switches, and usage of decoder features that include npvr, catchup, startover, as well as timestamps. We linked them to one another and grouped them accordingly to have an overview of the level of video quality played on the set-top boxes. Figure 10 shows an outline of the system designed to collect the necessary metrics from the decoders located in customers' homes. The part related to unsupervised learning, which was used in the first stage to make an overall assessment of the content being played, is highlighted in red. The MQTT protocol was used by the decoders to send data to the database cluster. A compute server used queuing mechanisms to retrieve data related to the playback of video streams, stalls, time stamps, the use of decoder features and other respective parameters from the worker nodes at regular intervals. The data were cleaned. Features were extracted and combined into new metrics, as described in Section 5.2. Once the target features were obtained, they were clustered in a continuous mode. For this, a two-stage algorithm based on DStream [41] and a hierarchical algorithm were used. In the first phase, a synopsis of the stream was computed. This synopsis was updated whenever a new input occurs. The second phase was to agglomerate the collected stream summary and obtain the target associations. DStream and agglomeration algorithms were chosen due to their speed and flexibility [24]. To reflect the Absolute Category Rating (ACR) and Degradation Category Rating (DCR) scales shown in Table 4, and described in ITU-T Recommendation P.913 [42]; the number of groups was set to five. A score of 5 indicates the best possible perceived video quality on the ACR scale. It corresponds to an imperceptible level of degradation on the DCR scale. A score of 1 on the ACR is the lowest possible video quality and corresponds to the maximum degradation on the DCR scale.

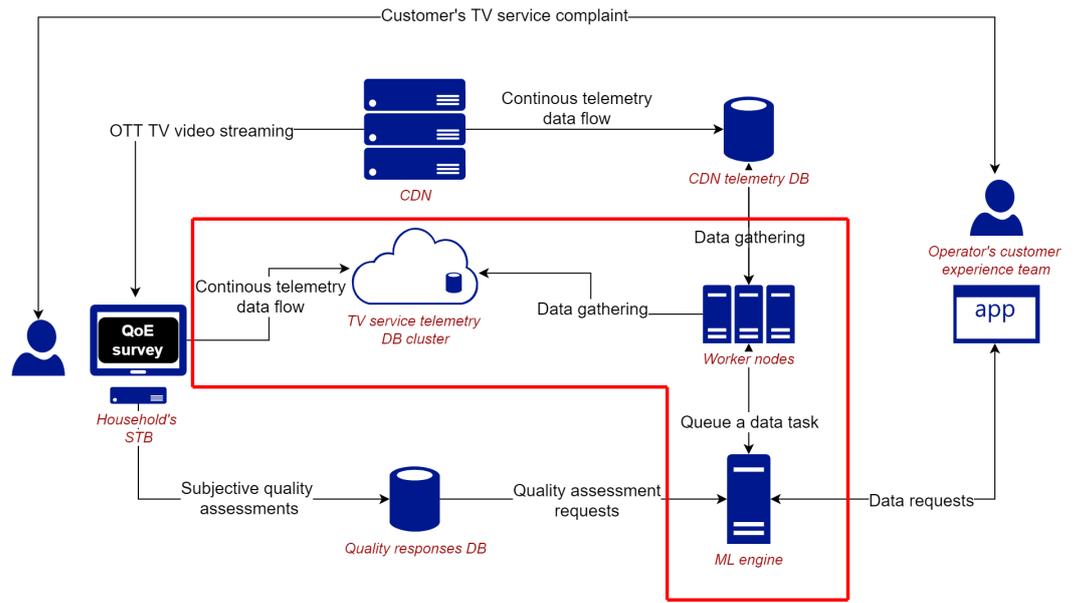


Figure 10. Outline of QoE assessment system for OTT TV service. The red shape indicates the unsupervised part.

Table 4. ITU-T 5 point ACR and DCR scale.

Grade	Estimated Quality (ACR)	Perceived Impairment (DCR)
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

5.2. Clustering

Unsupervised techniques were used due to the lack of reliable labels related to the assessment of perceived video quality. For stream clustering, we used the input features described in [24], and computed on the basis of the characteristics of adaptive video content streamed over the HTTP protocol. This is the technology used by the TV system to deliver video content to customers. Specific TV channels are delivered to clients at different quality levels. CDN servers can typically stream three or four profiles per channel. In general, the higher the bitrate, the better the quality and resolution of the video content. Standard channels can range from 800 kbps to 10 Mbps per profile. Beyond these values, there are usually 4 K (3840 × 2160) resolution channels. The *SCI* (1) parameter takes on values between the range $< -1; 1 >$ and determines the degree to which the video profile changes during playback. If the player switches to a higher profile, i.e., one with better quality from the viewer’s point of view, the parameter takes on positive values. In the opposite case, it will take on negative values. If the player starts playing the highest profile available for a given channel, then $SCI = 1$. If the player is unable to play any stream, e.g., due to network degradation at the customer’s home, then $SCI = -1$.

$$SCI = \begin{cases} -1 & \text{bitrate}_{kbps}^N = 0 \\ 1 & \text{bitrate}_{kbps}^N = ref_{bitrate} \\ \frac{\text{bitrate}_{kbps}^N - \text{bitrate}_{kbps}^{N-1}}{ref_{bitrate}} & \text{otherwise.} \end{cases} \quad (1)$$

SCTI (2) is a correlated coefficient with *SCI* to determine the level of distortion on a given channel over time. This coefficient takes into account the importance of a given channel viewership PI_{CH}^{TI} (3) by comparing the ratio of active decoders on the selected channel

CH to all active set-top boxes at a given time. Viewership is calculated at time intervals fixed at $TI = 15$ minutes but can be adjusted accordingly. SCTI takes into account the playback time of a given stream. This parameter will be negative when playing a lower bitrate profile. The parameter $edt_n = t_n - t_{n-1}$ indicates the playback time of the stream. If a stream with a lower quality than the maximum is played for a longer time, the SCTI parameter will be further amplified to reach lower values. The reason for the inclusion of channel viewership is that some channels may be more important than others from a provider perspective. When major events are broadcast on television, the number of viewers tends to increase significantly. A degradation or even failure during such an event could lead to a significant reduction in the viewer’s perceived quality of the overall service.

$$SCTI = \log_2(1 + edt_n) \cdot SCI \cdot \left(1 + \text{channel viewership}_{CH}^{TI}\right) \tag{2}$$

$$\text{channel viewership}_{CH}^{TI} = \frac{\text{active devices}_{CH}^{TI}}{\text{all active devices}_{CH}^{TI}}. \tag{3}$$

VSBC (4) is a memory factor that takes into account the frequency of profile changes and stalls. A large number of events in a short period of time can have a greater effect on the perceived quality of the video than a small number of individual events that occur over the entire playback session. Any negative change in quality will cause this ratio to deteriorate. Conversely, a passage of time without negative events causes the coefficient to increase. A large number of events in a short period of time will cause this indicator to decrease significantly. When the decoder fully recovers, the values of the characteristics start to improve and, after a certain time, reach the maximum value until the next quality deterioration. VSBC takes into account the negative event counter bcc (5), which increases when the decoder switches to a lower profile than the one that is currently being played. The $tsbc$ (6) factor is the time elapsed since the last negative event.

$$VSBC = \begin{cases} 1 & b_c = 0 \\ -1 \cdot \frac{\log_2(1+b_c)}{tsbc} & b_c > 0 \end{cases} \tag{4}$$

$$b_c = \begin{cases} b_c & \text{bitrate}_{kbps}^N > \text{bitrate}_{kbps}^{N-1} \\ b_c + 1 & \left(\text{bitrate}_{kbps}^N \neq \text{ref}_{bitrate}\right) \leq \text{bitrate}_{kbps}^{N-1} \\ b_c & \text{bitrate}_{kbps}^N = \text{ref}_{bitrate} \end{cases} \tag{5}$$

$$tsbc = \begin{cases} t_{n-x} & \text{bitrate}_{kbps}^N > \text{bitrate}_{kbps}^{N-1} \\ t_n & \text{bitrate}_{kbps}^N \leq \text{bitrate}_{kbps}^{N-1}. \end{cases} \tag{6}$$

Stalls are the events related to the saturation of the playback buffer. At this point, the decoder stops playing the content. Stalls can be caused by network conditions, problems with CDN servers or internal problems of the TV platform. The bitrate value for the stream is then 0 Mbps. We decided to use characteristics to determine the number of stalls for a single event, channel and session. We considered the ratio of the sum of the stall duration to the duration of the event, channel and session. However, we decided that the stall counter would better represent the behavior of the decoder when a stall occurs. In our case, the TV service was only provided on a GPON network, which is characterized by fast data transmission and low latency. The decoders are delivered to the customers together with optical terminals. The connection to these terminals can be wired or wireless. When stalling occurs, the decoder displays a black screen with an error message. We have observed that, in most cases, this behavior lasts between 1 and 3 s, as the performance of the optical network allows the decoders to download successive video segments quickly. In this particular case, it will be more annoying for the customer to have interchangeable stalls with a black board that appears alternating with the video content than to have a single

black screen that lasts for a longer time. In addition, we needed to take into consideration that events on TV services tend to last much longer than short videos on popular streaming services. Sports events, for example, can last for more than 3 h as a single show. The number of stalls was determined for a single event $avErr_{event}$ (7), a channel $avErr_{channel}$ (8), and the entire viewing session, i.e., from start to shutdown of the decoder $avErr_{session}$ (9).

$$avErr_{event} = \sum_{n=1}^{i_{event}} n \tag{7}$$

$$avErr_{channel} = \sum_{n=1}^{i_{channel}} n \tag{8}$$

$$avErr_{session} = \sum_{n=1}^{i_{session}} n. \tag{9}$$

A min-max normalization (10) was used to scale the different metrics. The value of the metric x was returned as x_{scaled} . This value was scaled in the range $\langle a; b \rangle$. Since the parameters SCI , $SCTI$ and $VSBCT$ can refer to both positive behavior, i.e., quality improvement, and negative behavior, i.e., quality degradation, they took values within the range $\langle -1; 1 \rangle$. The parameters $avErr_{event}$, $avErr_{channel}$ and $avErr_{session}$ were scaled to values $\langle 0; -1 \rangle$. This is because they were only associated with the negative behavior of stalling.

$$x_{scaled} = a + \frac{(x - \min(x))(b - a)}{\max(x) - \min(x)}. \tag{10}$$

5.3. Results

The continuous nature of television requires constant evaluation of video quality. The viewer is typically watching linear channels on which video events are presented according to the TV schedule. We divided the quality assessment into levels that included session, channel and individual events on a time scale in order to obtain as much information relevant to video playback as possible. Figure 11 presents the adopted approach. A session is the time from startup to shutdown of the decoder. It can include ratings during playback of different channels, recordings, catchup and startover content. The rating at the channel level is related to the viewing of a single linear channel. At the lowest level is the rating for a single programme, i.e., the events that are displayed in the EPG grid or in the TV programmes that are available on various websites. The rating starts 30 s after the start of playback to avoid channel-hopping, where the customer is looking for a channel that they prefer to watch.

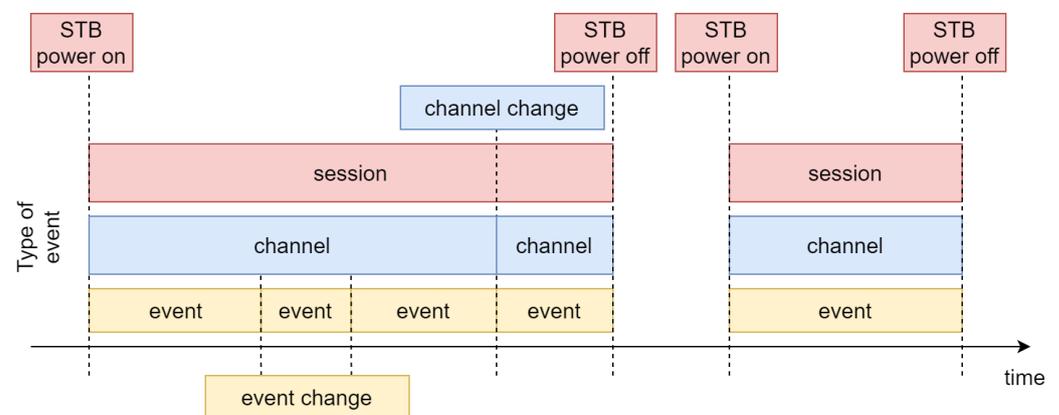


Figure 11. Timeline for session assessment. The customer can watch many events on a single TV channel. The session begins when the STB power is on, and lasts until shutdown.

The ratings per channel for two selected decoders are shown in Figures 12 and 13. Figures 14 and 15 show the rating for the whole of the viewing session for the same set of decoders. The ratings are for 24 h. The vertical lines are drawn in each graph to indicate the channel change (Figures 12 and 13) and the start of the viewing session associated with the start of the set (Figures 14 and 15). The decoder shown in Figures 12 and 14 is the reference device, located in the data center. In general, it should always have the best rating but, in this case, especially around the peak traffic hours, i.e., between 7 p.m. and midnight, ratings lower than 5 appeared. This was associated with the degradation of one of the edge servers, and the increased number of degradations appeared on the other decoders. This can also be seen in the numbers associated with the other device. A detection of such degradations is important because such streaming server failures may not be visible on typical monitoring systems, as they may be related to server software issues.

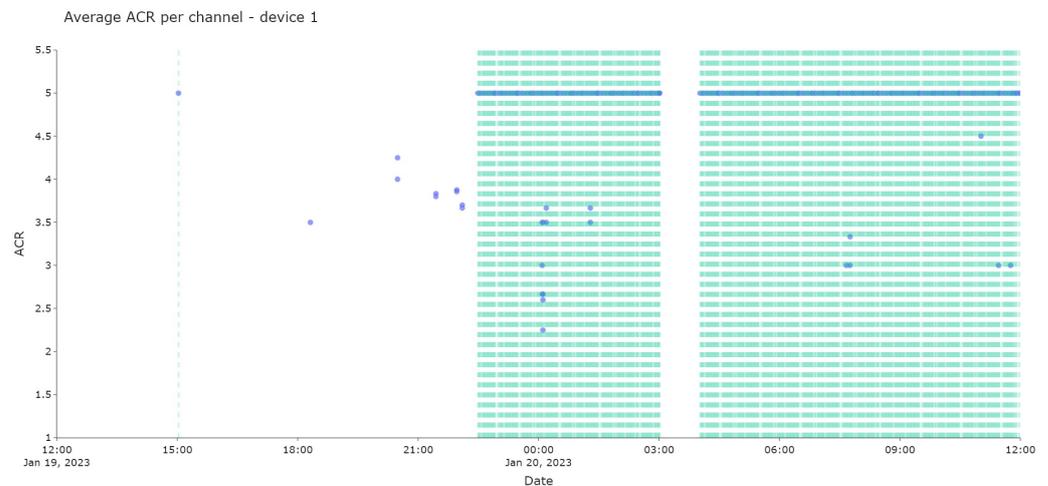


Figure 12. Average ACR grade for given channel for first device.

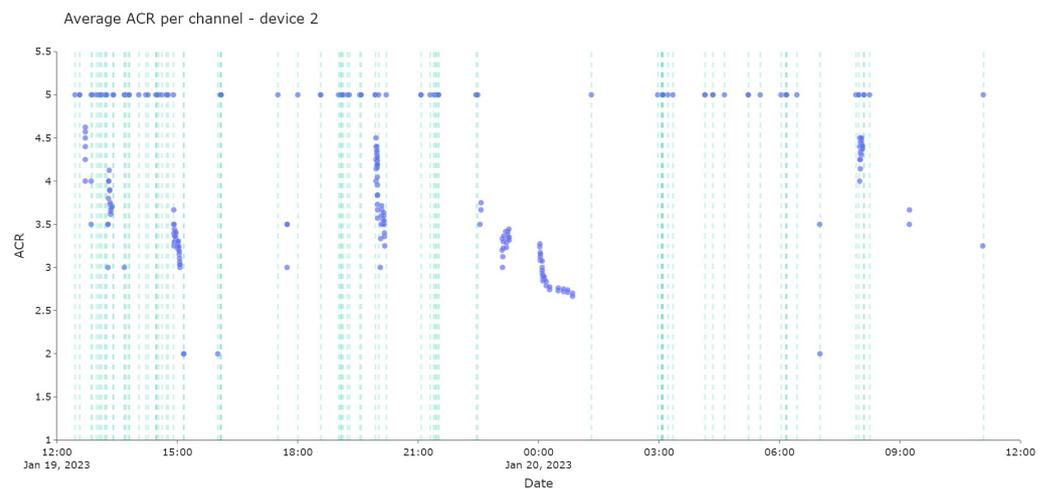


Figure 13. Average ACR grade for given channel for second device.

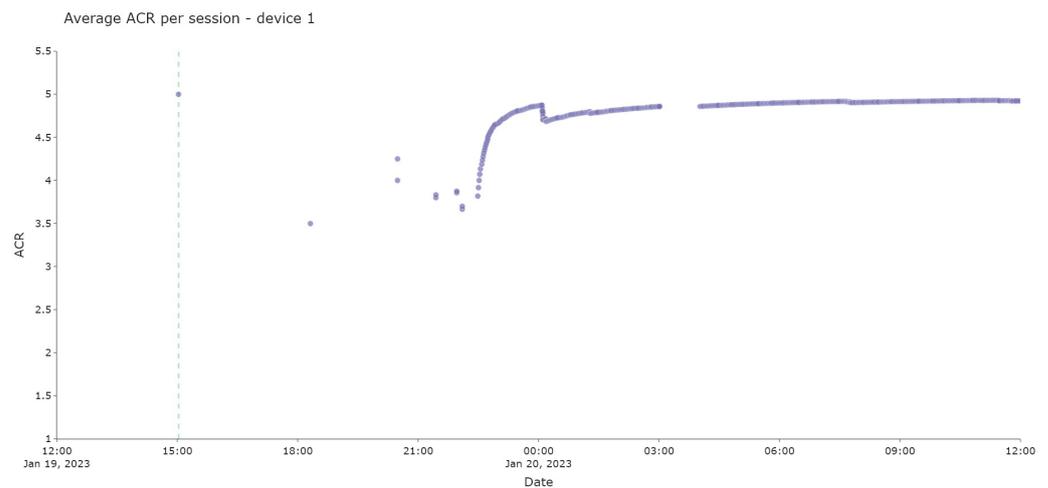


Figure 14. Average ACR grade for entire viewing session for first device.

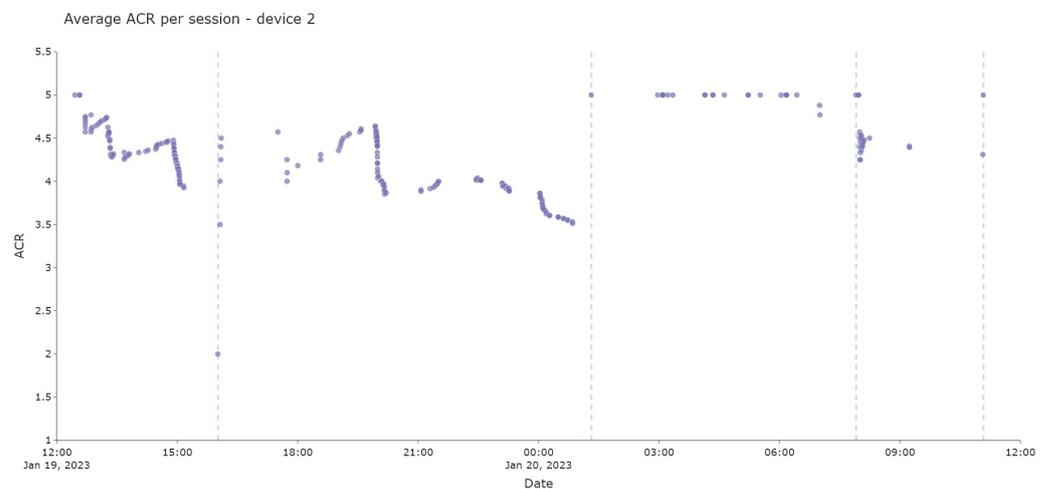


Figure 15. Average ACR grade for entire viewing session for second device.

Figures 13 and 15 show the evaluation for the regular decoders in the households. The time periods between 12:00 and 15:00, 18:00 and 01:00 and after 07:00 are particularly characteristic. It can be seen that there is an increase in the degradation of the various decoders. Between the hours of 12:00 and 15:00, errors appeared on the client devices, whereas this was not observed on the reference device. After 07:00 in the morning, individual errors appeared on the reference unit, but this was much more clearly visible on the client devices. These errors may have been related to maintenance work carried out by an engineer on the CDN server infrastructure. The most problematic period is between 18:00 and 01:00. In this case, significant errors appeared on all devices, and were related to problems in the operator's infrastructure. Identifying this type of degraded connection can help determine whether the failure was global, affecting the entire TV infrastructure, or local, reaching the subscribers' homes. When maintenance work is carried out, it is possible to assess the extent to which viewers are being affected. Searching for individual faulty devices allows for proactive quality improvement on a per-customer basis, which can significantly increase satisfaction with both the content watched by the customer and the service provided by the operator.

6. Plans for Implementation of Supervised Learning

Figure 10 shows a diagram of the Multi-layer QoE learning system in the TV service, which includes both the AX and DX parts. The AX part is implemented in two ways and involves the collection of labels linked to customer satisfaction with the service. The ratings

collected from customers are fed into a database of quality ratings, which is accessed by an engine that estimates ratings based on telemetry data. The decoders retrieve the content from CDN servers located behind the load balancer. The choice of a CDN server depends on the rules set on the load balancer and can be based on location, type of content being played, channel and many other operator defined criteria. CDN servers generate their own telemetry logs which are stored in a database for the CDN servers. Decoders located at customer premises record individual content playback actions. They can report on the type of content that is being played, the bit rate of the stream, stalls or the state of the wireless network, among others. A server running the machine learning algorithms queues individual tasks on worker nodes that have access to data collected from the decoders, CDN servers and rating labels. If necessary, the consultant can use the application to check the condition of the installation and service at the customer's home. Based on individual assessments, the system can conclude overall TV platform performance. A supervised machine learning approach would provide a more accurate quality rating than stream clustering. With DX labels that take into account individual feelings of customers, it would be possible to create an algorithm that evaluates not only the video, but also subjective feelings about the video service. The prerequisite for this is that the labels are of high quality. We planned to collect the labels through a short survey about the perceived quality. The data would arrive in the system in two ways. The first will be a simple single-page application with a questionnaire with six questions that will be displayed as a menu item on the decoders Figure 16. The user can answer the questions, which are then sent to the database via an API. A specific user group Friendly User Test (FUT) will be included in the survey.

Welcome in the video quality survey!
Thanks to your commitment we can continuously improve the quality of our TV service.

How do you rate the quality of video today?
1 2 3 4 5

How do you perceive the level of image quality of the service?
1 2 3 4 5

How much do the image stalls affect your satisfaction with the service?
1 2 3 4 5

How much do the video quality changes affect your satisfaction?
1 2 3 4 5

How much does the delay in starting playback affect your satisfaction?
1 2 3 4 5

How would you rate the overall performance of the service?
1 2 3 4 5

Send

Figure 16. Application displayed on STBs that contains survey for video quality assessment and label gathering.

When a customer calls the helpline and reports a problem related to the TV part, a DX survey is collected. The customer is asked a number of questions. These are listed in Table 3 and relate to the questions displayed on the one-page application on STBs. The expected response is a rating on an ACR scale within the range of 1 to 5, where 1 is the most negative feeling and 5 is the most positive feeling. In the first question, we want to find out how the customer feels about the TV service on the day of the survey. The evaluation should cover the entire time the customer watched the content on this particular day. In the second question, we want to discover whether the bitrates of the video streams, mainly the maximum ones, which account for more than 65% of the total traffic on the decoders, have been appropriately adjusted (Figure 5). By inquiring on the preferences of the channels being watched and collecting a sufficient number of ratings, it is possible to deduce what group of channels should have an increased bitrate or a more efficient video coding algorithm applied, which will directly affect the increased QoE level.

In questions three, four and five, we want to determine how customers are affected by the previously described problems specific to adaptive streaming. The questions related to stalls, video quality changes and initial delay will allow us to prioritize quality improvements in these areas. The overall evaluation of the system included in the last question allows us to assess the overall health and quality of the service from the customer's perspective.

The first five questions are more related to picture quality, while the fifth may also relate to impressions of the decoder's software or GUI, the operating speed of additional functions (timeshift, catchup, startover, playback of recordings), messages displayed in the decoder's GUI, or the quality and number of movies and VOD services available alongside linear content. Obtaining ratings from customers and correlating them with the parameters of the STB, the network environment and the logs related to activities on the set-top box should make it possible to create an algorithm for estimating QoE based on supervised learning, provided a sufficient number of ratings are collected.

Our system currently uses unsupervised learning methods. However, unsupervised learning is vulnerable to incorrect data, which may be produced as an error in measurements, wrong implementation, etc. Therefore, we plan to extend its learning capabilities to include the implementation of a supervised learning channel through QoE surveys, as shown in Table 3.

Even though the supervised learning will occur infrequently (as compared to the number of logs and events reported by the network), we are still of the opinion that it will provide a valuable input which will reflect the end-user perspective (expected quality of service, user experience and preferences).

In addition to the QoE surveys presented in Table 3, the operator can use traces of observed user behavior as an indicator of user dissatisfaction. The idea for this is presented in [43]. For example, if a customer performs many bandwidth tests, thus may indicate a problem with the service. However, we have observed that users usually perform a great number of speed tests when they have just purchased a service or are having problems with it. Therefore, it is crucial to distinguish these two completely different user behaviors. Additionally, repeated pressing of the reload button on the TV remote by the customer may mean that there is a problem with the service or the responsiveness of the service is not provided at the level expected by the user. However, when analyzing these events, it should be noted that not all users know which button on the TV remote is responsible for the reload function, and some users who do know may simply abuse this function. Therefore, we need to implement an appropriate layer that will filter and interpret user behavior. We consider these issues to be a very interesting part of our work related to QoE and the correct use of machine learning methods, and we intend to continue research in this area in the future.

7. Conclusions

In this article, we introduced a multi-layer QoE learning system implemented by Fiberhost to perform QoE assessment. The system uses a compound QoE Metric, which is called CX (Customer eXperience). The system allows us to compare the QoE level offered by various technologies, experienced at different geographic locations and influenced by human interactions with the operator's maintenance or sales teams. This allows a proactive fault management to be appropriately implemented.

Work on the system is still in progress, though the system already implements machine learning methods to identify anomalies in the network performance. For this purpose, unsupervised learning is mostly used. However, we also plan to extend the system's learning capabilities by implementing a supervised learning channel through QoE surveys. Another interesting area is the use of traces of user behavior as the input that would help us train our machine learning models, which we also consider for our future work.

Author Contributions: Conceptualization, K.K., P.A., B.P. and P.Z.; methodology, K.K., P.A., B.P. and P.Z.; software, P.A.; validation, K.K., P.A. and P.Z.; formal analysis, K.K. and P.Z.; investigation, K.K., P.A., B.P. and P.Z.; resources, P.A.; data curation, P.A.; writing—original draft preparation, K.K.; writing—review and editing, K.K., P.A. and P.Z.; visualization, K.K. and P.A.; supervision, K.K. and P.Z.; project administration, K.K. and P.Z.; funding acquisition, P.A. and P.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The authors thank the Polish Ministry of Education and Science for financial support (Applied Doctorate Program, No. DWD/4/24/2020). This research was funded in part by the Polish Ministry of Science and Higher Education (No. 0313/SBAD/1307).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset was provided by Fiberhost S.A. Data are unavailable due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dynatrace. Available online: <https://www.dynatrace.com/> (accessed on 31 October 2022).
2. Huawei: IMaster NCE Autonomous Network Management and Control System. Available online: <https://e.huawei.com/en/products/network-management-and-analysis-software> (accessed on 31 October 2022).
3. Nokia: Network Services Platform. Available online: <https://www.nokia.com/networks/products/network-services-platform/> (accessed on 31 October 2022).
4. Witbe. Available online: <https://www.witbe.net/technology/> (accessed on 31 October 2022).
5. New TestTree StreamProbe 2110 for 24/7 QoS & QoE Monitoring of IP-Based Media Production. Available online: <https://www.enensys.com/product/streamprobe-2110-launch/> (accessed on 31 October 2022).
6. Campanella, A. Intent Based Network Operations. In Proceedings of the 2019 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 3–7 March 2019; pp. 1–3.
7. Han, Y.; Li, J.; Hoang, D.; Yoo, J.H.; Hong, J.W.K. An intent-based network virtualization platform for SDN. In Proceedings of the 2016 12th International Conference on Network and Service Management (CNSM), Montreal, QC, Canada, 31 October–4 November 2016; pp. 353–358. [\[CrossRef\]](#)
8. Banjanin, M.K.; Stojčić, M.; Danilović, D.; Čurguz, Z.; Vasiljević, M.; Puzić, G. Classification and Prediction of Sustainable Quality of Experience of Telecommunication Service Users Using Machine Learning Models. *Sustainability* **2022**, *14*, 17053. [\[CrossRef\]](#)
9. Terra Vieira, S.; Lopes Rosa, R.; Zegarra Rodríguez, D.; Arjona Ramírez, M.; Saadi, M.; Wuttisittikulkij, L. Q-Meter: Quality Monitoring System for Telecommunication Services Based on Sentiment Analysis Using Deep Learning. *Sensors* **2021**, *21*, 1880. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Taha, M.; Canovas, A.; Lloret, J.; Ali, A. A QoE adaptive management system for high definition video streaming over wireless networks. *Telecommun. Syst.* **2021**, *77*, 1–19. [\[CrossRef\]](#)
11. Laiche, F.; Ben Letaifa, A.; Aguilu, T. QoE-aware traffic monitoring based on user behavior in video streaming services. *Concurr. Comput. Pract. Exp.* **2021**, e6678. [\[CrossRef\]](#)
12. Tao, X.; Duan, Y.; Xu, M.; Meng, Z.; Lu, J. Learning QoE of Mobile Video Transmission with Deep Neural Network: A Data-Driven Approach. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 1337–1348. [\[CrossRef\]](#)
13. Gadaleta, M.; Chiariotti, F.; Rossi, M.; Zanella, A. D-DASH: A Deep Q-Learning Framework for DASH Video Streaming. *IEEE Trans. Cogn. Commun. Netw.* **2017**, *3*, 703–718. [\[CrossRef\]](#)
14. Zhang, H.; Dong, L.; Gao, G.; Hu, H.; Wen, Y.; Guan, K. DeepQoE: A Multimodal Learning Framework for Video Quality of Experience (QoE) Prediction. *IEEE Trans. Multimed.* **2020**, *22*, 3210–3223. [\[CrossRef\]](#)
15. Vega, M.T.; Mocanu, D.C.; Famaey, J.; Stavrou, S.; Liotta, A. Deep Learning for Quality Assessment in Live Video Streaming. *IEEE Signal Process. Lett.* **2017**, *24*, 736–740. [\[CrossRef\]](#)
16. Kossi, K.; Coulombe, S.; Desrosiers, C.; Gagnon, G. No-Reference Video Quality Assessment Using Distortion Learning and Temporal Attention. *IEEE Access* **2022**, *10*, 41010–41022. [\[CrossRef\]](#)
17. Liu, X.; Song, W.; He, Q.; Mauro, M.D.; Liotta, A. Speeding Up Subjective Video Quality Assessment via Hybrid Active Learning. *IEEE Trans. Broadcast.* **2022**, 1–14. [\[CrossRef\]](#)
18. Motaung, W.; Ogudo, K.A.; Chabalala, C. Real-Time Monitoring of Video Quality in a DASH-based Digital Video Broadcasting using Deep Learning. In Proceedings of the 2022 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), Durban, South Africa, 4–5 August 2022; pp. 1–6. [\[CrossRef\]](#)
19. RFC 2544: *Benchmarking Methodology for Network Interconnect Devices*; RFC 2544, RFC Editor; IETF: Fremont, CA, USA, 1999. [\[CrossRef\]](#)

20. Y.1564: *Ethernet Service Activation Test Methodology*; Recommendation Y. 1564; Telecommunication Standardization Section of ITU: Geneva, Switzerland, 2016.
21. Fedor, M.; Schoffstall, M.L.; Davin, J.R.; Case, D.J.D. *RFC 1157: Simple Network Management Protocol (SNMP)*; RFC 1157, RFC Editor; IETF: Fremont, CA, USA, 1990. [[CrossRef](#)]
22. *TR-069 CPE WAN Management Protocol*; Technical Report Issue: 1 Amendment 6; The Broadband Forum: Fremont, CA, USA, 2018.
23. Kowalik, K.; Partyka, B.; Andruloniw, P.; Zwierzykowski, P. Telecom Operator's Approach to QoE. *J. Telecommun. Inf. Technol.* **2022**, *2*, 26–34. [[CrossRef](#)]
24. Andruloniw, P.; Kowalik, K.; Zwierzykowski, P. Unsupervised Learning Data-Driven Continuous QoE Assessment in Adaptive Streaming-Based Television System. *Appl. Sci.* **2022**, *12*, 8288. [[CrossRef](#)]
25. Zabbix. Available online: <https://www.zabbix.com/> (accessed on 31 October 2022).
26. *TR 103-559 V1.1: Speech and multimedia Transmission Quality (STQ)*; Best Practices for Robust Network QoS Benchmark Testing and Scoring; Recommendation TR 103-559 V1.1; European Telecommunications Standards Institute (ETSI): Valbonne, France, 2019.
27. *Speech and Multimedia Transmission Quality (STQ); QoS Aspects for Popular Services in Mobile Networks; Part 2: Definition of Quality of Service Parameters and Their Computation*; Technical Specification ETSI TS 102 250-2 V2.7.1; European Telecommunications Standards Institute (ETSI): Valbonne, France, 2019.
28. Okerman, E.; Vounckx, J. Fast Startup Multicast Streaming on Operator IPTV Networks using HESP. In Proceedings of the 2021 IEEE International Symposium on Multimedia (ISM), Naples, Italy, 29 November–1 December 2021; pp. 79–86. [[CrossRef](#)]
29. Doverspike, R.; Li, G.; Oikonomou, K.N.; Ramakrishnan, K.; Sinha, R.K.; Wang, D.; Chase, C. Designing a Reliable IPTV Network. *IEEE Internet Comput.* **2009**, *13*, 15–22. [[CrossRef](#)]
30. Stockhammer, T. Dynamic Adaptive Streaming over HTTP –: Standards and Design Principles. In Proceedings of the Second Annual ACM Conference on Multimedia Systems, MMSys '11, San Jose, CA, USA, 23–25 February 2011; Association for Computing Machinery: New York, NY, USA, 2011; pp. 133–144. [[CrossRef](#)]
31. Seufert, M.; Egger, S.; Slanina, M.; Zinner, T.; Hoßfeld, T.; Tran-Gia, P. A Survey on Quality of Experience of HTTP Adaptive Streaming. *IEEE Commun. Surv. Tutorials* **2015**, *17*, 469–492. [[CrossRef](#)]
32. *P.1203: Parametric Bitstream-Based Quality Assessment of Progressive Download and Adaptive Audiovisual Streaming Services Over Reliable Transport*; Recommendation P.1203; Telecommunication Standardization Section of ITU: Geneva, Switzerland, 2017.
33. Bampis, C.G.; Bovik, A.C. Learning to Predict Streaming Video QoE: Distortions, Rebuffering and Memory. *arXiv* **2017**. [[CrossRef](#)]
34. Duanmu, Z.; Liu, W.; Chen, D.; Li, Z.; Wang, Z.; Wang, Y.; Gao, W. A Knowledge-Driven Quality-of-Experience Model for Adaptive Streaming Videos. *arXiv* **2019**. [[CrossRef](#)]
35. Li, Z.; Aaron, A.; Katsavounidis, I.; Moorthy, A.; Manohara, M. Toward A Practical Perceptual Video Quality Metric. 2016. Available online: <https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652> (accessed on 31 October 2022).
36. Bentaleb, A.; Begen, A.C.; Zimmermann, R. SDNDASH: Improving QoE of HTTP Adaptive Streaming Using Software Defined Networking. In Proceedings of the 24th ACM International Conference on Multimedia, MM '16, Amsterdam, The Netherlands, 15–19 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 1296–1305. [[CrossRef](#)]
37. Mok, R.K.P.; Luo, X.; Chan, E.W.W.; Chang, R.K.C. QDASH: A QoE-Aware DASH System. In Proceedings of the 3rd Multimedia Systems Conference, MMSys '12, Chapel Hill, NC, USA, 22–24 February 2012; Association for Computing Machinery: New York, NY, USA, 2012; pp. 11–22. [[CrossRef](#)]
38. Xue, J.; Zhang, D.Q.; Yu, H.; Chen, C.W. Assessing quality of experience for adaptive HTTP video streaming. In *2022 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, Durban, South Africa, 4–5 August 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–6. [[CrossRef](#)]
39. *G.711: Pulse Code Modulation (PCM) of Voice Frequencies*; Recommendation G.711; Telecommunication Standardization Section of ITU: Geneva, Switzerland, 1990.
40. Berger, J.; Sochos, J.; Stoilkovic, M. *Network Performance Score*; Technical Report; Rohde & Schwarz: Munich, Germany, 2020.
41. Chen, Y.; Tu, L. Density-Based Clustering for Real-Time Stream Data. In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '07, San Jose, CA, USA, 12–15 August 2017; Association for Computing Machinery: New York, NY, USA, 2007; pp. 133–142. [[CrossRef](#)]
42. *P.913: Methods for the Subjective Assessment of Video Quality, Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in Any Environment*; Recommendation P.913; Telecommunication Standardization Section of ITU: Geneva, Switzerland, 2021.
43. *Fifth Generation Fixed Network (F5G) F5G High-Quality Service Experience Factors Release #1*; Group Specification ETSI GS F5G 005 V1.1.1; European Telecommunications Standards Institute (ETSI): Valbonne, France, 2022.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.