applied sciences

MDPI

Article

Unsupervised Learning Data-Driven Continuous QoE Assessment in Adaptive Streaming-Based Television System

Paweł Andruloniw [1,2,*], Karol Kowalik [2] and Piotr Zwierzykowski [1]

[1] Institute of Communication and Computer Networks, Poznań University of Technology, 60-965 Poznań, Poland
[2] Fiberhost S.A., 60-211 Poznań, Poland
* Correspondence: pawel.andruloniw@doctorate.put.poznan.pl or pawel.andruloniw@fiberhost.com

Featured Application: In modern television systems based on adaptive streaming technology, an assessment of customer contentment might be necessary to deliver the highest possible audio and video quality. Data-driven 'quality of experience' methods based on continuous clustering can be a solution for the problem of service level assessment from the perspective of customers.

Abstract: The quality of experience (QoE) assessment of adaptive video streaming may be crucial for detecting degradations impacting customer satisfaction. In a telecommunication environment, eliminating failure points may be the highest priority. This study aims to assess the QoE level of the video played by the STB device connected to the production TV system. The evaluation has been based on the stalling effects, video quality changes, and the time related to the last decreased bitrate change occurrence. The two-phase continuous clustering approach has been studied to assess the QoE level based on the ACR scale. The number of devices with grades 1 or 2 is relatively low, but those devices generate significantly more events than adequately functioning devices. STBs try to play the highest possible bitrate, and there is no possibility of setting the intermediate bitrate level. The STB player does not have the button to set the quality level, usually available in pure over-the-top applications. Hence the bitrate fluctuations that can annoy customers appear for the lowest grades. The boundary cases can be easily assessed. The outcome should be challenged by the customers' opinions to find the proper QoE threshold. Continuous clustering may allow telecom operators to assess customer satisfaction with their TV service.

Keywords: artificial intelligence; unsupervised learning; clustering; quality of experience; adaptive streaming; over-the-top

check for updates

Citation: Andruloniw, P.; Kowalik, K.; Zwierzykowski, P. Unsupervised Learning Data-Driven Continuous QoE Assessment in Adaptive Streaming-Based Television System. *Appl. Sci.* **2022**, *12*, 8288. https://doi.org/10.3390/app12168288

Academic Editors: Katarzyna Węgrzyn-Wolska, Piotr Prokopowicz and Maciej Piechowiak

Received: 12 July 2022
Accepted: 16 August 2022
Published: 19 August 2022

1. Introduction

By 2022, Cisco estimated that video network flow would account for 82% of total network traffic. In addition, 4K video resolution will be associated with 22% of the video traffic, while high-definition content (HD) will be related to 57% of the total video bandwidth. This is an increase of 19% and 11%, respectively, compared to 2017. The number of TV devices capable of serving 4K content will represent nearly two-thirds of the installed sets in 2022 and will increase from 162 to 799 million worldwide [1].

According to the rise of video traffic share, assessing the quality of service (QoS) and the quality of experience (QoE) of provided services is necessary. The QoS metric can be used to assess service parameters from a technical point of view and may include the measurement of throughput, jitter or delay, and packet loss. The QoE measurement reflects the human perception of video services that QoS-related metrics cannot reveal. Hence, the QoE is a more reliable video quality assessment from the customer's point of view when customer dissatisfaction is considered [2]. Over-the-top (OTT) video technology is based on the HTTP Adaptive Streaming application. The main factors influencing the QoE level for adaptive video streaming are stalls, video quality changes, and initial delay [3]. The stallings are connected to buffer saturation and, as a result, a black screen or rebuffering may occur. Video quality changes are associated with changing the level of video bitrate. The higher bitrate, the better the video quality. The video player can decrease the bitrate level during network impairments when the bandwidth is insufficient. Hence the video will still be played. The initial delay is related to the time needed to fill the player buffer above the specified threshold that allows starting playback. In pure OTT applications, the initial delay may also be connected to commercials played before the relevant video starts. However, the initial delay might not be as harmful as stalling or bitrate change. It was stated that the occurrence of stalling might be six times more significant than the initial delay in leaving the video [4].

Several studies on the QoE in adaptive video streaming environments have been conducted. The k-means clustering algorithm and classification approach was used to assess whether the viewer may experience a positive or negative QoE level. Mainly stallings, bitrate changes, initial delay, and duration metrics were considered. The proposed algorithm has a precision of over 96%. The study covers batch clustering with a classification of the data collected from mobile networks [5]. The video ATLAS algorithm based on a support vector regressor (SVR) can assess the QoE level based on video quality, the occurrence of stallings as well as memory-related to the last distortion interval. However, it is limited to subjective data used for algorithm training [6]. Reinforcement learning was used to increase the adaptive video streaming QoE level in mobile devices used in transport. The quality was measured on a tram, ferry, and bus. TV operators can consider QoE measurements for mobile devices since mobile TV services are often complementary to regular services [7]. The prediction approach named Streaming QoE Index considers the quality degradation due to perceptual video impairment and stallings, initial buffering, and interactions between them. The study indicates that video content coded with a constant bitrate could have different presentation qualities that might influence the QoE level [8]. Adaptive streaming QoE evaluation algorithms based on calculated peak to signal noise ratio (PSNR), video multimethod assessment fusion (VMAF), and bitrate have been proposed to consider the encoding quality, rebufferings, quality changes, and initial delays. The study points out that many short stalls can be more annoying than a single long stall. Additionally, 70.83% of test subjects answered that stalls are the most relevant metric when evaluating the quality of the video. 16.67% of users considered quality the most relevant aspect [9]. The Hammerstein–Wiener predictor has been used to create a QoE evaluator called time-varying QoE Indexer, which accounts for interactions between stalls, analyzes video content and perceptual video quality, and predicts continuous-time QoE. The proposed predictor considers the number of stalls and their length, time since the last stall, frequency of stalls, rebuffering rate, buffer model, scene criticality, and perceptual quality [10]. An unsupervised learning approach has been studied to assess the QoE in IPTV services along with self-organizing maps. That approach included a full-reference model, visual distortion, and packet loss considerations and was suitable for assessing the video quality in broadcasting TV systems [11]. Another study that investigated unsupervised learning for QoE measurement presented the real-time algorithm deployed on the server side and the offline no-reference assessment on the customer side. A restricted Boltzmann machine (RBM) was trained based on, inter alia, bitrate, video motion, or blur mean metrics and sent to the customer side to execute the detailed measurement in real-time [12]. The event-based perceptual quality (EPQ) framework has been introduced to estimate the QoE concerning rebufferings and memory mechanisms. The measurement is performed in a real-time scenario, and the perceptual quality can be returned at any time. The EPQ output is consistent with the logarithmic nature of the human perception system regarding video impairments insight [13]. Back-propagation neural networks and random forests were applied to assess stallings and initial buffering delays in the QoE estimation. Metrics evaluated based on encrypted HTTP traffic are common in video transmission, specifically from the perspective of network operators [14]. Back-propagation neural networks have likewise been utilized to assess video quality based on quantum placet, bitrate, and motion vectors. Research has mainly analyzed the compression and network transmission damage and mapped it to the mean opinion score (MOS) grade [15]. Ridge regression, a three-dimensional convolutional neural network, and LSTM were used to build a QoE assessment model that analyzed video quality, fluency, and volatility. During the publication, the proposed model can outperform state-of-the-art models. However, the model complexity may be a drawback for real-time analysis by network operators [16]. Random forest, neural network, and LSTM algorithms were used to create a QoE assessment model to analyze YouTube videos. Those models applied the initial playback delay, video streaming changes, quality, and buffering to return the quality prediction [17]. The SSIM metric and neural network have been implemented to predict the quality of videos encoded with H.264 and H.265 codec by one mapping function. The model is dedicated to IPTV systems and returns the ACR grade [18]. Motion vectors, spatial image features, and transmission impairments were utilized to develop the assessment model that includes the SVR regression to return the mean opinion score value [19]. LSTM network was also utilized to assess the quality of HTTP streaming sessions based on bitstream-level parameters, stalling effects, and padding of adaptive streaming segments. The segment MOS (S-MOS) has been found as the best metric to return the segment quality grade. The proposed model has an RMSE of less than 0.479, depending on the analyzed dataset [20]. Another neural network approach forecasts the streaming video quality and degradation before the user notices it. The paper proposed a time series solution based on BiLSTM-CNN and compared it to solutions based on SVR, MLP LSTM, and BiLSTM methods. The RMSE of the proposed algorithm was less than 0.1 regarding used metrics, including playtime, average buffering, or buffering frequency [21].

This study aimed to investigate whether the online clustering methods can be used to assess the customer QoE level in the production TV system based on adaptive streaming technology. The evaluation was grounded on the data-driven approach and parameters that include stall occurrence, bitrate level, frequency of quality switches, time on the decreased bitrate, or stalling ratio. The event logs gathered from the set-top-box devices are related to bitrate level, and rebufferings were employed. A continuous data processing method which refers to online clustering was chosen since it analyzes the devices operating in a production television system in real-time. In the end, the QoE level outcome was applied to study the degradation within specified geographical areas based on aggregation routers across Poland. The primary motivation for this was to create a simple system that would allow service providers to assess the perceived quality of live TV streaming in their system. The assessment would be performed in real-time based on data collected from set-top boxes. The active number of devices at a particular time requires an analysis based on logs indicating what happens on the decoder. Of all the data, the most relevant is information about the video bitrate level, stallings, the start and end of watching an event, and the use of additional features. The analysis should be performed in real-time and return the result that can be checked against customer requests later. Based on the collected data, the operator would geographically locate faults and, as a result, in the time of an increased influx of errors, more resources can be redirected by the operator to correct the degradation which has occurred.

The article is divided into four sections. In the second section, the gathered dataset, quality of experience assessment scale, as well as television background followed by clustering features has been provided. The third section covers the results of the geographical area study. The last section provides a discussion along with the proposed scope of future work.

2. Materials and Methods

2.1. Television Background

The principal television over-the-top (OTT) platforms provide at least live channel services. The adaptive streaming television might employ the MPEG-DASH [22], Apple HLS [23], or Microsoft Smooth Streaming [24] protocols to deliver multimedia content.

Additional features include content playback in the given time (catchup), timeshifting, personal video recording in device storage (PVR) or the cloud (nPVR), and finally, video-on-demand content (VOD). Set-top-boxes (STB) offered by telecoms frequently run the Linux or Android-based software that serves additional options, such as content recommenders, profile management, and third-party applicants (Netflix, YouTube, Disney, and others). However, out of all the listed capabilities of the STB, the most important ones from the TV provider's perspective are those related to living content.

2.1.1. Quality of Experience

The quality of experience (QoE) measure has been proposed to more accurately estimate perceived quality and reflect viewers' perception of multimedia content more precisely than the quality of service (QoS) metrics. The QoE assessment can be subjective or objective. The first approach calculates the arithmetic mean of assessments collected from the subjects. The subjective assessment might be difficult to conduct on a large scale. Furthermore, such judgment is not considered when the assessment has to be gathered continually. The objective QoE can be computed by a reference algorithm adjusted to service provider infrastructure [2]. The QoE algorithms in IPTV services will be based on the other parameters in pure OTT applications or OTT-based television services deployed on telecom operator-managed networks. The proposed solution employs the reduced-reference approach. Applied reference is represented by the highest available bitrate for a given video adaptation set. The feature is extracted from the configuration available on the CDN server.

2.1.2. Category Rating

Absolute category rating (ACR) is a judgment that can be used to estimate the subjective quality of audio and video content. The impairment of service level might be expressed in the degradation category rating (DCR) (Table 1). ITU-T proposed that the grading scales assess the subjective level of perceived video and audio quality. However, listed gradations can be adopted for data-driven objective assessments [25]. The highest ACR grade describes excellent service quality, while the lowest means the perceived quality of content is terrible. The highest DCR grade assumes that the degradation is imperceptible, and the lowest impairment reaches annoying levels.

Table 1. Absolute category rating (ACR) and degradation category rating (DCR) scale levels.

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

2.1.3. OTT Service in the Managed Network

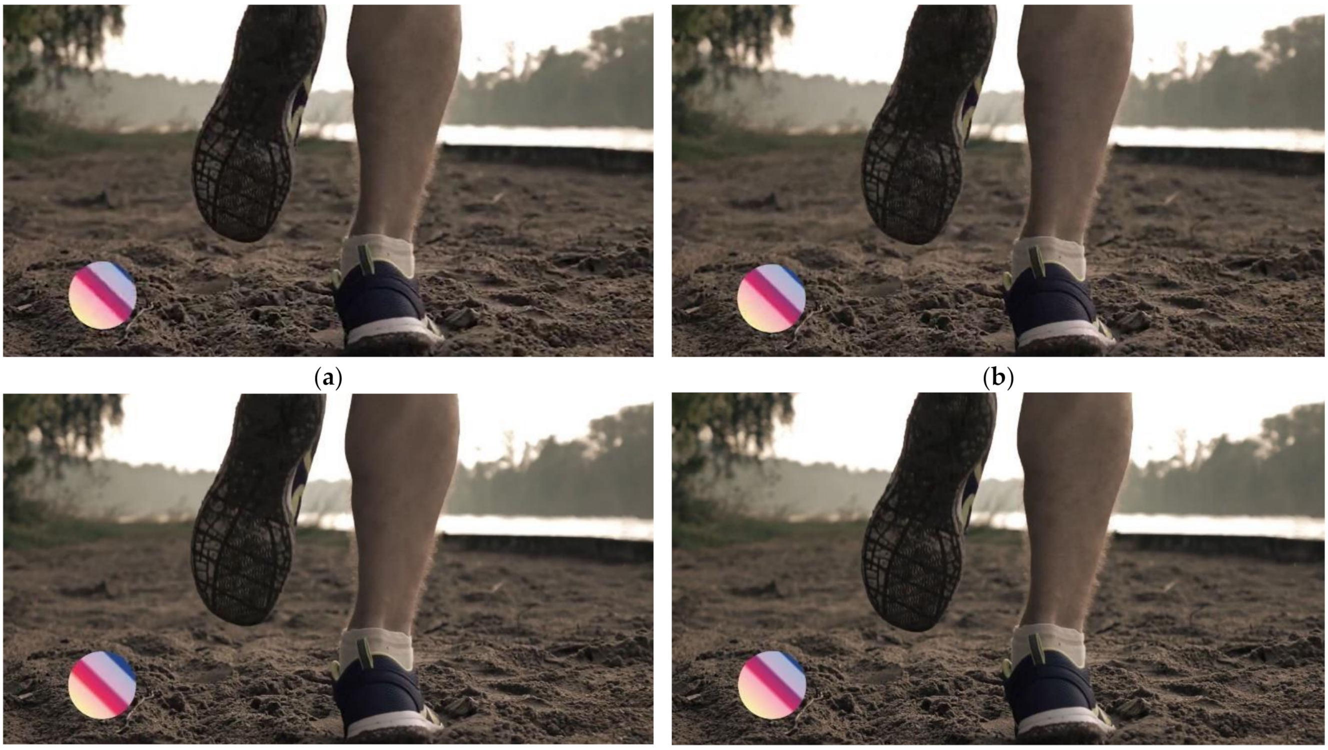
The research has been conducted based on data collected from the Hiway TV [26] production system owned by the Polish telecom operator INEA [27]. TV services are only provided on the INEA's gigabit passive optical network (GPON), which is more reliable than the hybrid fiber-coaxial (HFC) network. TV features include live streaming, catchup, timeshift, video-on-demand, and given days back and forth recording. Live channels have different resolutions (SD, HD, and 4K) and bitrates (Table 2). A single channel can have multiple bitrate profiles, commonly from two to four, and resolutions from SD up to 4K. The SD and HD channels are encoded by H.264 codec with Main or High profile, level 4.1, 4:2:0 color sampling. The 4K content is encoded by H.265 Main 10 profile. The number of frames depends on the input received from the content providers. The frame number is passed through to the output. The TV operator distinguishes between ordinary and premium channels due to the importance of channels based on viewership, content consumption, or agreements with content providers.

Table 2. Channel profiles for the specific resolution.

Channel Quality	Bitrate Range [kbps]	Resolution
SD	800–2500	640 × 480
HD (ordinary channels)	3000–6000	1920 × 1080
HD (premium channels)	6000–8000	1920 × 1080
4K	>10,000	3840 × 2160

The STB has Linux-based software that allows customers to create favorite channel lists, profiles with their settings, and multi-room options. Events connected to catchup, timeshift, or recordings inherit the parameters from the channel settings. Thus, in a single household, up to five devices can be established. Decoders can be connected to the optical network terminal (ONT) by Wi-Fi (802.11 b/g/n/ac) or ethernet cable.

The live channels and timeshift, VOD, or nPVR events can be available with up to 4 video profiles (Figure 1). The 800 kbps (Figure 1a) and 1500 kbps (Figure 1b) are unacceptable to display on the TV screen. Those qualities may be sufficient for mobile. The 3000 kbps (Figure 1c) and 6000 kbps (Figure 1d) are qualities adequate to watch on TV screens. The highest video profile for a particular part of the content is treated as a reference bitrate.



(**a**) (**b**) (**c**) (**d**)

Figure 1. Different video profiles are available for the viewers: (**a**) 800 kbps, (**b**) 1500 kbps, (**c**) 3000 kbps, and (**d**) 6000 kbps. The video screens were derived from the Fiberhost S.A. video test file.

A black screen with a notification message indicates the stallings related to the buffer saturation (Figure 2). After the black screen appears, the STB can recover autonomously.

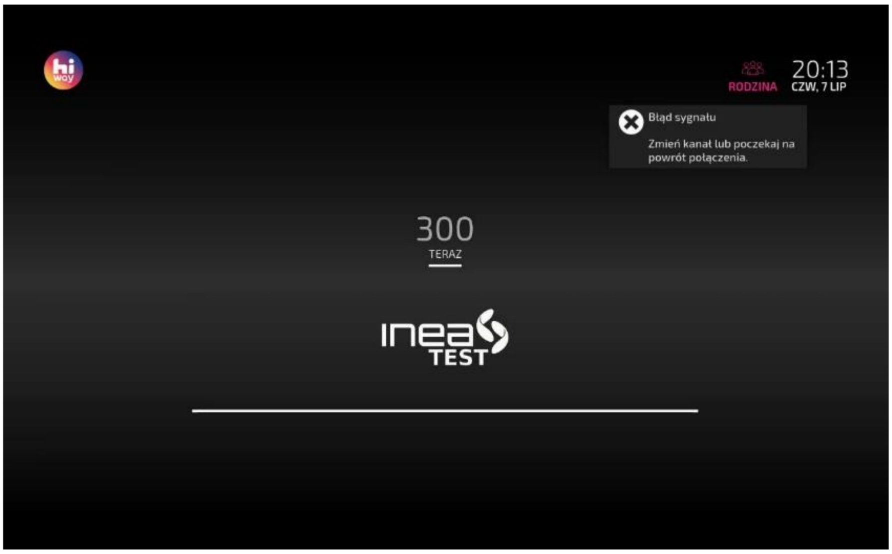


Figure 2. Stalling occurrence on the operator's STB.

2.2. System Scheme

The simplified system scheme (Figure 3) includes the TV headend, GPON access network, and households section representing STBs connected to the ONT.

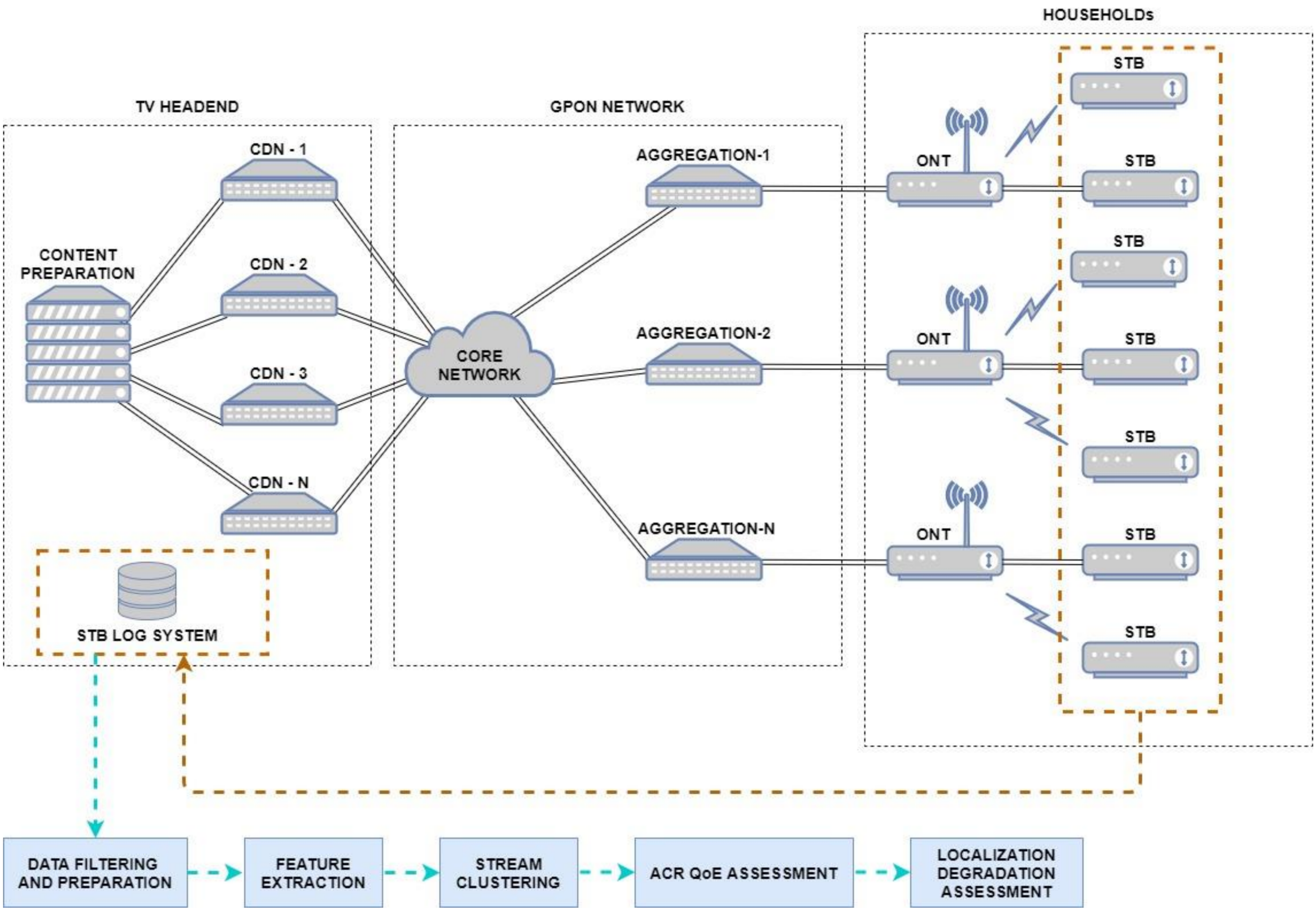


Figure 3. The simplified OTT television scheme with the stream clustering approach to QoE estimation.

A single household has at least one television set-top-box and optical network terminal used to connect to Internet and TV services. Through the backbone and GPON network that includes aggregation routers, the video content is delivered from the TV headend to the customer site. The main task of the headend is to receive multicasts that bear linear channels from content providers, transcode them, and generate HTTP-based adaptive streams. Activities that occur on the STBs are transmitted to the central log system. Logs can contain entries related to channel playback, bitrate level, stalling errors, additional feature utilization, or even network parameters. Due to the amount of data, it has to be filtered. Feature engineering is applied to obtain data related to the played stream, bitrate level, stallings, or playback time of a specific bitrate. During the phase, the input features for the continuous clustering process are calculated. Based on clustering output, QoE assessment and degradation localization are performed. The GPON aggregation routers are located across Poland. After the QoE assessment, the external IP addresses used by STBs to receive the content from CDN servers to discover the aggregation router assignment, as the IP addresses are not stored in the television log system.

The current data-driven QoE level to the particular STB is computed during the stream clustering. After, the IP mapping of the region with the lowest level of experience can be discovered.

2.3. Gathered Data

A total of 37,283 devices were analyzed during the 12 h period from 12 a.m. to 12 p.m. on the 14 June 2022. The dataset consists of 4,211,336 entries. From the QoE assessment perspective, only the events connected to changing the STB state, event playing, channel profile playback, and stalling occurrences are essential for further processing. Event logs contained incidents related to starting or shutting down the decoder, changing a channel or event, playing a live channel, and using auxiliary functions such as timeshift, catchup, and nPVR. Bitrate changes, frequency, and the playback time of a particular video profile were also taken into account, as were the occurrences of rebufferings and their duration.

Features

The clustering algorithm input consists of four main features calculated based on heuristic measurements gathered from production devices. Authors have proposed newly calculated parameters to connect the stalls occurrence, frequency, duration, and bitrate switch metrics, including occurrence, frequency, and duration.

The proposed parameter SCI (Streaming Change Importance) will indicate whether the change in bitrate playback is positive or negative in the present time n to the previous time $n-1$. The negative change means that the set-top-box in the time n plays lower video quality regards time $n-1$ and vice versa. The bitrate level in the given time and reference bitrate is gathered directly from the STBs logs.

The stalling occurrence for the analyzed system has been illustrated in the recorded video attached to the Supplementary Files. The sequence presents stalling from the production set-top-box that occurred while watching a football match. When a signal loss occurs, the customer can notice the black screen with the audio/video error message. The value "−1" generally represents the loss of signal. The formula of SCI is as follows:

$$SCI = \begin{cases} -1, bitrate^{N}_{kbps} = 0 \\ \dfrac{bitrate^{N}_{kbps} - bitrate^{N-1}_{kbps}}{ref_{bitrate}} x \geq 0 \\ 1, bitrate^{N}_{kbps} = ref_{bitrate} \end{cases} \quad (1)$$

In Formula (1), the bitrate is the bitrate level in kilobytes per second in the present time (n), as well as in the previously gathered entry ($n-1$). The $ref_{bitrate}$ determines the highest available bitrate for a given channel gathered from the CDN configuration.

When the stalls occur, the current bitrate level will decrease to 0, or the system error will occur (i.e., $bitrate^{N}_{kbps} = 0$) in which case SCI will take the value -1. This means that the significant change in the stream occurs in the negative direction. The customer will perceive that the black screen that is not desired. On the other hand, when the current bitrate level reaches the maximum intended level, i.e., the $bitrate^{N}_{kbps} = ref_{bitrate}$, the SCI will take the value 1. This means that the positive change in stream occurred, or in the given sampling time, the set-top-box plays the highest available quality. When the currently played stream is higher than 0 and lower than the maximum available value, which can be different for every channel, and the SCI will be assigned to the values from the $(-1; 1)$ range. If the bitrate in time n is improving, and it is higher than the bitrate in time $n-1$, the SCI will have positive values. On the other hand, the SCI will take the negative values. The SCI parameter is to provide information on whether, in the given time, the video quality improved or deteriorated.

When set-top-box starts playing, it can reach the highest possible bitrate level (Figure 4). In the example time t_{n-2}, the SCI will reach the value "1", as the degradation is not observable. When the bitrate decreases by half in time t_{n-1}, the SCI will reach the value "-0.5". In time t_n the SCI will become "-1" due to the stall occurrence, which is highly observable, and the degradation is significant. After the given period of time, the STB can recover the signal, and at times t_{n+x}, the SCI will take positive values.

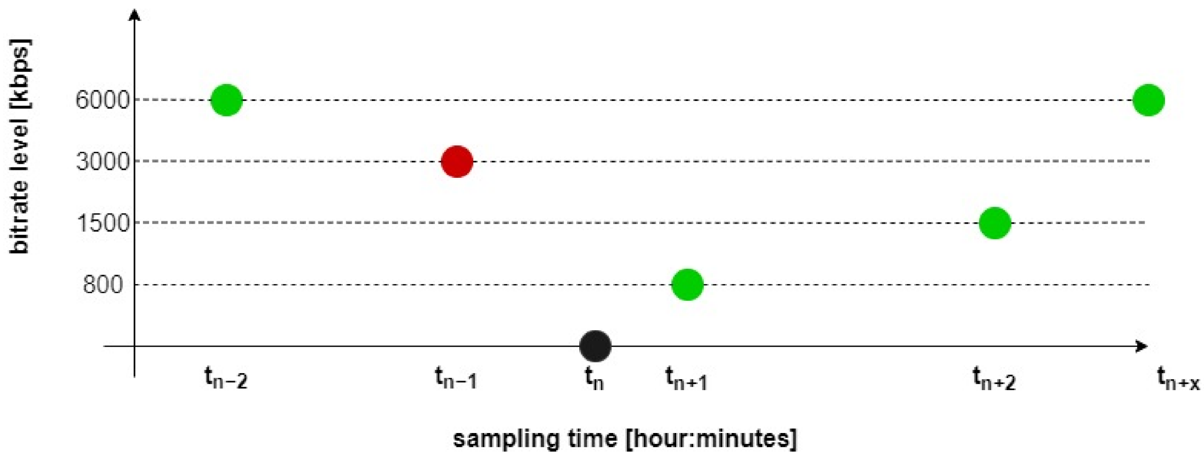


Figure 4. The example of bitrate change sequence, green—improvement, red—degradation, black—stall.

The SCTI (stream change importance in the given time) parameter is proposed based on the SCI parameter. The arranged feature aims to link the positive or negative quality change with the given TV event or channel importance. From the TV operator's perspective, specified channels can be more important than others. The importance can be identified by the occupied number on the channel list, increased maximum bitrate, or audience. The channel importance is set as the current viewership in the given model. The current viewership is calculated every 15 min time interval (TI) for every channel (CH) available to customers. The audience calculated in such a way is simplified since there is no mechanism to assess the number of people jointly watching the TV screen. In Formula (2), the simplified audience is assessed as the number of active devices divided by all of the running devices.

$$channel\ viewership^{TI}_{CH} = \frac{active\ devices^{TI}_{CH}}{all\ active\ devices^{TI}_{CH}} \quad (2)$$

The edt_n (Figure 5) expressed in Formula (3) determines the playback time in seconds for the currently played bitrate level. It is calculated to measure how long the STB plays the video quality. Extended edt_n for decreased bitrate indicates that there can be significant problems with the network used by the STB. The decoder downloads the lower-quality chunks since the network performance cannot be adequate.

$$edt_n = t_n - t_{n-1} \tag{3}$$

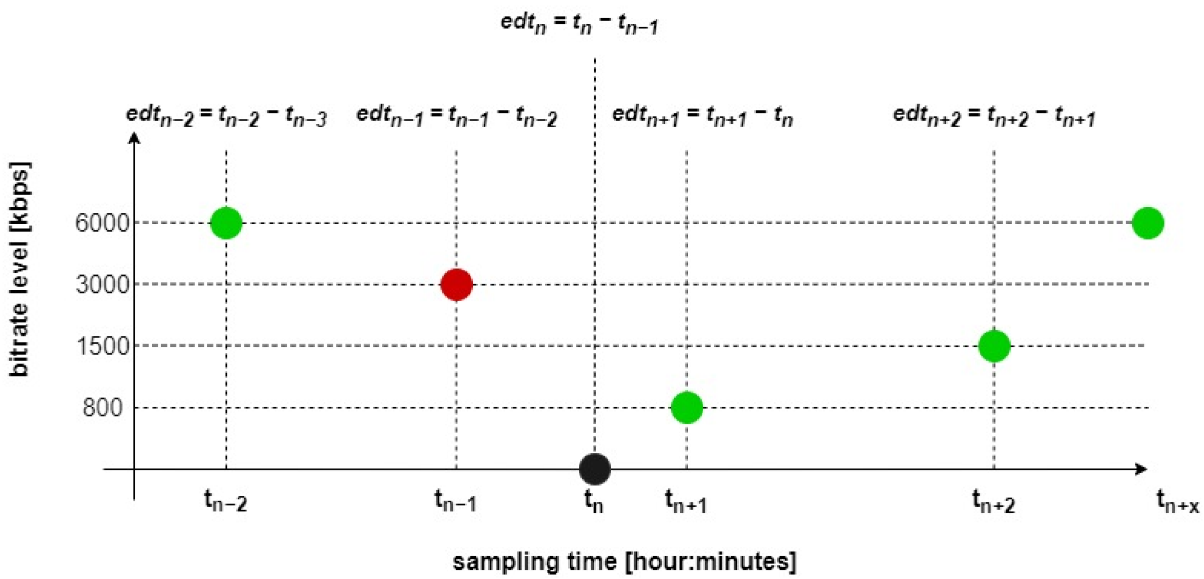


Figure 5. The example of event difference time (*edt*) parameter analysis, green—improvement, red—degradation, black—stall.

SCTI can determine how long the STB plays the video with decreased quality. Secondly, it assesses how effectively the playout of the decreased bitrate concerns current channel viewership. In Formula (4), the SCTI links the audience, playback time for the current bitrate level, and SCI parameter. The low values of SCTI (around 0) can indicate that the problems do not occur on the vital channel since the audience is low. Similar values can mean that the device is quickly changing bitrate qualities. Thus, there is a customer network instability or only temporary problems. On the other hand, many devices with increased SCTI values might indicate a global problem with vital channels. When the SCTI returns increased values during prime time, it can indicate to a TV operator that a specific channel may need on-duty engineer intervention.

$$SCTI = log_2(1 + edt_n) * channel\_viewership_{CH} * SCI \tag{4}$$

The usage of stalling ratio for video quality assessment has been proposed by Huawei [28]. Stalling time regarding the playing session time is represented by STCSI (stalling coefficient within session increasingly) parameter. The STCSI tracks how long the decoder displays the black screen with audio/video error. From the perspective of the clustering approach, the stalling ratio is expressed as a negative value, as the occurrence of video freezes is undesired.

$$STCSI = \frac{-1 * \sum stalling\_time_s}{\sum session\_event\_time_s} \tag{5}$$

In Equation (5), the $stalling\_time_s$ is the summary time in seconds of stalling events, and the $session\_event\_time_s$ determines the time since the customer has enabled the STB.

The memory feature (8), VSBCT (viewing session bitrate counter), is proposed to assess the significance of degradation based on the time since the last bitrate change occurred. According to [9], many negative quality switches, and stalls over short intervals, can be more annoying than a single more extended quality change. The proposed parameter aims to link the number of occurred problems, including bitrate switches and stalls, to the time that passed since the last problem occurred. In the given proposal, the bc_c (bitrate change counter) (6) is increased every time the negative bitrate switch occurs, and the stalls are considered zero value bitrate. The counter is not increased when the current bitrate level is equal to the maximum bitrate or greater than the previous sample's video quality. The counter can increase during the watching session. When the customer power off the device, the counter is reset. It is related to the soft reset of STB, or ONT, which can often help with network environment problems.

$$bc_c = \begin{cases} bc_c, bitrate_{kbps}^{N} > bitrate_{kbps}^{N-1} \\ bc_c + 1, (bitrate_{kbps}^{N} \neq ref_{bitrate}) \leq \ bitrate_{kbps}^{N-1} \\ bc_c, (bitrate_{kbps}^{N} = ref_{bitrate}) \end{cases} \tag{6}$$

The *tslbc* (time since last bitrate change) is the time passed from the last problem occurrence. The *tslbc* updates when the currently played bitrate is lower than the bitrate read in the previously gathered sample.

$$tslbc = \begin{cases} t_{n-x}, & bitrate_{kbps}^{N} > bitrate_{kbps}^{N-1} \\ t_n, & bitrate_{kbps}^{N} \leq \ bitrate_{kbps}^{N-1} \end{cases} \tag{7}$$

The proposed VSBCT parameter decreases when the bitrate quality switches and stalls frequently occur over a short interval. Mentioned device behavior may mean significant problems with the customer's TV set (ONT + STB). The VSBCT will highly decrease if the number of problems is excessive. On the other hand, if the impairment was intensive, but the device recovered the appropriate parameters, the VSBCT will tend to the maximum 0 value. The primary purpose of the proposed parameter is to detect situations when high degradation occurs at the beginning (Figure 6). After a specific time, the STB plays well, and degradation disappears.

$$VSBCT = \begin{cases} 0, bc_c = 0 \\ -1 * \dfrac{log_2(1 + bc_c)}{tslbc}, bc_c > 0 \end{cases} \tag{8}$$

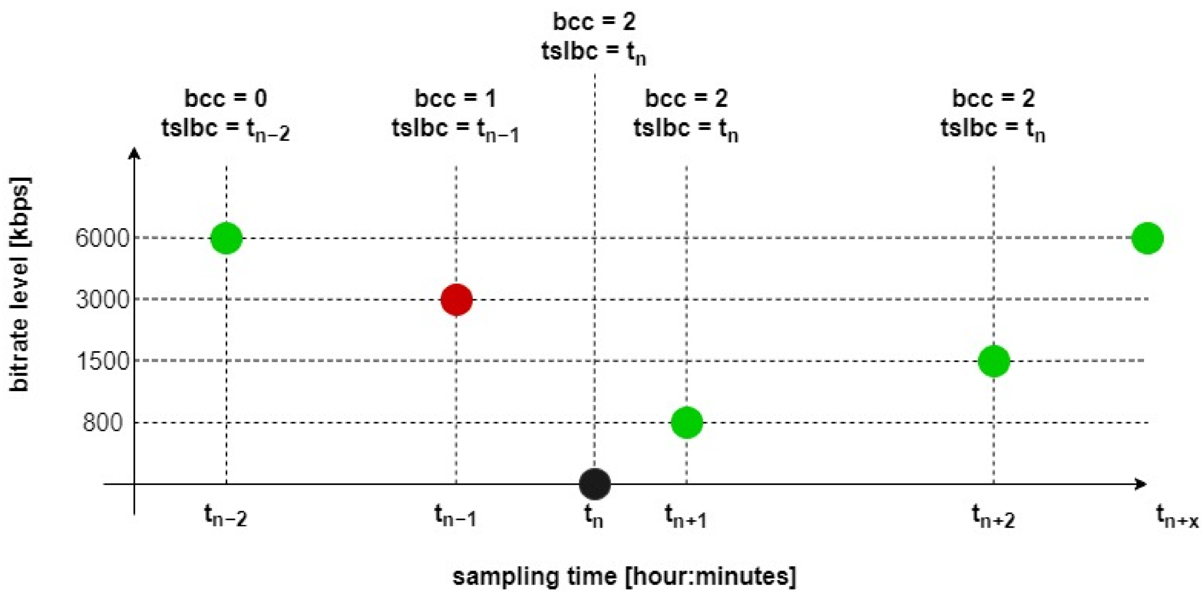


Figure 6. The example of bitrate change counter (*bcc*), and time since last bitrate change evaluation (*tslbc*), green—improvement, red—degradation, black—stall.

The Pearson correlation coefficient of input parameters has been calculated for the entire dataset. (Figure 7). Only SCI and SCTI parameters have an increased positive relationship.

	SCI	SCTI	STCSI	VSBCT
SCI	N/A	0.57	0.08	−0.32
SCTI	0.57	N/A	0.07	−0.12
STCSI	0.08	0.07	N/A	0.19
VSBCT	−0.32	−0.12	0.19	N/A

Figure 7. Pearson correlation of input parameters for the entire dataset.

2.4. Clustering

STBs send event logs to the data collector when plugged in. Data can be sent on standby and during content playback. It is necessary to use a continuous approach to assess the QoE level continuously. The stream clustering algorithms might be a solution to that issue. The incoming data stream can be shown as $DS = \{x_1, x_2, x_3, \ldots, x_N\}$, where the x_i is the single event and n goes to infinity [29]. Each plugged-in STB is the source of the x_i data instances. Data stream algorithms can use a single-phase approach, but most use a two-phase approach [30]. In the first phase, the synopsis of the stream is calculated and updated when a new entry appears. The second phase is used to cluster the data on the ground of a calculated summary.

Five clustering algorithms were considered during the experiments (Table 3). The evoStream [31], DBStream [32], as well as D-Stream [33] were tested. Those algorithms are among the most popular stream clustering methods. The evoStream implements the evolutionary algorithm to optimize the macro cluster assignment. evoStream utilizes the idle time between new stream objects' appearance to build and refine macro clusters incrementally. The idle time optimization might be an effective solution for the data stream that comes from the TV system due to the variability of data arrival. The heuristic approach in the offline phase might be a significant asset regarding the QoE assessment in grade optimization. The DBStream is a micro-cluster-based algorithm that can utilize the shared density mechanism for reclustering. It has been shown that DBStream can outperform other stream clustering algorithms in most settings [34]. The D-Stream is a density-based algorithm that uses the grid structure to associate incoming data instances with grid cells. The D-Stream can adjust the returned groups in real time and capture the changing streaming character due to the usage of the density decaying technique.

Table 3. Time comparison of chosen stream clustering algorithms.

Algorithm	Duration [s]
DBStream/Hierarchical	338
DBStream/K-means	195
evoStream	1837
D-Stream/Hierarchical	252
D-Stream/K-means	186

All of them returned comparable output. The final choice was based on the processing time, which may be a crucial metric for clustering rapidly emerging data instances from the production TV system. The evoStream algorithm was the slowest during the clustering of the dataset, split into 32-instance batches (Table 3). The K-means [35] in the second phase returned a similar output to DBStream [32] and DStream [33] in the first phase. The K-means overcame the hierarchical algorithm, but the DStream and hierarchical algorithm were eventually chosen. The output of K-means was difficult to map into the ACR scale, as the K-means can return the other cluster assignment in every clustering process due to initial algorithm points. The agglomerative algorithm will always return the same clusters for the given dataset and thus can be considered sufficient during the current research. The selection of the most efficient stream clustering algorithm for QoE assessment for production TV systems may be the subject of further research and will not be considered in this article. The continuous TV system development that results in enormously increased data stream size has to be deliberated, as well as the idle times between the appearance of new data points.

3. Results

3.1. Clustering Results

All 361,746 events have been recognized as entries that may impact the QoE level of viewers. A total of 49,629 events were generated by 212 devices and marked by the algorithm as the lowest grade. A total of 1093 devices reached an unsatisfactory penultimate grade. In total, 1305 (0.036%) devices were affected by the two lowest quality levels. On the other hand, 32,319 (86.7%) decoders reported at least one event with the highest possible quality (Table 4). The perceptible but not annoying level was determined for 13,017 STBs, and a slightly annoying stage was characteristic for 7738 devices. The single STB may report several grades during the study.

Table 4. ACR assignment of unique devices and clustered events.

Grade	No. Devices	No. Events
5	32,319	103,270
4	13,017	43,073
3	7738	69,334
2	1093	96,440
1	212	49,629

The ACR assignment is reflected by cluster allocation referenced to SCI, STCSI, SCTI, and VSBCT features. Every single entry consists of four features during the clustering process, and the returned cluster is considered ACR grade. The VSBCT feature implies the most on the returned cluster during considering the given dataset (Figure 8). The VSBCT parameter can reach lower values than SCI, SCTI, and STCSI since in the dataset can be found devices that log an enormous number of negative bitrate switches. Those devices are often connected wirelessly with decreased signal power. The VSBCT for 5 varies between -1 and -2.585. Those values mean that no degradation occurred or the significance of degradation related to a smaller bitrate than the source is low. Single bitrate changes could occur, but if there is no following negative bitrate switch in time, the VSCBT is improving and aims to -1. For the lowest cluster 1, the values are between -11.1693 and -7.5314, which means that degradation is evident. Many bitrate switches occur over the whole watching session. There is no period when the signal quality can improve, and as a result the customer can observe degradation almost over the whole watching session.

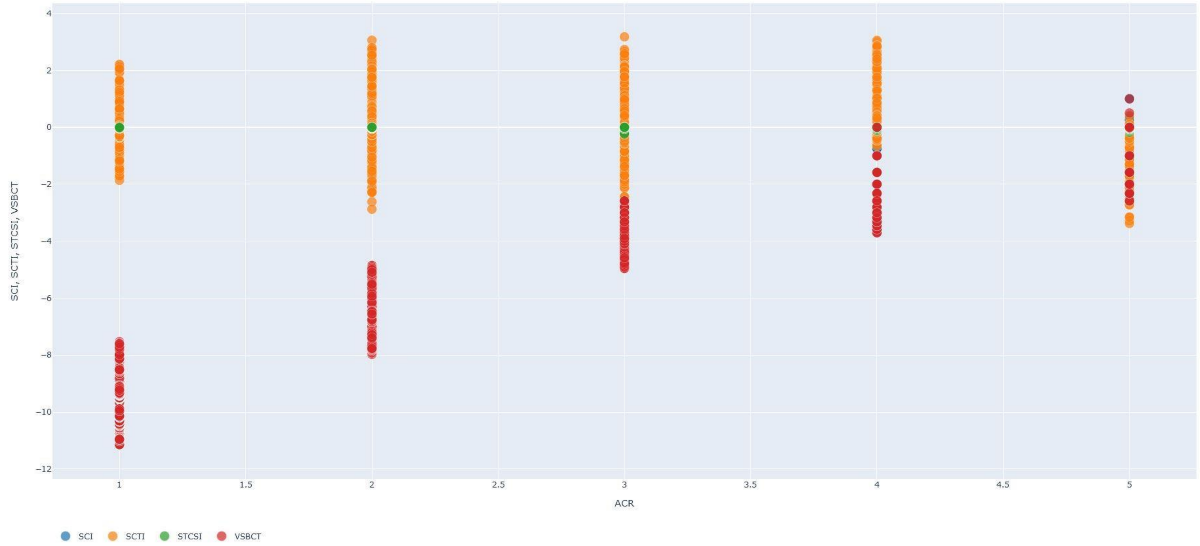


Figure 8. ACR assignment by SCI, SCTI, STCSI, and VSBCT.

3.2. Boundary Cases

The boundary cases have been determined to examine the results. The reference device with the excellent quality level has been a test device connected closely to the TV headend infrastructure. On the other hand, the worst scenario has been found within the examined devices located at the customer's site.

3.2.1. Reference Device

The reference device is the STB connected by wire to an optical network terminal (ONT) in the close neighborhood of the content-delivery-network (CDN) servers. The reference STB is managed by software that automatically switches channels to check for loss of signal or stream degradation occurrences. Due to the environment, the STB maintains the highest possible bitrate for every channel.

The data-driven quality assessment proved that the applied clustering approach returned the highest grade, indicating excellent video quality (5) for the reference device (Figure 9). The 725 events have been recognized as excellent grades. Thus, the reference device will be considered an excellent boundary case.

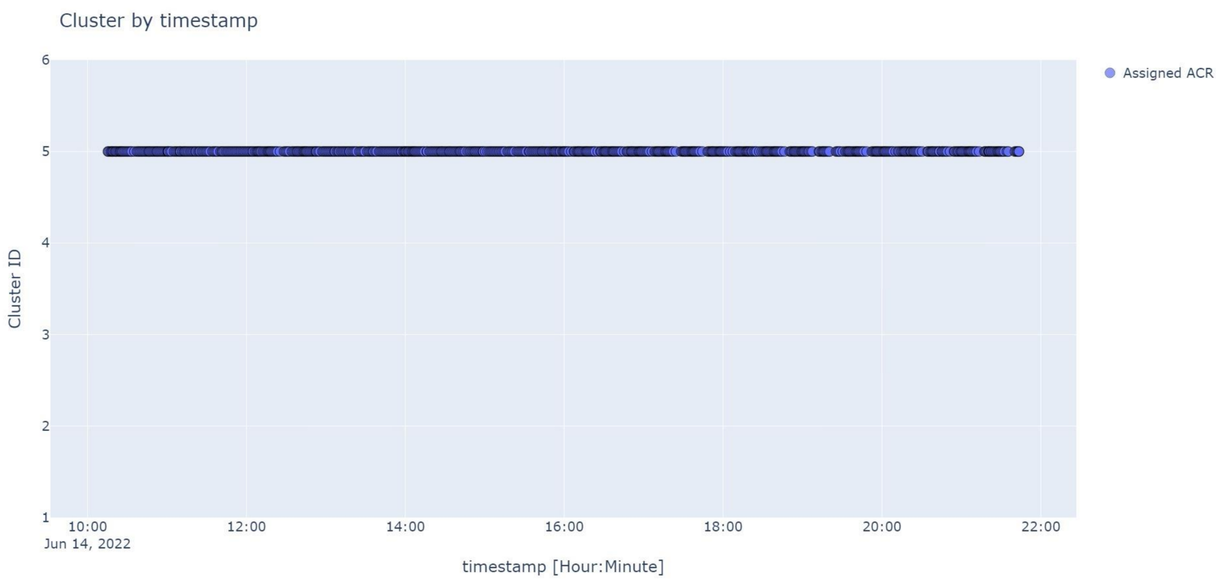


Figure 9. Reference device clustering results.

3.2.2. The Worst Case

The reference device can be considered as the appliance with the finest video quality. The explicit indication might be challenging. On the contrary, the worst or significantly bad case has to be considered. Hence, the device with the most frequent grade 1 (Bad) will be discussed.

The device with the most significant number of negative clusters has been assigned each time the stream change or stalling event occurred 3059 times (Figure 10). Only twice has the device received the 'excellent' cluster, both after the device restarted. The first time the customer started watching the session, the second, the device was restarted due to many perceived degradations. During the entire watching time, cluster 1 or 2 appeared in ca. 98.1% of assessments (Table 5).

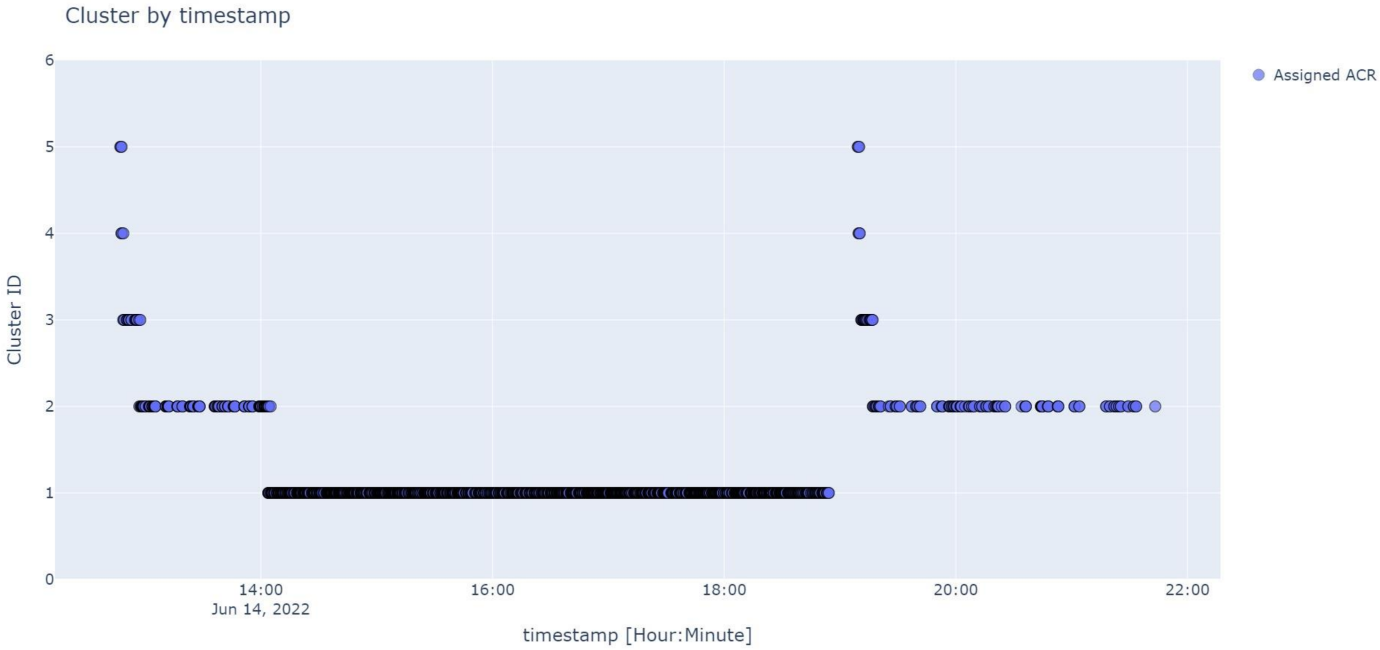


Figure 10. The device with the frequent assignment of bad grades. After the device reboot, grade 5 appeared.

Table 5. The distribution of cluster assignments for the worst case in the evaluated dataset.

Grade	No. Events
5	6
4	6
3	45
2	317
1	2685

A total of 3059 cluster assignments have been found for that device.

3.3. Standard Scenario

The standard scenario may be described as the regular STB utilized at the customer's home. Those devices often maintain good or excellent grades (Table 1). However, example devices report even poor quality (Figure 11). It can be seen that grade (marker X) is correlated with the VSBCT parameter (red dot). When several bitrate changes occur, the grade goes down.

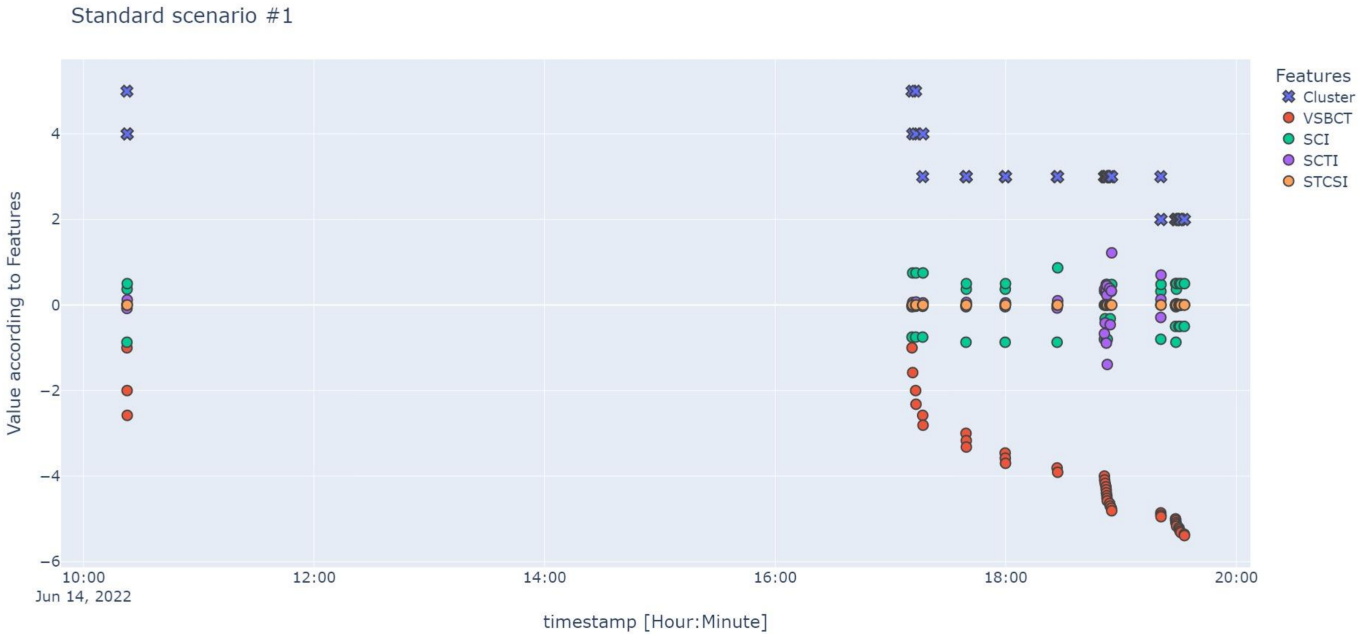


Figure 11. Clustering results for standard scenario #1. *X*-axis–hour and minute timestamp, *Y*-axis–cluster and features used during clustering.

The bitrate can change every particular number of seconds when a new chunk has to be downloaded. The SCI and SCTI parameters frequently vary between negative and positive values (orange and green dot, respectively).

Fluctuations are directly related to the bitrate changes caused by network impairments. The STB tries to download the MPEG-DASH chunk with the highest possible bitrate. The lower bitrate cannot be set consistently, similarly to pure OTT applications (YouTube, Netflix). There is no button to decrease the bitrate to omit the bitrate hopping. Hence the telecom operator has to provide the most satisfactory possible network environment.

Standard scenario #1 assessment is based on the 44 gathered events (Table 6). In the beginning, the STB starts with the 800 kbps playback. Within the next 10 s, the bitrate increased to the maximum for the given channel and reached 6000 kbps. Although the maximum bitrate was achieved, grade 4 was assigned. The quality of the video delivered by 800 kbps is significantly worse than 3000 or 6000 kbps. Therefore, the lower grade was stated. It has to be mentioned that STBs can switch to higher or lower bitrate after at least 5 s due to the MPEG-DASH chunk length. After the morning session, another playback started after 5 p.m. During the hardly 2 h and 30 min, the customer may have witnessed 40 bitrate switches, but the channel was changed only five times. In an excellent scenario, the number of bitrate switches may be equal to channel changes, and the video referenced to bitrate lower than the maximum will not be displayed. For the KinoPolska channel, the 11 bitrate changes appeared in the ca. 4-min time span and fluctuated mainly between 3000 kbps and 6000 kbps.

Another standard scenario (Figure 12) was assessed based on 14 events (Table 7). In the beginning, the STB played the video with 1255.2 kbps over 16 s, decreasing the grade from 5 to 4. After almost 1 h, the bitrate jumping appeared, and the cluster switched between 3 and 4. At 6:57 PM, the audio/video error occurred, the customer may recognize the black screen and error notification, and the grade suddenly decreased to 3.

Table 6. Events studied for standard scenario #1 shown in Figure 5.

No.	Timestamp	Channel	Bitrate	SCI	SCTI	STCSI	VSBCT	Cluster
1	2022-06-14T10:22:35.587Z	PolsatNewsHD	800.0	−0.87	−0.08	0.00	−1.00	5
2	2022-06-14T10:22:40.894Z	PolsatNewsHD	3000.0	0.37	0.03	0.00	−2.00	4
3	2022-06-14T10:22:45.589Z	PolsatNewsHD	6000.0	0.50	0.12	0.00	−2.58	4
4	2022-06-14T17:11:11.798Z	WTK	1500.0	−0.75	−0.04	0.00	−1.00	5
5	2022-06-14T17:11:32.787Z	WTK	6000.0	0.75	0.06	0.00	−1.58	4
6	2022-06-14T17:13:04.789Z	WTK	1500.0	−0.75	−0.03	0.00	−2.00	5
7	2022-06-14T17:13:16.784Z	WTK	6000.0	0.75	0.07	0.00	−2.32	4
8	2022-06-14T17:16:44.793Z	WTK	1500.0	−0.75	−0.03	0.00	−2.58	3
9	2022-06-14T17:16:54.790Z	WTK	6000.0	0.75	0.05	0.00	−2.81	4
10	2022-06-14T17:39:16.806Z	Polsat2HD	800.0	−0.87	−0.04	0.00	−3.00	3
11	2022-06-14T17:39:25.805Z	Polsat2HD	3000.0	0.37	0.01	0.00	−3.17	3
12	2022-06-14T17:39:27.825Z	Polsat2HD	6000.0	0.50	0.06	0.00	−3.32	3
13	2022-06-14T17:59:40.895Z	Polsat2HD	800.0	−0.87	−0.04	0.00	−3.46	3
14	2022-06-14T17:59:45.844Z	Polsat2HD	3000.0	0.37	0.02	0.00	−3.58	3
15	2022-06-14T17:59:51.827Z	Polsat2HD	6000.0	0.50	0.05	0.00	−3.70	3
16	2022-06-14T18:26:42.842Z	Puls2HD	800.0	−0.87	−0.07	0.00	−3.81	3
17	2022-06-14T18:27:00.839Z	Puls2HD	6000.0	0.87	0.10	0.00	−3.91	3
18	2022-06-14T18:51:19.865Z	TVP1HD	1255.2	−0.80	−0.67	0.00	−4.00	3
19	2022-06-14T18:51:26.879Z	TVP1HD	3298.4	0.32	0.37	0.00	−4.09	3
20	2022-06-14T18:51:43.849Z	TVP1HD	1255.2	−0.32	−0.42	0.00	−4.17	3
21	2022-06-14T18:52:07.864Z	TVP1HD	3298.4	0.32	0.28	0.00	−4.25	3
22	2022-06-14T18:52:15.864Z	TVP1HD	6364.0	0.48	0.35	0.00	−4.32	3
23	2022-06-14T18:52:20.862Z	TVP1HD	1255.2	−0.80	−0.89	0.00	−4.39	3
24	2022-06-14T18:52:35.853Z	TVP1HD	3298.4	0.32	0.23	0.00	−4.46	3
25	2022-06-14T18:52:40.850Z	TVP1HD	6364.0	0.48	0.45	0.00	−4.52	3
26	2022-06-14T18:52:49.852Z	TVP1HD	1255.2	−0.80	−1.39	0.00	−4.58	3
27	2022-06-14T18:54:02.850Z	TVP1HD	3298.4	0.32	0.39	0.00	−4.64	3
28	2022-06-14T18:54:22.052Z	TVP1HD	1255.2	−0.32	−0.46	0.00	−4.70	3
29	2022-06-14T18:54:55.853Z	TVP1HD	3298.4	0.32	0.33	0.00	−4.75	3
30	2022-06-14T18:55:07.970Z	TVP1HD	6364.0	0.48	1.22	0.00	−4.81	3
31	2022-06-14T19:20:40.888Z	TVP1HD	1255.2	−0.80	−0.29	0.00	−4.86	3
32	2022-06-14T19:20:43.871Z	TVP1HD	3298.4	0.32	0.13	0.00	−4.91	2
33	2022-06-14T19:20:47.886Z	TVP1HD	6364.0	0.48	0.70	0.00	−4.95	2
34	2022-06-14T19:28:20.873Z	KinoPolska	3000.0	−0.50	−0.01	0.00	−5.00	2
35	2022-06-14T19:28:24.891Z	KinoPolska	6000.0	0.50	0.02	0.00	−5.04	2
36	2022-06-14T19:28:32.875Z	KinoPolska	800.0	−0.87	−0.04	0.00	−5.09	2
37	2022-06-14T19:28:47.874Z	KinoPolska	3000.0	0.37	0.01	0.00	−5.13	2
38	2022-06-14T19:28:52.933Z	KinoPolska	6000.0	0.50	0.03	0.00	−5.17	2
39	2022-06-14T19:30:05.882Z	KinoPolska	3000.0	−0.50	−0.01	0.00	−5.21	2
40	2022-06-14T19:30:21.879Z	KinoPolska	6000.0	0.50	0.01	0.00	−5.25	2
41	2022-06-14T19:30:40.908Z	KinoPolska	3000.0	−0.50	−0.01	0.00	−5.29	2
42	2022-06-14T19:31:08.911Z	KinoPolska	6000.0	0.50	0.01	0.00	−5.32	2
43	2022-06-14T19:32:54.881Z	KinoPolska	3000.0	−0.50	0.00	0.00	−5.36	2
44	2022-06-14T19:32:59.882Z	KinoPolska	6000.0	0.50	0.01	0.00	−5.39	2

Table 7. Events studied for standard scenario #1 showed in Figure 6.

No.	Timestamp	Channel	Bitrate	SCI	SCTI	STCSI	VSBCT	Cluster
1	2022-06-14T13:30:02.272Z	TVN24HD	1255.2	−0.62	−0.23	0.0000	−1.00	5
2	2022-06-14T13:30:18.270Z	TVN24HD	3298.4	0.62	0.56	0.0000	−2.00	4
3	2022-06-14T14:36:54.475Z	PolsatNewsHD	800.0	−0.87	−0.12	0.0000	−2.58	3
4	2022-06-14T14:36:59.317Z	PolsatNewsHD	6000.0	0.87	0.39	0.0000	−3.00	4
5	2022-06-14T15:14:00.354Z	PolsatNewsHD	800.0	−0.87	−0.13	0.0000	−3.32	3
6	2022-06-14T15:14:11.354Z	PolsatNewsHD	6000.0	0.87	0.35	0.0000	−3.58	3
7	2022-06-14T18:57:02.785Z	TVPINFO	800.0	−0.87	−0.02	0.0000	−1.00	5
8	2022-06-14T18:57:05.066Z	TVPINFO	0.0	−1.00	−0.03	−0.0024	−2.58	3
9	2022-06-14T18:57:05.504Z	TVPINFO	6000.0	1.00	0.08	−0.0024	−3.00	4
10	2022-06-14T19:15:23.527Z	TVN7HD	800.0	−0.90	−0.09	−0.0007	−3.32	3
11	2022-06-14T19:15:38.538Z	TVN7HD	8000.0	0.90	0.22	−0.0007	−3.58	3
12	2022-06-14T19:26:14.576Z	TVN7HD	1500.0	−0.81	−0.06	−0.0005	−3.81	3
13	2022-06-14T19:26:21.525Z	TVN7HD	8000.0	0.81	0.20	−0.0005	−4.00	3
14	2022-06-14T19:26:21.525Z	TVN7HD	8000.0	0.00	0.00	−0.0005	−4.09	3

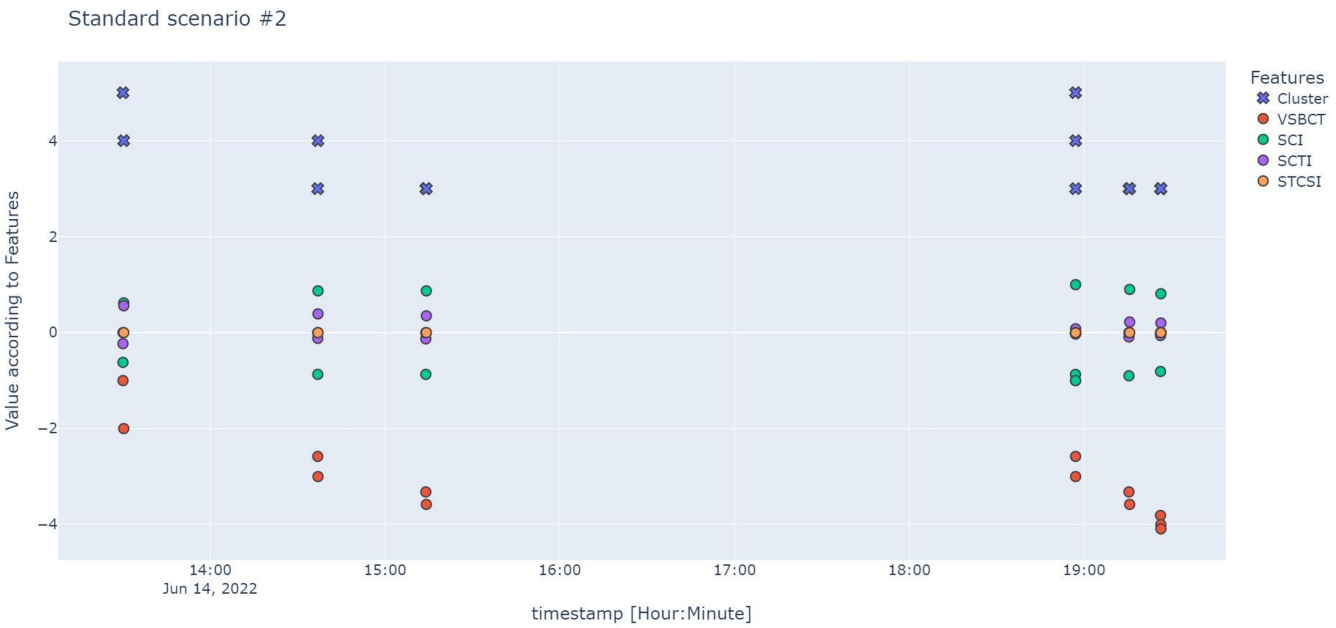


Figure 12. Clustering results for a standard scenario #2. *X*-axis–hour and minute timestamp, *Y*-axis–cluster and features used during clustering.

After half a second, the STB recovered the signal and started playing the highest possible bitrate. However, the stalling is severe degradation, so the stream recovery does not provide the highest cluster. The grade is an assessment at the moment of occurrence. In future works, constant sampling might be applied to assess the grade at every specified interval. The STBs with the detected stallings should be monitored by the TV operator. Those events might be connected with decreased wireless network quality or signal power.

3.4. Aggregation Router Assignment

Gathered QoE assessment data may be applied to estimate the degradation level of the geographical area.

STBs are connected to the aggregation terminals through the ONTs. The aggregation router names have been replaced at the operator's request. The assignment of unique devices and generated events to aggregation routers shows that many devices reach the highest possible grade (Table 8). The heatmap of devices with the highest assessment shows that most of them are connected to the single aggregation router located in the center of Poznan, the fifth-largest city in Poland (Figure 13). The common feature of all aggregation routers is that the highest viewership is noted between 7 PM and 9 PM. The higher number of devices with grade 5, the better the TV system performance.

However, to improve the quality of provided services, it is necessary to localize the source of problems with the devices that reports the Bad QoE level (Figure 14). Naturally, the highest degradation can be connected to the region with the highest population in the number of unique devices. Nonetheless, the interesting might be Aggregation-23, Aggregation-27, or Aggregation-33. In these locations, the unique devices with Bad grades were increased in reference to the unique devices with Excellent quality. The percentage share of negatively (Bad) assessed devices to positively (Excellent) was 1.76%, 1.56%, and 1.32%, respectively. Aggregation-35 has 1.06% negatively assessed devices. The ratio for the rest of the aggregation routers was less than 1%.

The number of unique devices and events have a linear relation (Figure 15). The devices are working correctly. Hence the number of events is minimal. On the other hand, the small number of malfunctioning devices can generate as many events as all devices located within a particular area that works correctly. A total of 212 devices with degradation generated almost 50,000 events. In the comparison, a similar number of events (50,198) were generated by 15,895 devices that reported Excellent quality in a given time.

Eliminating devices with impairment might effectively impact TV system performance and customer satisfaction.

Table 8. The assignment of devices and events to aggregation routers.

Aggregation Router Name	Unique Number of Devices					Number of Events				
	5	4	3	2	1	5	4	3	2	1
Aggregation-1	3911	1709	1002	144	26	12,398	5679	9120	11,547	5986
Aggregation-2	786	326	201	27	7	2526	1173	2053	3055	672
Aggregation-3	1082	476	288	45	8	3408	1638	2790	3646	2981
Aggregation-4	766	301	174	26	5	2483	955	1779	3098	978
Aggregation-5	695	306	182	23	5	2198	1060	1709	2086	464
Aggregation-6	579	191	111	12	2	1648	603	903	827	16
Aggregation-7	1039	356	199	26	5	2909	1040	1495	2000	1894
Aggregation-8	504	203	126	18	1	1614	702	1241	1552	9
Aggregation-9	604	245	156	19	3	2098	837	1284	1730	2226
Aggregation-10	374	133	78	7	1	1141	408	450	469	203
Aggregation-11	1675	649	344	58	7	5382	2036	3062	4039	1128
Aggregation-12	450	162	109	20	4	1574	585	1079	1922	1244
Aggregation-13	334	119	70	8	1	1001	354	428	691	265
Aggregation-14	921	369	206	26	3	2875	1209	1681	1716	768
Aggregation-15	1196	432	247	30	8	3720	1371	1986	3151	919
Aggregation-16	189	68	47	3	0	569	234	240	42	0
Aggregation-17	412	160	100	17	2	1395	539	909	1101	464
Aggregation-18	378	156	89	14	5	1259	505	768	1499	473
Aggregation-19	396	142	72	8	0	1230	435	602	231	0
Aggregation-20	68	25	16	3	0	204	81	114	64	0
Aggregation-21	907	336	198	33	9	2977	1042	1885	2891	880
Aggregation-22	625	248	142	15	3	1992	793	1085	736	640
Aggregation-23	1417	554	342	77	25	4049	1882	4040	8971	8933
Aggregation-24	598	288	172	32	7	1994	1019	1767	3252	1445
Aggregation-25	1732	724	442	51	12	5549	2404	3614	4748	1423
Aggregation-26	450	158	87	8	1	1424	534	658	751	274
Aggregation-27	128	53	31	7	2	392	159	317	786	205
Aggregation-28	968	431	247	33	6	3145	1382	2048	2530	722
Aggregation-29	201	87	55	4	1	657	253	283	285	90
Aggregation-30	406	128	76	9	1	1248	431	679	886	313
Aggregation-31	720	281	161	28	7	2385	1015	1954	4113	1691
Aggregation-32	1571	646	366	45	9	5140	2200	3129	3886	2690
Aggregation-33	2145	915	558	87	13	6898	3049	5087	6368	2674
Aggregation-34	846	324	207	26	3	2813	1055	1711	2366	642
Aggregation-35	846	403	314	38	9	3184	1439	3085	3628	601
Aggregation-36	325	111	62	7	2	989	346	637	863	670
Aggregation-37	539	193	113	15	1	1721	723	949	1037	179
Aggregation-38	968	361	197	25	5	3234	1133	1503	2217	4343
Aggregation-39	568	248	151	19	3	1847	770	1210	1660	524

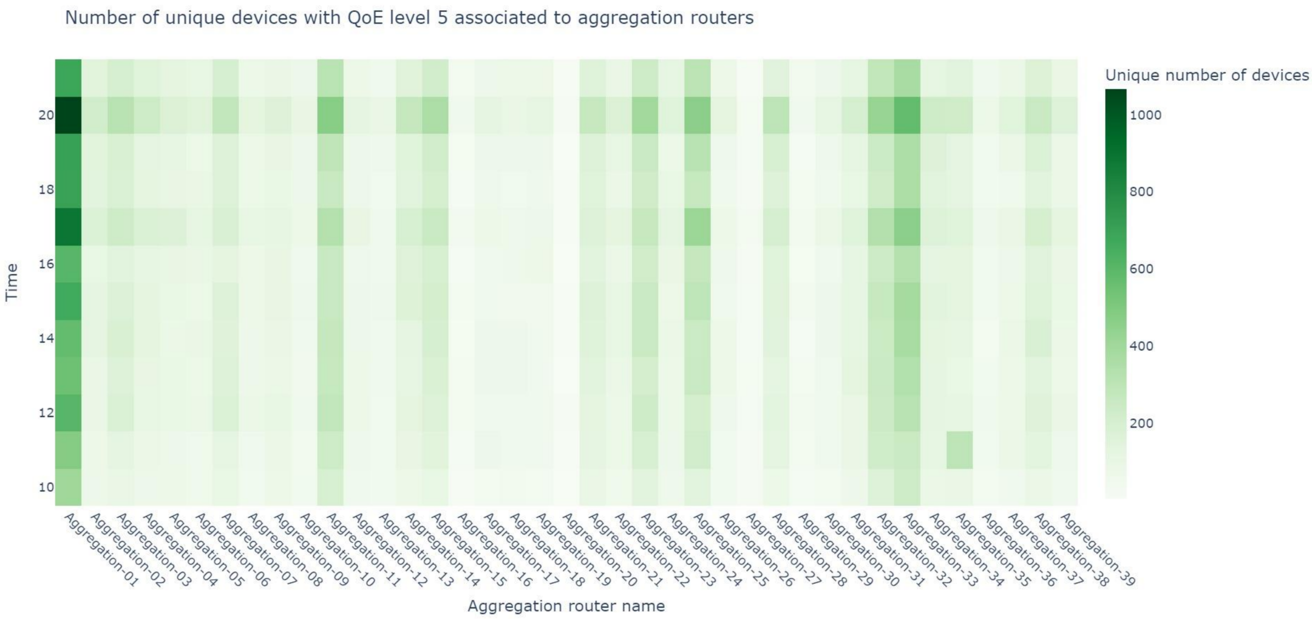


Figure 13. Heatmap with the unique count of STBs with the assigned Excellent (5) cluster. *X*-axis-aggregation router, *Y*-axis-time span (hours), *Z*-axis-count of unique devices connected to the aggregation router.

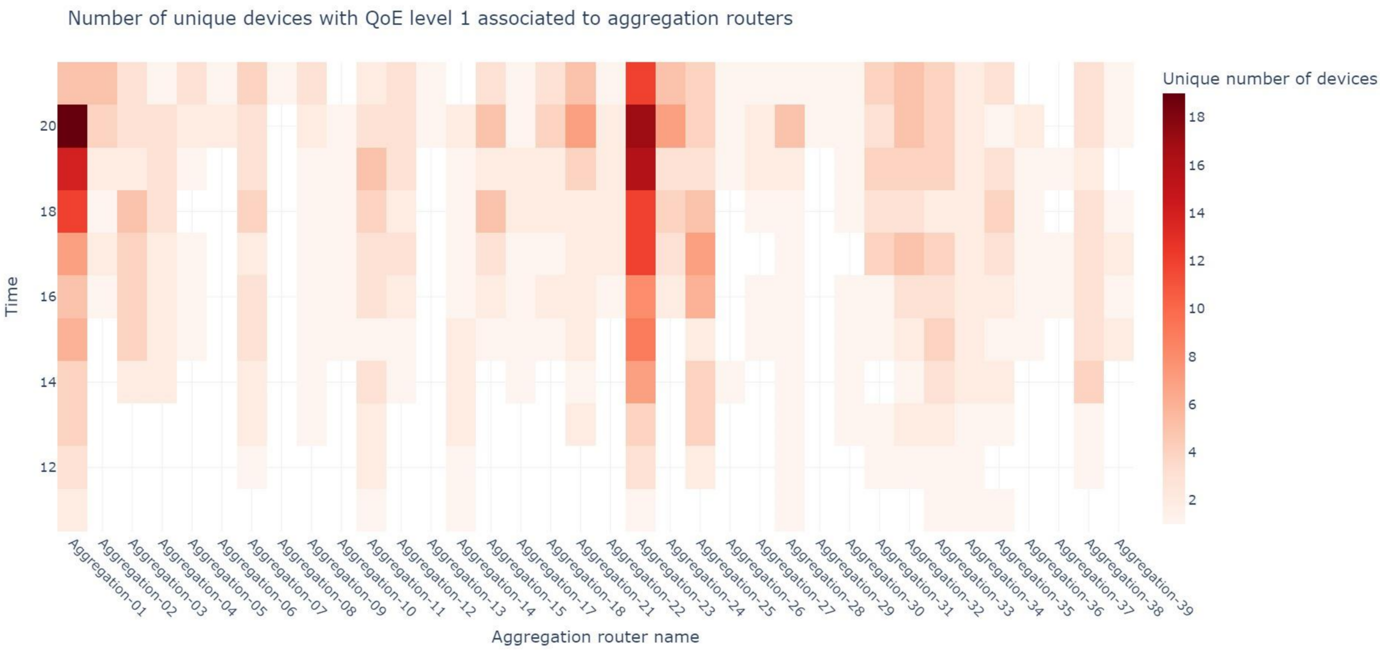


Figure 14. Heatmap with the unique count of STBs with the assigned Bad (1) cluster. *X*-axis-aggregation router, *Y*-axis-time span (hours), *Z*-axis-count of unique devices connected to the aggregation router.

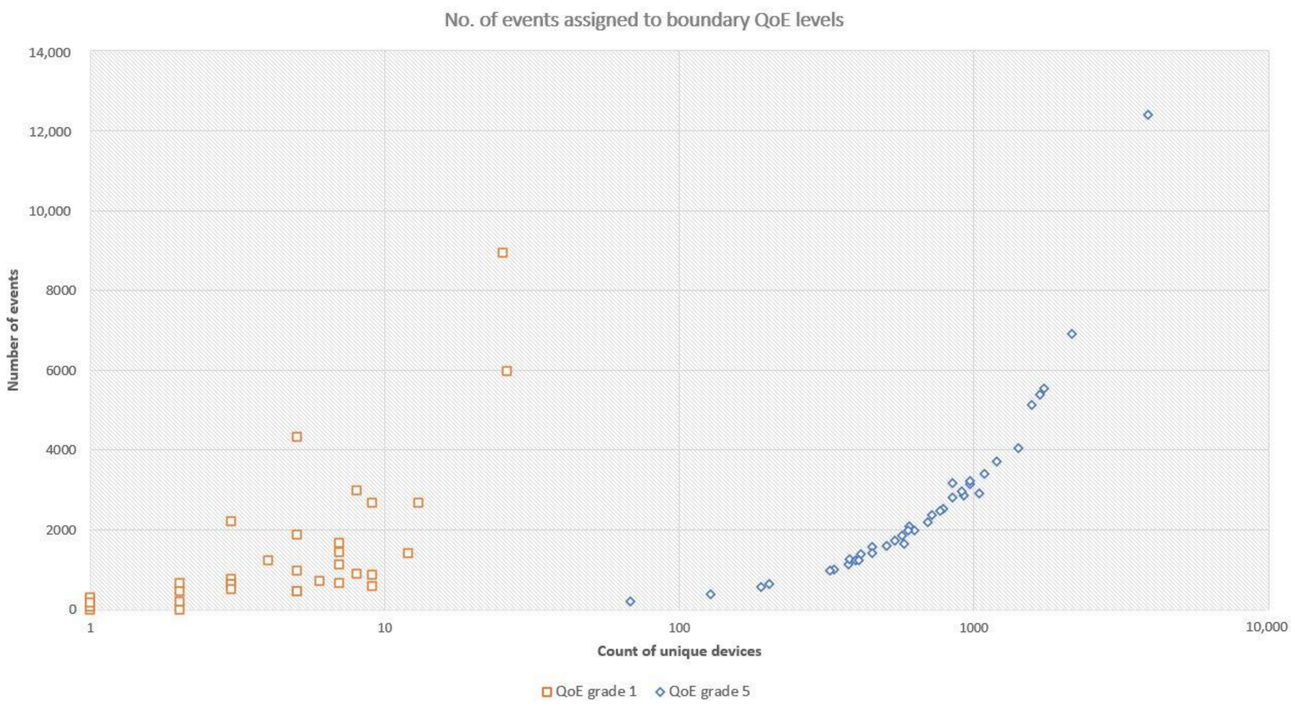


Figure 15. The number of events assigned to grade 1 (orange) or grade 5 (blue).

Devices with impairment are mainly located within Greater Poland voivodeship (Figure 16). Over 20 devices are working in Poznan or suburban rings. A device placement map can be used to send the fitters assigned to specific areas to deploy proactive monitoring solutions. The main goal ought to be to eradicate problematic devices and increase customer satisfaction as a result.

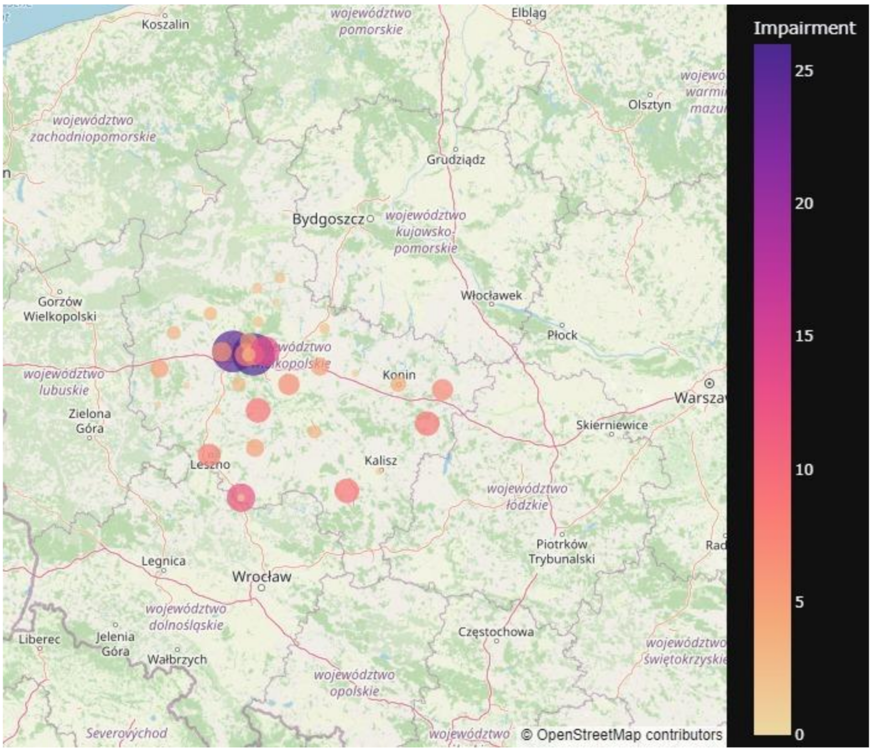


Figure 16. Localization of devices with increased impairment.

4. Discussion

The QoE assessment for television devices that uses adaptive streaming technology might be similar to algorithms that assess pure OTT applications. Continuous clustering can be used to list devices with a decreased level of video playback continuously. The STB devices work similar to OTT applications, but the main difference is that there is no possibility to set the intermediate level of bitrate to omit the bitrate switches. The STB player does not have the button to set the quality level, usually available in pure over-the-top applications. QoE assessment was made on a TV system that uses Linux-based STBs. However, the solution can be applied to other TV platforms based on Android or other software devices. Currently used STBs will always tend to play the highest possible bitrate. Hence the bitrate switches will occur in the impaired network environment. Grade 1 or 2 can be considered a significant decrease in the QoE level. The number of devices with impairments is relatively low. The elimination of problems affecting those devices should be considered the highest priority for TV operators due to decreased customer satisfaction. Proactive monitoring based on the proposed solution will be an asset in increasing the television video experience. The VSBCT parameter is highly correlated with the outcoming cluster. The SCI and SCTI parameters might not be meaningful since STBs always try to play the highest bitrate, and several fluctuations appear during the playback. The fluctuations trigger the events with positive and negative SCI and SCTI interchangeably.

The limitations of the work are connected to the data-driven approach. It is necessary to generate logs that describe the current level of bitrate and the occurrence of stallings either as bitrate with 0 kbps or separate logs. The proposed system cannot infer this based on the player's buffer state or network parameters from the ONT's perspective. If the pure OTT applications are available on the set-top box, it may not be possible to assess the QoE of content served by providers such as YouTube, Netflix, Disney, or others due to the lack of logs on the system associated with applications that complement TV features. However, the proposed approach allows real-time QoE assessment for many connected devices playing content related to linear channels or nPVR, catchup, and timeshift functions offered by the TV provider.

Another limitation is that it is hard to compare with related work at the current research stage due to the limitations regarding gathering data from customers. Mentioned algorithms in the introduction are frequently compared with RMSE, SROCC, or PCC parameters. However, preparing the assessment gathered from TV system users is in progress. Unfortunately, it is a long process since the data will be gathered from the customers calling the call center or second line of support and compared with the output returned by the QoE assessment system. Along with data collection, the assessment system may be improved, and the outcome will be compared with the current and future state-of-the-art works.

5. Conclusions

The paper proposes an unsupervised learning approach that utilizes continuous clustering to assess the QoE level of video services the TV operator delivers to the customers' households. The estimation is based on the logs that signalize the various aspects of STB's behavior. Data related to the bitstream level, stall occurrences, viewership, and the number of active devices have been reforged into SCI, SCTI, STCSI, and VSBCT features used during the clustering process. The ACR scale was used to mark the QoE level on the STBs. Grade 1 and 2 determined the worst video quality while the 4 or 5 pointed mainly the sightless degradations. Gathered assessments were applied to search the localization with the highest impairment level. The proposed approach might be effectively used by the TV operators to proactively eliminate the weak points in the network to overcome the increased degradation and, as a result, increase the customers' contentment.

In future works, the outcoming clusters should be challenged with the customer feelings to find a better threshold between the positive and negative clusters concerning TV service. The adjustment of the clustering algorithm might be a subject of future work due to the rapid increase of data stream instances related to TV system development. The infrastructure expansion leads to utilized device expansion. The QoE measurements may be a key indicator in assessing the network performance regards the video services.

To the best of our knowledge, this is the first study that utilizes continuous clustering with reclustering methods to assess the QoE level of devices connected to production television systems based on adaptive streaming technology.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app12168288/s1.

Author Contributions: Conceptualization, P.A., K.K. and P.Z.; methodology, P.A., K.K. and P.Z.; software, P.A.; validation, P.A., K.K. and P.Z.; formal analysis, K.K. and P.Z.; investigation, P.A., K.K. and P.Z.; resources, P.A.; data curation, P.A.; writing—original draft preparation, P.A.; writing—review and editing, P.A., K.K. and P.Z.; visualization, 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.

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

References

1. Cisco. *Cisco Visual Networking Index: Forecast and Trends, 2017–2022*; Cisco: San Jose, CA, USA, 2019.
2. Kowalik, K.; Andruloniw, P.; Partyka, B.; Zwierzykowski, P. Telecom Operator's Approach to QoE. *J. Telecommun. Inf. Technol.* **2022**, *2*, 26–34. [CrossRef]
3. Rodríguez, D.Z.; Wang, Z.; Rosa, R.L.; Bressan, G. The impact of video-quality-level switching on user quality of experience in dynamic adaptive streaming over HTTP. *EURASIP J. Wirel. Commun. Netw.* **2014**, *2014*, 216. [CrossRef]
4. Nam, H.; Kim, K.-H.; Schulzrinne, H. QoE matters more than QoS: Why people stop watching cat videos. In Proceedings of the IEEE INFOCOM 2016—The 35th Annual IEEE International Conference on Computer Communications, San Francisco, CA, USA, 10–14 April 2016.
5. Wang, Q.; Dai, H.N.; Wu, D.; Xiao, H. Data analysis on video streaming QoE over mobile networks. *EURASIP J. Wirel. Commun. Netw.* **2018**, *2018*, 173. [CrossRef]
6. Christos, C.; Bampis, G. Feature-based prediction of streaming video QoE: Distortions, stalling, and memory. *Signal Process. Image Commun.* **2018**, *68*, 218–228.
7. Wei, A.; Song, H.; Katto, J. High-QoE DASH live streaming using reinforcement learning. In Proceedings of the 2021 IEEE/ACM 29th International Symposium on Quality of Service (IWQOS), Tokyo, Japan, 25–28 June 2021.
8. Duanmu, Z.; Zeng, K.; Ma, K.; Rehman, A.; Wang, Z. A Quality-of-Experience Index for Streaming Video. *IEEE J. Sel. Top. Signal Process.* **2017**, *11*, 154–166. [CrossRef]
9. Lava, I.d.; Ortega, R.B.; Cebollada, J.C.G. New objective QoE models for evaluating ABR algorithms in DASH. *Comput. Commun.* **2020**, *158*, 126–140.
10. Ghadiyaram, D.; Pan, J.; Bovik, A.C. Learning a Continuous-Time Streaming Video QoE Model. *IEEE Trans. Image Process.* **2018**, *27*, 2257–2271. [CrossRef]
11. Frnda, J.; Nedoma, J.; Martinek, R.; Friedrich, M. Predicting Perceptual Quality in Internet Television Based on Unsupervised Learning. *Symmetry* **2020**, *12*, 1535. [CrossRef]
12. 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]
13. Zou, W.; Yang, F.; Song, J.; Wan, S.; Zhang, W.; Wu, H.R. Event-Based Perceptual Quality Assessment for HTTP-Based Video Streaming with Playback Interruption. *IEEE Trans. Multimed.* **2018**, *20*, 1475–1488. [CrossRef]
14. Tang, S.; Qin, X.; Wei, G. Network-Based Video Quality Assessment for Encrypted HTTP Adaptive Streaming. *IEEE Access* **2018**, *6*, 56246–56257. [CrossRef]
15. Yuan, Y.; Wang, C. IPTV video quality assessment model, based on neural network. *J. Vis. Commun. Image Represent.* **2019**, *64*, 102629. [CrossRef]
16. Du, L.; Zhuo, L.; Li, J.; Zhang, J.; Li, X.; Zhang, H. Video Quality of Experience Metric for Dynamic Adaptive Streaming Services Using DASH Standard and Deep Spatial-Temporal Representation of Video. *Appl. Sci.* **2020**, *10*, 1793. [CrossRef]
17. Loh, F.; Poignée, F.; Wamser, F.; Leidinger, F.; Hoßfeld, T. Uplink vs. Downlink: Machine Learning-Based Quality Prediction for HTTP Adaptive Video Streaming. *Sensors* **2021**, *21*, 4172. [CrossRef] [PubMed]
18. Frnda, J.; Nedoma, J.; Vanus, J.; Martinek, R. A hybrid QoS-QoE Estimation System for IPTV Service. *Electronics* **2019**, *8*, 585. [CrossRef]
19. Zhang, H.; Li, F.; Yan, Z. Feature fusion quality assessment model for DASH video streaming. *IET Image Process.* **2021**, *15*, 127–142. [CrossRef]
20. Tran, H.T.T.; Nguyen, D.V.; Ngoc, N.P.; Thang, T.C. Overall Quality Prediction for HTTP Adaptive Streaming Using LSTM Network. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *31*, 3212–3226. [CrossRef]
21. Dinaki, H.E.; Shirmohammadi, S.; Janulewicz, E.; Côté, D. Forecasting Video QoE with Deep Learning From Multivariate Time-Series. *IEEE Open J. Signal Process.* **2021**, *2*, 512–521. [CrossRef]
22. ISO. *ISO/IEC 23009–1:2019 Information Technology—Dynamic Adaptive Streaming Over HTTP (DASH)*; ISO/IEC JTC 1/SC 29; ISO: Geneva, Switzerland, 2019.
23. Pantos, R.; May, W. *HTTP Live Streaming. Internet Engineering Task Force*; Apple Inc.: Cupertino, CA, USA, 2017.
24. Microsoft. *[MS-SSTR]: Smooth Streaming Protocol*; Microsoft: Redmond, WA, USA, 2020.
25. ITU-T. P.910: Subjective Video Quality Assessment Methods for Multimedia Applications. October 2021. Available online: https://www.itu.int/rec/T-REC-P.910 (accessed on 30 June 2022).
26. Dziadul, C. Inea Launches New TV Platform. 12 February 2019. Available online: https://www.broadbandtvnews.com/2019/02/12/inea-launches-new-tv-platform/ (accessed on 18 June 2022).
27. Dziadul, C. Poland's Inea Joins Global Broadband Elite. 14 October 2018. Available online: https://www.broadbandtvnews.com/2018/10/14/polands-inea-joins-global-broadband-elite/ (accessed on 28 June 2022).
28. Huawei. *Requirements of Mobile Bearer Network for Mobile Video Service*; Huawei: Shenzhen, China, 2016.
29. Zubaroğlu, A.; Atalay, V. Data Stream Clustering: A Review. *Artif. Intell. Rev.* **2020**, *54*, 1201–1236. [CrossRef]
30. Silva, J.A.; Faria, E.R.; Barros, R.C.; Hruschka, E.R.; de Carvalho, A.C.; Gama, J. Data stream clustering: A survey. *ACM Comput. Surv.* **2013**, *46*, 1–31. [CrossRef]
31. Carnein, M.; Trautmann, H. evoStream—Evolutionary Stream Clustering Utilizing Idle Times. *Big Data Res.* **2018**, *14*, 101–111. [CrossRef]

32. Hahsler, M.; Bolaños, M. Clustering Data Streams Based on Shared Density between Micro-Clusters. *IEEE Trans. Knowl. Data Eng.* **2016**, *28*, 1449–1461. [CrossRef]
33. Chen, Y.; Tu, L. Density-based clustering for real-time stream data. In Proceedings of the KDD '07: 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, CA, USA, 13–15 August 2007; pp. 133–142.
34. Carnein, M.; Assenmacher, D.; Trautmann, H. An empirical comparison of stream clustering algorithms. In Proceedings of the CF'17: Proceedings of the Computing Frontiers Conference, New York, NY, USA, 15–17 May 2017; pp. 361–366.
35. Arthur, D.; Vassilvitskii, S. k-means++: The advantages of careful seeding. In Proceedings of the SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, New Orleans, LA, USA, 7–9 January 2007; pp. 1027–1035.