Rebuffering but not Suffering: Exploring Continuous-Time Quantitative QoE by User's Exiting Behaviors

Rebuffering but not Suffering: Exploring Continuous-Time Quantitative QoE by User's Exiting Behaviors

Sheng Cheng, Han Hu, Xinggong Zhang*, Zongming Guo
Peking University
Email: {chaser_wind, huhan951753, zhangxg, guozongming}@pku.edu.cn

***Abstract*—Quality of Experience (QoE) is one of the most important quality indicators for video streaming applications. But it is still an open question how to assess QoE value objectively and quantitatively over continuous time both for academia and industry. In this paper, we carry out an extensive data study on user behaviors in one of the largest short-video service providers. The measurement data reveals that the user's exiting behavior in viewing video streams is an appropriate choice as a continuous-time QoE metric. Secondly, we build a quantitative QoE model to objectively assess the quality of video playback by discretizing the playback session into the Markov chain. By collecting 7 billion viewing session logs which cover users from 20 CDN providers and 40 Internet service providers, the proposed state-chain-based model of State-Exiting Ratio (SER) is validated. The experimental results show that the modeling error of SER and session duration are less than 2% and 10s respectively. By using the proposed scheme to optimize adaptive video streaming, the average session duration is improved up to 60% to baseline, and 20% to the existing black-box-like machine learning methods.**

I. Introduction

With the popularity of online video applications, it has become a critical issue to assess the quality of network video streaming [1]. Widely used metrics, such as packet loss, delay, etc. in the network layer are not sufficient to represent the quality of applications [2], [3], [4]. Quality of experience (QoE), defined by ITU-T [5], is sounded as "the overall acceptability of an application or service, as **perceived** subjectively by the end-user". However, due to the subjective character, QoE is impacted by network, environment, video content, user subjective psychology etc [6], [7]. It is also infeasible to survey users' opinions on the quality of large-scale online video services [8], [9]. For academia and industry, it is still a grand challenge to measure the QoE of video streaming in practice [10].

Continuous-time objective quantitative QoE metric. Our vision is to provide a continuous-time objective and quantitative tool to measure the QoE of video streaming. It seems to be an impossible task since QoE is a subjective metric and it is hard to assess QoE at all time points. The first challenge *objective* means it can be measured without an explicit user opinion survey. The second challenge *quantitative* indicates QoE could be scored numerically with one single scalar. The third challenge *continuous-time* means QoE shall be able to be measured at any time point during the playback.

Many previous efforts have been devoted to developing objective QoE metrics. Most of the researchers and online video service providers tend to use application layer metrics [11], [12], [13], such as startup delay, rebuffering time and bitrate. Of course, they are objective and easy to be measured from viewing logs [14]. However, these **application layer metrics cannot truly meet the definition of QoE** which requires the **perception** process. Moreover, their joint impacts on QoE are implicit [15]. Other works [16], [17], [11] try to build subjective models mapping these metrics to user's quality opinion. But they are not objective which makes it skeptical to be used.

Among user engagement metrics, exiting behavior (also called abandonment in previous research) is a promising potential QoE metric. User engagement metrics are more favored than application layer QoE metrics due to their stronger relationship with user satisfaction [18]. Among user engagement metrics, exiting is a kind of user behavior that can be captured objectively, and more importantly, it can happen at any time point of a playback session. Consequently, the exiting behavior is a potential QoE metric for video streaming for its continuous-time features which are not owned by other traditional user engagement metrics like play ratio [19] and session duration [20], [21], [22].

There are several challenges in building a QoE model based on user's exiting behavior: 1) how to deal with non-streaming factors which could also impact user engagement [23], 2) how to establish a readily usable analytic model to predict exiting behavior given an unfinished (may be very long or very short) video session with playing and stalling events and 3) how to assign QoE scores for diverse network contexts and further optimize QoE during video streaming.

To address these problems above, this paper proposes **ExQoE**, a Markov-chain-based exiting-behavior model, which assumes that the user's exiting at any time point is a probability event and depends on the historical playback pattern, i.e. the transition between the events of *playing* and *stalling*. More *playing* events increase users' experience, consequently reducing the probability of users' exiting behavior. Therefore, the user's exiting ratio is considered an explicit QoE metric which is modeled as a function of playback state transitions.

In summary, ExQoE handles technical challenges in model user's exiting ratio by:

- The statistical method that quantifies the user's exiting behavior. We propose a quantification scheme of exiting behavior that shields the effects of non-streaming factors by randomly sampling and calculating the exiting ratio over these factors.

*Dr. Xinggong Zhang is the corresponding author.

- The intrinsic mechanism about how playback pattern impacts on exiting ratio. We propose a Markov-chain model to transform playback session samples into Markov chains, and reveal the insight from extensive data studying (Section IV).
- How to score and further optimize the quality of the network using the proposed QoE metric. We propose a probabilistic method to depict state transitions caused by network contexts and a layered implementation of QoE optimization (Section V).

We have carried out extensive experiments with large datasets of short videos collected from Douyin [24][1]. The results prove that our proposed scheme can accurately predict users' exiting ratios and session durations. For network contexts such as access type, location, CDN and resolution, the modeling error of exiting ratio and session duration are less than 2% and 10s respectively. By using the proposed QoE score to optimize CDN and resolution decisions in streaming, the average session duration is improved up to 60% to baseline, and 20% to the existing black-box-like machine learning methods.

II. Motivations and Insights

In this section, we will summarize the shortcomings and problems of previous works about QoE, and provide some ideas on how to solve these problems, thus introducing the basic idea of our QoE modeling.

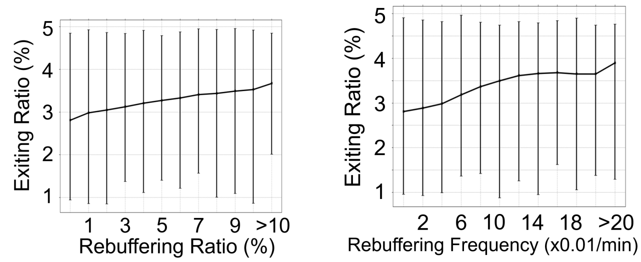
A. Problem 1: Selecting Metric for QoE Measurement

Ideally, QoE metric shall be **objective**, which means it can be measured objectively without explicit inquiry, **quantitative**, which means it can be represented by a real number and is convenient for optimization, and **continuous-time**, which means it can be assessed at any point of time during video playback.

There are quantities of works about improving the quality of video streaming, with various metrics set as optimization objectives, from network parameters like throughput or packet loss ratio [25], [26] to playback metrics such as video bitrate and rebuffering events [27], [28]. Nevertheless, what all these metrics reflect is the quality of video playback provided by the video player, not the true quality that user **perceives**. So these metrics can not meet the definition of QoE, and it is not sure whether these optimization operations can truly maximize QoE [29]. On the contrary, only the metrics about **user engagement** can exactly represent QoE [18]. For instance, the longer duration of a view session certainly means that the viewer is more desired to continue watching [30], hence better user's QoE [6]. Play ratio (proportion of the played duration to video length) can also reflect users' satisfaction with what they have watched [19]. However, these user engagement metrics do not fulfill the requirement of continuous time.

Solution: exiting behavior. Among user engagement metrics, the user's exiting behavior is a promising QoE candidate,

[1]Chinese version of TikTok.



(a) Rebuffering Ratio (b) Rebuffering Frequency

Fig. 1. The influence of rebuffering events on exiting ratio. There is only ambiguous correlation, without any precise quantitative relationship.

which meets the **objective** and **continuous-time** conditions of QoE. Users' exiting behavior can happen at any time during video playback and can be captured objectively and easily. Intuitively, exiting behavior indicates user dissatisfaction, hence the potential QoE metric. However, owing to the lack of quantification, exiting itself is rarely adopted as a QoE metric directly, but its derivative metrics are common in use. Session duration is a derivative metric of exiting that quantifies exiting behavior most simply but loses the feature of continuous time. Session duration is not continuous time because we can not measure the duration of an unfinished session, and we can only measure it at the end of a session. Session duration has been studied as QoE metrics recently [20], [21], [22].

B. Problem 2: Non-streaming factors.

User engagement is not only determined by streaming quality but also affected by many other factors like the user's mood, the user's preference for the video content and even the user's time planning at that time. Recent research have proved that user engagement is also influenced by other factors beyond streaming quality [23]. Among these non-streaming factors, some of them are not easy to fully explore like content preference, while some of them may be nearly infeasible to study like time planning. Therefore, we need a way to properly deal with non-streaming factors that impact users' exiting behavior.

Solution: statistical quantification with random sampling. It is trivial to quantify exiting behavior with a statistical method to ensure the three conditions of the QoE metric. After collecting quantities of samples, we will be able to calculate the probability of the user's exiting (also named as "exiting ratio") at any time according to the Law of Large Numbers. More importantly, we could conduct random sampling over non-streaming factors and calculate their average exiting ratio. In this way, the effects of non-streaming factors are considered at the very beginning of our modeling and we do not need extra considerations for them in the modeling process of our QoE model.

C. Problem 3: Correlation between Exiting and Rebuffering

Having chosen the exiting ratio as the objective metric for QoE measurement, the remained problem is how to analyze the influencing factors, especially the rebuffering events [31],

[32], [33]. After long explorations, some work found that application layer metrics like rebuffering events influence viewer engagement. For example, a higher ratio and frequency of rebuffering events will lead to a lower video play ratio [19].

However, **statistical features about rebuffering may not be sufficient to accurately predict exiting**. We conduct data analysis to study the correlation between exiting ratio and statistical rebuffering features. For each batch of sessions longer than 61 seconds, we calculate the ratio of the sessions that end in the 61st second and the statistical rebuffering features in the very first 60s. By repeated sampling batches from the database, we reached the results in figure 1. As is shown in figure 1, the influence of rebuffering events on the exiting ratio cannot be accurately measured due to large variance. There is certainly a correlation between rebuffering events and the exiting ratio, but the quantitative relationship is not precise.

Solution: discretization over time. Discretization over time is the best method of solving the problem above. In most of the previous works, rebuffering events are summarized as statistical metrics for a whole session. However, the influence of rebuffering events may reflect in smaller units. Ricky et al. [12] found that users tend to pause the playback around 2 seconds after the rebuffering event happens, which inspires us to divide the video playback into consecutive units to analyze their effects in detail.

D. Insights

Given the problems and corresponding results above, two insights can be concluded:

- Statistical quantification with random sampling over non-streaming factors could help us study exiting behavior numerically without interference.
- Traditional statistical metrics like rebuffering ratio and frequency are insufficient to accurately predict exiting behavior.

On the whole, we offer two ideas according to the insights: 1) choosing the average exiting ratio as the continuous-time objective quantitative metric for QoE measurement, and 2) establishing a QoE model by transforming the video sessions into state chains.

III. System Overview

We have designed **ExQoE**, an exiting-ratio-based QoE modeling, scoring and optimizing system, which can precisely predict the exiting ratio as the measurement of QoE, scoring different network contexts by deduced metric session duration, and optimize QoE by selecting the most appropriate CDN and resolution level. Figure 2 shows the diagram of **ExQoE**.

QoE modeling: By transforming the video playback sessions into state chains, it is possible to measure QoE for every time step and further estimate session durations.

The process of QoE modeling has three steps. At first, we divide the sessions into units with unified length (1 second is chosen in this paper), and determine the state of each unit according to its playback smoothness, thus transforming the

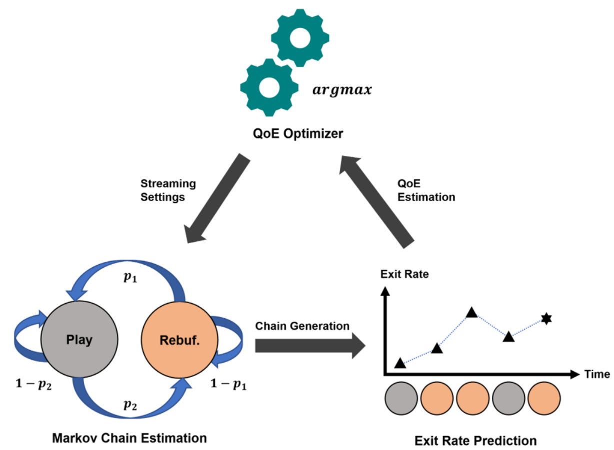


Fig. 2. The diagram of **ExQoE**. It mainly has three modules: QoE modeling, QoE scoring and QoE optimization.

sessions into state chains. Secondly, for each pattern of state chains, we calculate viewers' tendency to exit at each time step, which represents their real-time dissatisfaction. Finally, since the exit action is the signal of ending a session, we can calculate the mathematical expectations of session duration for each pattern of state chains. For industrial applications, session duration is a commonly used Key Performance Indicator (KPI), so we take a further step from exiting ratio to session duration.

QoE scoring & optimization: We focus on the network contexts, assigning QoE scores for them by calculating the expectations of session duration. Based on QoE scores, we conduct optimization by selecting the best CDN and resolution level.

The process of QoE scoring also consists of three steps. It starts with calculating state transition probabilities for different network contexts. Then we can estimate the occurrence probability of each state chain pattern. Finally, we can calculate the weighted expectations of session duration for network contexts. In this way, for a given user, the QoE scores for different network contexts can be obtained.

In the process of QoE optimization, we use five features to classify different network contexts: access type, location, CDN, resolution level and time. All of the features influence state transition probabilities, while other factors are proved unnecessary to be considered. Among these features, CDN and resolution level is changeable for a given user. We apply the ϵ-greedy bandit algorithm to make optimal decisions.

IV. QoE Modeling

In this section, we present the details of QoE modeling and reveal how we solve the problems in section II.

A. Discretization of Playback Session

There are quantities of metrics for us to describe and measure QoE, but these metrics are still not enough for us to accurately measure QoE in time. Figure 1 has already revealed that it is difficult to make quantitative analysis only based on

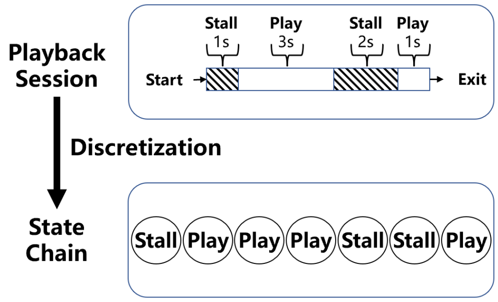


Fig. 3. The playback sessions are transformed into state chains.

the statistical metrics of a whole session due to large variances. Therefore, it is necessary to provide a better QoE modeling scheme.

Discretization of playback session. As a result, we decide to split the sessions into smaller units to analyze the state of playback and users' action at each time step, as is shown in figure 3. By transforming the playback sessions into state chains, we can find the recent states after which users are most likely to exit to accurately figure out how startup delay and rebuffering events influence QoE. For any playback session sample, we can discretize it into a state chain:

$$\boldsymbol{c} = < s_0, s_1, s_2, \ldots, s_t, \ldots > \tag{1}$$

where $s_t \in \{Play, Stall\}$ is the state of time step t.

Definition of state. The length of a state unit is 1 second in this paper. Owing to that rebuffering events are relatively rare for most of the playback samples, we define $s_t = Play$ if there is less than 0.3-second rebuffering in time step t, otherwise $s_t = Stall$.

It should be emphasized that s_t only represents the objective state of playback streaming, while users' actions are not included. Therefore, the states are only influenced by network qualities, while subjective exits of playback sessions are controlled by the users themselves.

B. Measurement Study on Exiting Ratio

Session duration is one of the most important metrics that measure a user's feeling about watching streaming video, and the active exit action, as the signal of the end of session duration, clearly reflects the user's dissatisfaction. Therefore, if we figure out the situations in which users will choose to exit, we can understand how QoE is influenced.

Definition of state-exiting ratio (SER). User may be disturbed by startup delay and rebuffering events, so session duration will vary with the pattern of the state chain. To analyze the tendency of giving up watching, for any batch of prefix chain samples (for every two chains, the short one is a prefix of the long one.) denoted as set S, we define SER to quantify users' instant probability of exit at each time step $t > 1$:

$$SER_t = \frac{|S_{t-1}| - |S_t|}{|S_{t-1}|} \tag{2}$$

where S_t and S_{t-1} are the subsets of S which contain samples longer than t and $t-1$ seconds respectively. Formally, $S = S_1 \supseteq S_2 \supseteq \ldots \supseteq S_t \supseteq \ldots$. In the remainder of this section, it is always implicitly assured that $t > 1$ if t is present in an equation, and we omit it for simplicity.

Analysis of SER. To figure out the factors that influence SER, it is natural to analyze the relation between SERs of adjacent time steps. We have carried out an extensive measurement study of a commercial online short video service provider for 6 months and collected 7 billion viewing session logs which cover users from 20 CDN providers and 40 different network operators.

We calculated SERs of each pair of adjacent time steps of over 120 batches with different patterns of state chains and scattered the results in figure 4. The data points are divided into 4 parts by the type of state transition, since the states of playback may reverse users' desire of watching. It can be spotted that when the type of state transition is fixed, SERs of adjacent time steps will show an obvious linear correlation.

C. Modeling Exiting Ratio by State Transition

Before modeling, we heuristically believe that SER is continuous over time and the SER variation is determined by the most recent video play state transition. So we explore the correlation between the SERs of adjacent time steps and the state transitions. Figure 4 reveals that SERs of adjacent time steps have linear correlation regardless of time. Based on this finding, the following formula can be summarized temporarily:

$$SER_t = k\left(s_{t-1}, s_t\right) SER_{t-1} + b\left(s_{t-1}, s_t\right) \tag{3}$$

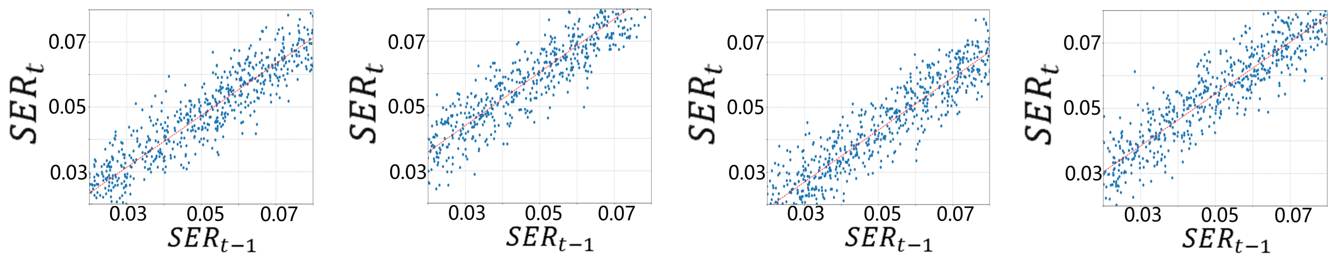
where (s_{t-1}, s_t) stands for the latest state transition, with four values in total. k and b represent the slope and the intercept of linear correlation.

Analysis of *k*. To further determine the parameters, we have traversed the dataset and calculated the values for each batch of $100K$ samples. Firstly, for k, the results are shown in figure 5. It can be seen that whichever the type of state transition is, the value ranges are relatively fixed for k. Almost all values of k are between 0.6 and 0.8, and over 80% of values fall between 0.7 and 0.8.

Analysis of *b*. As for b, the results are shown in figure 6. It can be spotted that the value ranges are still very small. For $b(Play, Play)$, all the values are between 0.005 and 0.007. $b(Play, Stall)$ has the maximum lower bound, with about 80% of values falling between 0.020 and 0.022. For $b(Stall, Play)$, the values are the smallest among the 4 types of state transitions, and over 90% of values fall in the range of 0 and 0.005. Finally, for $b(Stall, Stall)$, although it has the maximum value range, over 90% of the values are between 0.012 and 0.015. However, b changes with the type of state transition.

Therefore, we can approximately determine that the value of k is fixed as γ whichever the type of state transition is, while the value of b is fixed only when the type of state transition is certain, then the formula can be further simplified as:

$$SER_t = \gamma SER_{t-1} + b\left(s_{t-1}, s_t\right) \tag{4}$$



(a) Play-Play (b) Play-Stall (c) Stall-Play (d) Stall-Stall

Fig. 4. Linear correlation of SERs between adjacent time steps. For each type of state transition, SERs of adjacent time steps show an obvious linear correlation

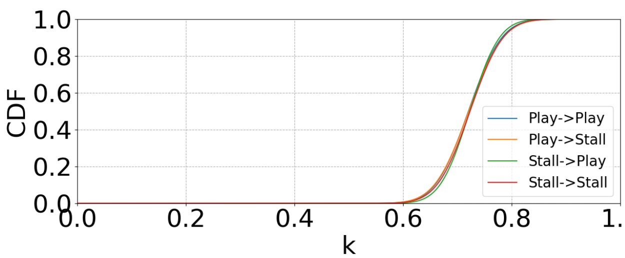


Fig. 5. As for k, whichever the type of state transition is, the value ranges are small and relatively fixed.

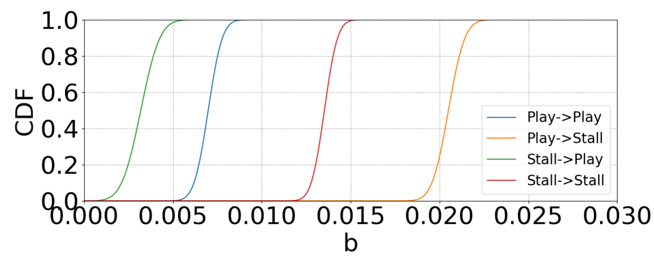


Fig. 6. As for b, the value ranges are still small, but they change with the type of state transition.

Parameter values. Directly using the linear regression method over the whole dataset, we calculate the precise values of parameters, which are listed in table I.

D. *QoE by State-Exiting Ratio*

Based on the findings above, we can finally derive the formula of QoE.

For any state chain $\mathbf{c} =< s_0, s_1, s_2, \ldots, s_t, \ldots >$, SER will be calculated as:

$$SER_t = \gamma SER_{t-1} + b\left(s_{t-1}, s_t\right) = \gamma^n SER_{t-n} + \sum_{i=1}^{n} \gamma^{i-1} b\left(s_{t-i}, s_{t-i+1}\right) \quad (5)$$

where $n \leq t$.

When n and t are very large, since $\gamma < 1$, the above formula can be approximated as:

$$SER_t = \sum_{i=1}^{n} \gamma^{i-1} b\left(s_{t-i}, s_{t-i+1}\right) \quad (6)$$

which means that SER is only influenced by the nearest several state transitions.

Mathematical expectation of session duration. With the measurement of SER at each time step, we can already quantify users' QoE in time. Based on the results above, it is practicable to calculate the mathematical expectation of session duration for each pattern of state chains:

$$E(\boldsymbol{c}) = \lim_{t\to\infty} \sum_{i=1}^{t} [i \cdot SER_i \cdot \prod_{j=1}^{i-1} (1 - SER_j)] \quad (7)$$

TABLE I
VALUES OF PARAMETERS

	γ	b(Play-Play)	b(Play-Stall)	b(Stall-Play)	b(Stall-Stall)
Values	0.78833	0.00698	0.02050	0.00319	0.01352

V. QoE SCORING & OPTIMIZATION

Based on the Markov-chain-based QoE model, we need to further score the network contexts and conduct QoE optimization accordingly. Concretely, it is necessary to derive the relation between network contexts and state chains to directly predict session duration in any network environment and conversely select the most appropriate CDN and resolution level to obtain the best QoE.

A. *Transition Probability*

Network contexts have a significant influence on playback sessions. Low quality of service (QoS) will lead to higher startup delays and more rebuffering events. In another word, different network environments will change the occurrence probability of state $s_t \in \{Play, Stall\}$ at each time step of a chain.

Definition of transition probability. Similar to the previous definition of SER_t, we define P_T to quantify the transition probabilities between adjacent states:

$$P_T\left(i, j\right) = \frac{N\left(i, j\right)}{N\left(i\right)} \quad (8)$$

where $i, j \in \{Play, Stall\}$. Each time we randomly choose $100K$ samples to build a batch, denote the number of sessions with sub-sequence (i, j) as $N\left(i, j\right)$, and denote the number of sessions with state i as $N\left(i\right)$.

Analysis of network contexts. As for network contexts, there are lots of factors that influence network conditions, including Location, Access Type, CDN and Resolution Level. Figure 7 shows the correlation between transition probabilities and these factors. For any of these four metrics, when the value changes, the difference in each transition probability is quite obvious. The results reveal that transition probabilities certainly change with network contexts.

Analysis of time. As a matter of fact, the network conditions change with time and the state of the network may fluctuate frequently. According to figure 8, apart from the network metrics above, the value of each transition probability will also change with time.

Analysis of other factors. Apart from network contexts, there are also user- and video-related factors. Therefore, it is necessary to evaluate whether these features have a significant influence on transition probabilities.

Figure 9 reveals the impacts of subjective factors on transition probabilities. We set <Location=SH, CDN=Bd, Access Type=Wi-Fi, Resolution Level=720p> and collect session samples within 12 hours. It can be spotted that when the values of network contexts are fixed, subjective factors such as user clusters and video tags have no remarkable influence on transition probabilities. We guess there are two reasons. First, the selected data comes from mainstream short video platforms, and the recommendation algorithm ensures that users have a high preference for the content currently played in most cases; Secondly, under the influence of the previous factor, even if there are a small number of videos that are not favored by users, users are likely to skip the video directly by skipping, so the impact of the unwanted videos on the playing session is further reduced. Consequently, the impact of other factors is not discovered in our experiment yet.

B. QoE Scoring for Network Contexts

Calculation of occurrence probability. In the previous section, we have known the calculation of transition probabilities in a certain network environment. Then for any pattern of state chain $\boldsymbol{c} =< s_0, s_1, s_2, \ldots, s_t, \ldots >$, its occurrence probability can be defined as:

$$P_C(\boldsymbol{c}) = \lim_{t\to\infty} \prod_{i=1}^{t} P_T\left(s_{i-1}, s_i\right) \tag{9}$$

Mathematical expectation of session duration. The expectation of session duration for a given network environment u=<Location, Access type, CDN, Resolution level, Time> can be obtained by the weighted mean values of patterns:

$$E_{SD}(u) = \sum_{\boldsymbol{c}} E(\boldsymbol{c}) P_C(\boldsymbol{c}) \tag{10}$$

C. QoE Optimization

After deriving the function between network contexts and session duration, the last step of optimization is to select the best CDN and resolution level according to the QoE scores. Denote u as the immutable components of network contexts including Location, Access Type and Time. a indicates the mutable components of network contexts including CDN and Resolution Level. Then what we need is a policy function $\pi(a|u)$ to guide the selection of CDN and resolution level.

ϵ-greedy bandit algorithm. [34] The problem can be regarded as a reinforcement learning problem since the various network contexts form the environment to be explored, and the available CDNs and resolution levels are the actions to be exploited.

There have been many classical algorithms in the field of reinforcement learning. To deal with the possible error of estimation results and the possible congestion of CDN, we decide to apply the ϵ-greedy bandit algorithm. Then the optimization problem can be formed as:

$$A^* \leftarrow \arg\max_{a} Q(u, a) \tag{11}$$

while $a \in \{< CDN, Resolution >\}$. For all $a \in \mathcal{A}(u)$:

$$\pi(a \mid u) \leftarrow \begin{cases} 1 - \varepsilon + \varepsilon / |\mathcal{A}(u)| & \text{if } a = A^* \\ \varepsilon / |\mathcal{A}(u)| & \text{if } a \neq A^* \end{cases} \tag{12}$$

VI. Evaluation

To validate the performance of **ExQoE**, we have carried out an extensive measurement study on a commercial online short video service provider for nearly nine months. We have collected logs of 7 billion viewing sessions which cover users from 20 CDN providers and 40 Internet service providers. Among them, 3 billion logs are chosen randomly for estimating the parameters of SER and session-duration model in section IV-D. The rest 4 billion logs will be used as a validation dataset to validate the proposed QoE model and CDN-resolution decision algorithm in the following sections.

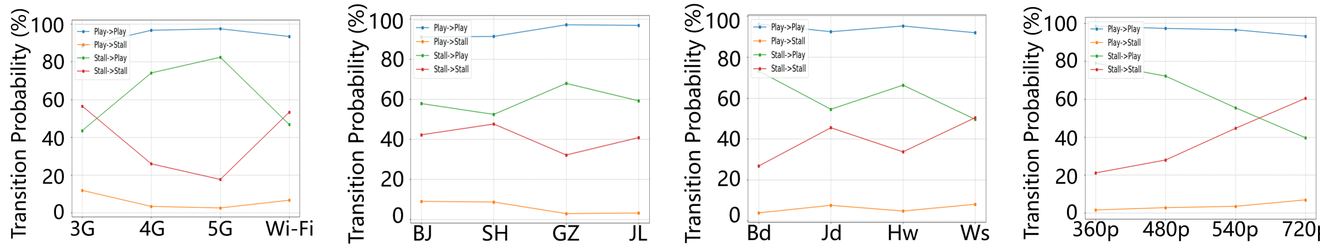
For a comprehensive evaluation, we have designed the following 4 parts of experiments.

Modeling SER by state-chain. To validate the accuracy of the QoE model by users' exiting behaviors, we measure SER on the validation dataset and compare it with the predicted value by SER model defined in section IV. The experimental results demonstrate the modeling error is less than 2% (exiting ratio). It also verified that the state-chain-based SER model is always held irrespective of various times, CDNs or video tags (Section VI-A).

QoE metrics by session-duration model. Session-duration model is also validated on the validation dataset. The predicted session durations are compared with those of video playback session samples. The results show that the duration error is less than 10s for more than 90% samples. The model is also validated on various datasets, such as rebuffering ratio, user cluster and video tag. The accuracy of the session-duration model is still high (Section VI-B).

QoE scoring on network contexts. The proposed QoE model is used to score network contexts, such as access type, location, CDN and resolution level. The QoE score is compared with the measured session duration across various network contexts. The results show that the modeling error of session duration is limited to less than 10s (Section VI-C).

Performance comparison with other algorithms. Compared with other machine learning algorithms, the proposed scheme



(a) Access Type (b) Location (c) CDN (d) Resolution Level

Fig. 7. The correlation between four network layer factors and transition probabilities. These factors can influence transition probabilities.

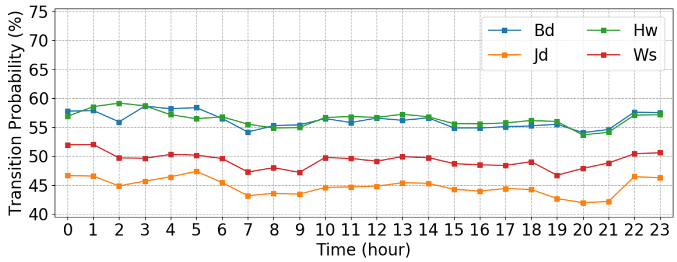
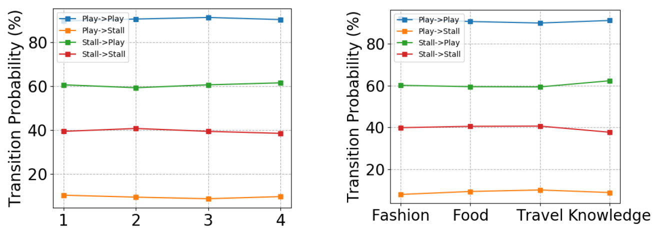


Fig. 8. The value of transition probability will change with time.



(a) User Cluster (b) Video Tag

Fig. 9. The influences of other features on transition probabilities.

can improve the average session duration up to 20%, with minimal time consumption (Section VI-D).

A. Validation of State-chain SER Model

The first part of our experiments is to evaluate the proposed state-chain-based QoE model of SER. Since SER represents the user's exiting ratio, only after a group of users passes the same pattern of state chain can we measure SER for each state. For each pattern, we randomly select $100K$ samples following the pattern and measure the real statistical SER values.

The modeling SER is directly predicted by the proposed SER model. For a specified state chain, each state's SER is calculated by equation.6 with the parameters in table I. The predicted and measured values of SER are illustrated in figure 10, for validating the accuracy of the Markov-chain-based SER model.

Overall accuracy of SER: The predicted and measured values are depicted on the scatter plot in figure 10(a). it can be seen that the prediction error is almost controlled within 2%. For most of the sample states, the differences between predicted and measured SERs are very small. For more than 80% of samples, the difference between the two values is less than 1% exiting ratio.

SER accuracy under various conditions: To validate whether the SER model only depends on the pattern of state chains, we also collected viewing samples under various times, CDNs and video tags. The prediction error is calculated by the difference between the measured value and modeling SER. The results are illustrated in figure 10(b), (c) and (d).

Figure 10(b) shows that the prediction error is slightly changed with time. Even for the period 8:00-16:00, it gets the worst prediction accuracy, but the errors are still less than 1% for almost 80% samples.

For different CDNs in figure 10(c), the SER model performs the best over the dataset of Hw, while the lowest accuracy is found among Jd users. However, for more than 95% of samples, the prediction error is less than 1.5%.

In figure 10(d), prediction error for different video tags is illustrated. The video tags represent the themes and main contents of videos. Figure 10(d) demonstrates the prediction error is less than 1% for 80% samples across video categories.

B. QoE Model by Session Duration

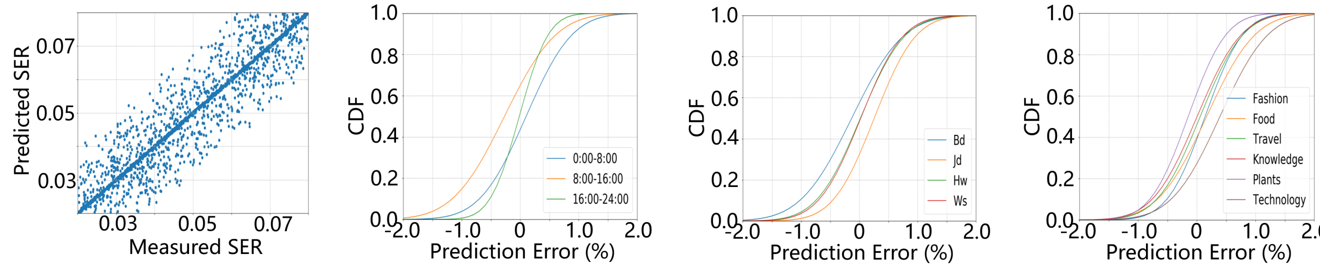
In this section, we evaluate our state-chain-based QoE model by session duration. Given a playback session sample, the session duration can be measured directly. For modeling value, we calculate SERs at each state, and estimate the mathematical expectation of session duration by equation.7. The results are demonstrated in figure 11.

Overall accuracy of session duration model: As is shown in figure 11(a), the differences between measured and predicted values are less than 10s. For more than 90% of samples, the prediction error is less than 10% of the measured value.

Model accuracy over various datasets: We have also evaluated the session-duration model over various datasets, which is divided by rebuffering ratio, user cluster, and video tag. The results are shown in figure 11(b), (c) and (d).

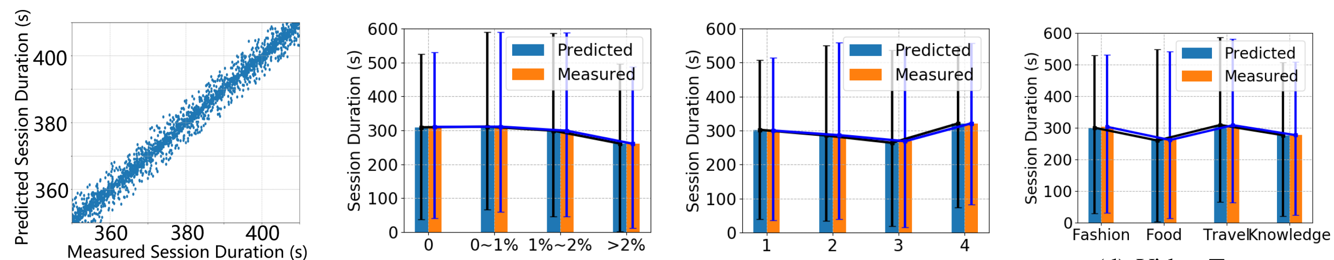
Figure 11(b) shows that the session duration is reduced with increasing rebuffering ratio, both for measured and predicted values. The differences between the predicted and measured values are limited to no more than 5%. This verifies that the model of session duration is accurate across datasets with different rebuffering ratios.

For various user clusters, figure 11(c) reveals that different users have various ranges of session duration in general. Possibly this phenomenon is caused by the different habits and preferences of users while watching short videos. Nevertheless, the prediction accuracy of session duration still maintains at a high level, proving that the model based on state chains is robust for any viewer.



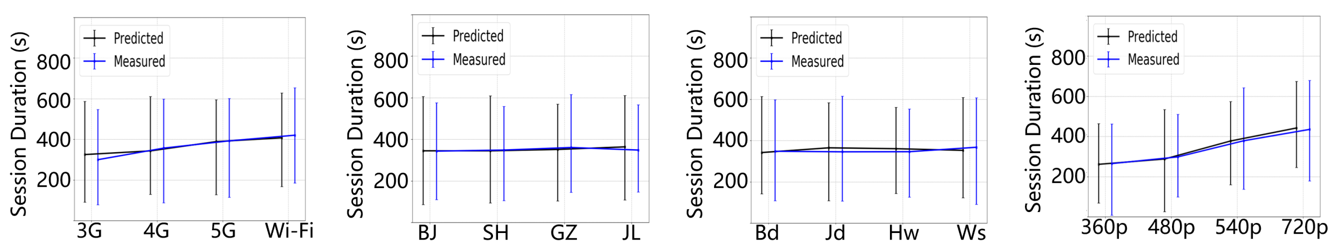
(a) Overall (b) Time (c) CDN (d) Video Tag

Fig. 10. Evaluation of State Exit Ratio Model. (a) plots overall scatter, showing that the differences between measured and predicted SERs are less than 2%. (b), (c) and (d) show the prediction error under different times, CDNs and video tags.



(a) Overall (b) Rebuffering Ratio (c) User Cluster (d) Video Tag

Fig. 11. Evaluation of session duration model. (a) plots overall scatter, showing that the prediction error of session duration is less than 5%. (b), (c) and (d) show the distributions of measured and predicted session durations under different rebuffering ratios, user clusters and video tags.



(a) Access Type (b) Location (c) CDN (d) Resolution Level

Fig. 12. The predicted and measured results of session duration under different network contexts. There are no obvious differences between locations or CDNs, but users of higher-speed access links or those watching videos of higher resolution levels are more likely to stay longer.

Figure 11(d) shows session duration for different video tags. The users watching videos of Knowledge are likely to stay longer than those watching videos about Travel. However, even if video tags can affect session duration, the difference between measured and predicted values is still small.

C. QoE Scoring for Network Contexts

In this section, we evaluate the performance of scoring QoE for different network contexts. Network contexts directly determine the pattern of state chains. Therefore, it can score network contexts by the proposed session-duration QoE model.

According to section V-A, the state-transition probability is measured for various network contexts, including access type, location, CDN, and resolution level. Then, the QoE score of session duration is calculated respectively. The results are illustrated in figure 12.

Access Type: Session logs are divided by the type of user's access link, i.e. 3G, 4G, 5G and Wi-Fi. The measured and predicted session durations are shown in figure 12(a). It is noticed that the prediction accuracy maintains at a relatively high level. The difference between average values of predicted and real session duration shall not exceed 5%.

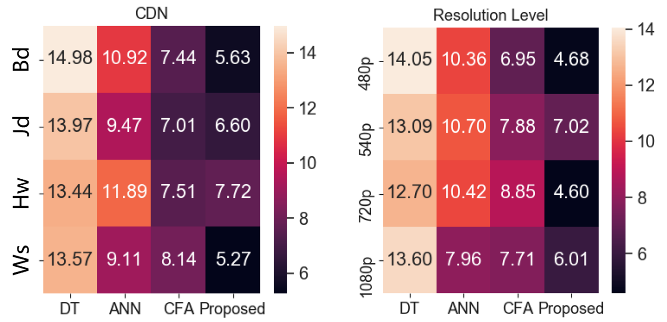
Location and CDN: Figure 12(b) and (c) demonstrated the values for various locations and CDNs. The predicted and measured values are almost the same.

Resolution-level: As is shown in figure 12(d), users would stay longer when they are watching videos with higher resolution, both for measured and predicted session duration. This verifies that the QoE model could accurately predict the user's session duration regardless of video resolution.

D. Comparison of Decision Algorithms

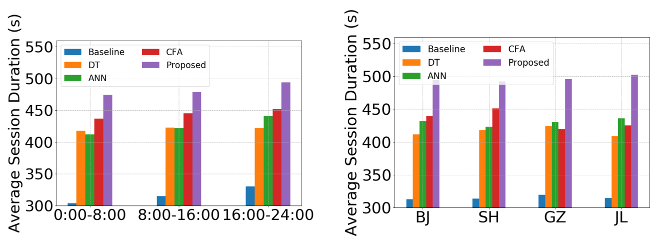
Finally, in this section, we evaluate the performance of the proposed scheme and compare it with other data-driven algorithms. Session duration is set as the optimization objective. We select the following methods for comparison:

- **Decision Tree (DT)** [35]: A fundamental machine-learning algorithm. It builds classification or regression models in the form of a tree structure. A decision node



(a) CDN (b) Resolution Level

Fig. 13. Comparison of prediction error (%) of session duration between different schemes.



(a) Time (b) Location

Fig. 14. Comparison of improvement on session duration. In general, the proposed scheme can improve the average session duration up to 60% compared with baseline, and 20% compared with other algorithms.

has two or more branches, and a leaf node represents a classification or decision.

- **Artificial Neural Network (ANN)** [36]: The most basic deep-learning algorithm. The numbers of neurons on each layer are 10x1000x32x2.
- **Critical Feature Analytics (CFA)** [37]: A latest CDN-resolution decision system for video QoE optimization. It automatically learns critical features from a lot of network metrics, and then derives the model by using K Nearest Neighbors (KNN) algorithm.

Since the comparison methods are dependent on the dimension of the feature vector, we supply the following attributes as features: *APP name, APP version, device model, operating system type* and *network operator*.

Comparison of session duration: The prediction accuracy of session duration is significant since the algorithms make their optimal decisions of CDNs and resolution levels based on the predicted session duration. Therefore, we calculated the average prediction error on each CDN/resolution level for all four schemes, and record them in figure 13. It is spotted that the proposed scheme usually obtains the highest accuracy, outperforming DT by up to 67%, ANN by 55% and CFA by 48%.

Comparison of CDN-resolution decision algorithms: As for the performance evaluation of optimization for the four schemes, the results are listed in figure 14.

It can be seen that due to the optimization objective being fixed as session duration, the performance of our proposed scheme on session duration is better than that of the comparison algorithms. In general, the proposed scheme can improve the average session duration up to 60% compared with the baseline, and 20% compared with other algorithms.

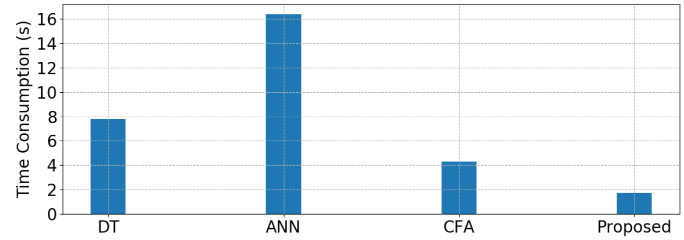


Fig. 15. Comparison of time consumption in the process of estimating parameters. The proposed scheme achieves the lowest complexity with its simple calculations.

The proposed scheme also outperforms other algorithms. CFA can sometimes compete with the proposed scheme, but mostly it is worse. DT usually obtains the worst performance. ANN is also unsatisfactory, mainly because it roughly treats all the input features as the same, while the lack of generalization ability also leads to bad performance.

System overhead: Finally, as for computation complexity, the evaluation results are shown in figure 15. As the only deep-learning-based algorithm, ANN takes the longest time in training. DT is also complex owing to its numerous decision nodes. CFA applies the KNN method to reduce calculation, but the consumption is still high. On the contrary, the proposed scheme achieves the lowest complexity with its simple calculations.

VII. Conclusion

It remains a critical challenge to choose a continuous-time objective and quantitative metric that reveals human-judge overall quality in video streaming. In this paper, we propose **ExQoE**, an exiting-ratio-based QoE modeling, scoring and optimizing system, which can precisely predict the exiting ratio and deduce session duration as the measurement of QoE, and optimize QoE by selecting the most appropriate CDN and resolution level. The extensive data study proves that user's exiting behavior is an appropriate choice as QoE metric. The discretization of playback sessions helps figure out the impacts of playing and stalling events on user's exiting behavior. The experimental results show that the modeling error of SER and deduced session duration are less than 2% and 10s respectively. By using the proposed scheme to optimize adaptive video streaming, the average session duration is improved up to 60% to baseline, and 20% to the existing black-box-like machine learning methods.

Acknowledgments

We thank the anonymous reviewers for their valuable feedback. This work was sponsored by the National Key R&D Program of China (2019YFB1802701), NSFC grant(U21B2012). We gratefully acknowledge the support of the State Key Laboratory of Media Convergence Production Technology and Systems, Key Laboratory of Intelligent Press Media Technology and ByteDance Inc.

References

[1] C. Aurrecoechea, A. T. Campbell, and L. Hauw, "A survey of qos architectures," *Multimedia systems*, vol. 6, no. 3, pp. 138–151, 1998.

[2] P. Juluri, V. Tamarapalli, and D. Medhi, "Measurement of quality of experience of video-on-demand services: A survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 401–418, 2015.

[3] Y. Chen, K. Wu, and Q. Zhang, "From qos to qoe: A tutorial on video quality assessment," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1126–1165, 2014.

[4] M. Alreshoodi and J. Woods, "Survey on qoe\qos correlation models for multimedia services," *arXiv preprint arXiv:1306.0221*, 2013.

[5] P. ITU-T RECOMMENDATION, "Subjective video quality assessment methods for multimedia applications," 1999.

[6] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "Developing a predictive model of quality of experience for internet video," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 339–350, 2013.

[7] D. Tsolkas, E. Liotou, N. Passas, and L. Merakos, "A survey on parametric qoe estimation for popular services," *Journal of network and computer applications*, vol. 77, pp. 1–17, 2017.

[8] M. Taha, A. Canovas, J. Lloret, and A. Ali, "A qoe adaptive management system for high definition video streaming over wireless networks," *Telecommunication Systems*, vol. 77, no. 1, pp. 63–81, 2021.

[9] S.-H. Chang, M.-L. Tsai, M.-H. Lee, and J.-M. Ho, "Optimal qoe scheduling in mpeg-dash video streaming." *International Journal of Interactive Multimedia & Artificial Intelligence*, vol. 6, no. 7, 2021.

[10] K. Brunnström, S. A. Beker, K. De Moor, A. Dooms, S. Egger, M.-N. Garcia, T. Hossfeld, S. Jumisko-Pyykkö, C. Keimel, M.-C. Larabi *et al.*, "Qualinet white paper on definitions of quality of experience," 2013.

[11] L. Yitong, S. Yun, M. Yinian, L. Jing, L. Qi, and Y. Dacheng, "A study on quality of experience for adaptive streaming service," in *2013 IEEE International Conference on Communications Workshops (ICC)*. IEEE, 2013, pp. 682–686.

[12] R. K. Mok, E. W. Chan, and R. K. Chang, "Measuring the quality of experience of http video streaming," in *12th IFIP/IEEE International Symposium on Integrated Network Management (IM 2011) and Workshops*. IEEE, 2011, pp. 485–492.

[13] W. Robitza, A. Ahmad, P. A. Kara, L. Atzori, M. G. Martini, A. Raake, and L. Sun, "Challenges of future multimedia qoe monitoring for internet service providers," *Multimedia Tools and Applications*, vol. 76, no. 21, pp. 22 243–22 266, 2017.

[14] M. J. Khokhar, T. Ehlinger, and C. Barakat, "From network traffic measurements to qoe for internet video," in *2019 IFIP Networking Conference (IFIP Networking)*. IEEE, 2019, pp. 1–9.

[15] R. Huang, X. Wei, Y. Gao, C. Lv, J. Mao, and Q. Bao, "Data-driven qoe prediction for iptv service," *Computer Communications*, vol. 118, pp. 195–204, 2018.

[16] A. W. Yusuf-Asaju, Z. M. Dahalin, and A. Ta'a, "A proposed analytical customer satisfaction prediction model for mobile internet networks," 2017.

[17] P. Spachos, T. Lin, W. Li, M. Chignell, A. Leon-Garcia, J. Jiang, and L. Zucherman, "Subjective qoe assessment on video service: Laboratory controllable approach," in *2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, 2017, pp. 1–9.

[18] F. Laiche, A. Ben Letaifa, I. Elloumi, and T. Aguili, "When machine learning algorithms meet user engagement parameters to predict video qoe," *Wireless Personal Communications*, vol. 116, no. 3, pp. 2723–2741, 2021.

[19] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, and H. Zhang, "Understanding the impact of video quality on user engagement," *ACM SIGCOMM computer communication review*, vol. 41, no. 4, pp. 362–373, 2011.

[20] A. Ahmed, Z. Shafiq, H. Bedi, and A. Khakpour, "Suffering from buffering? detecting qoe impairments in live video streams," in *2017 IEEE 25th International Conference on Network Protocols (ICNP)*. IEEE, 2017, pp. 1–10.

[21] H. Nam, K.-H. Kim, and H. Schulzrinne, "Qoe matters more than qos: Why people stop watching cat videos," in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 2016, pp. 1–9.

[22] F. Wang, C. Zhang, J. Liu, Y. Zhu, H. Pang, L. Sun *et al.*, "Intelligent edge-assisted crowdcast with deep reinforcement learning for personalized qoe," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 910–918.

[23] X. Zhang, Y. Ou, S. Sen, and J. Jiang, "{SENSEI}: Aligning video streaming quality with dynamic user sensitivity," in *18th USENIX Symposium on Networked Systems Design and Implementation (NSDI 21)*, 2021, pp. 303–320.

[24] Douyin, "Douyin," 2022, accessed December 2022. [Online]. Available: https://www.douyin.com

[25] M. H. Mazhar and Z. Shafiq, "Real-time video quality of experience monitoring for https and quic," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 1331–1339.

[26] C. Ge and N. Wang, "Real-time qoe estimation of dash-based mobile video applications through edge computing," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2018, pp. 766–771.

[27] H. B. Pasandi and T. Nadeem, "Latte: online mu-mimo grouping for video streaming over commodity wifi," in *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, 2021, pp. 491–492.

[28] T. Kimura, T. Kimura, A. Matsumoto, and K. Yamagishi, "Balancing quality of experience and traffic volume in adaptive bitrate streaming," *IEEE Access*, vol. 9, pp. 15 530–15 547, 2021.

[29] K. Fizza, A. Banerjee, K. Mitra, P. P. Jayaraman, R. Ranjan, P. Patel, and D. Georgakopoulos, "Qoe in iot: a vision, survey and future directions," *Discover Internet of Things*, vol. 1, no. 1, pp. 1–14, 2021.

[30] F. Bouali, K. Moessner, and M. Fitch, "Multi-timescale qoe provisioning for adaptive video streaming in heterogeneous deployments," *IEEE transactions on vehicular technology*, vol. 70, no. 2, pp. 1894–1909, 2021.

[31] Z. Li, B. Geng, X. Tao, Y. Duan, D. Gao, and S. Hu, "Video quality measurement for buffering time based on eeg frequency feature," in *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021, pp. 1424–1428.

[32] S. Meng, C. Qiao, J. Li, Y. Wang, and Z. Guo, "Dig: A data-driven impact-based grouping method for video rebuffering optimization," in *International Conference on Multimedia Modeling*. Springer, 2022, pp. 40–51.

[33] M. Ghosh, D. C. Singhal, and R. Wayal, "Desvq: Deep learning based streaming video qoe estimation," in *23rd International Conference on Distributed Computing and Networking*, 2022, pp. 19–25.

[34] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "The nonstochastic multiarmed bandit problem," *SIAM journal on computing*, vol. 32, no. 1, pp. 48–77, 2002.

[35] A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, "An introduction to decision tree modeling," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 18, no. 6, pp. 275–285, 2004.

[36] S.-C. Wang, "Artificial neural network," in *Interdisciplinary computing in java programming*. Springer, 2003, pp. 81–100.

[37] J. Jiang, V. Sekar, H. Milner, D. Shepherd, I. Stoica, and H. Zhang, "{CFA}: A practical prediction system for video {QoE} optimization," in *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)*, 2016, pp. 137–150.