Research Article

An Engagement Model Based on User Interest and QoS in Video Streaming Systems

Xiaoying Tan,[1] Yuchun Guo [iD],[1] Mehmet A. Orgun,[2] Liyin Xue,[3] and Yishuai Chen[1]

[1]*Beijing Jiaotong University, China*
[2]*Macquarie University, Australia*
[3]*Australian Taxation Office, Sydney, NSW 2000, Australia*

Correspondence should be addressed to Yuchun Guo; ychguo@bjtu.edu.cn

With the surging demand on high-quality mobile video services and the unabated development of new network technology, including fog computing, there is a need for a generalized quality of user experience (QoE) model that could provide insight for various network optimization designs. A good QoE, especially when measured as engagement, is an important optimization goal for investors and advertisers. Therefore, many works have focused on understanding how the factors, especially quality of service (QoS) factors, impact user engagement. However, the divergence of user interest is usually ignored or deliberatively decoupled from QoS and/or other objective factors. With an increasing trend towards personalization applications, it is necessary as well as feasible to consider user interest to satisfy aesthetic and personal needs of users when optimizing user engagement. We first propose an *Extraction-Inference (E-I)* algorithm to estimate the user interest from easily obtained user behaviors. Based on our empirical analysis on a large-scale dataset, we then build a *QoS and user Interest based Engagement (QI-E) regression model*. Through experiments on our dataset, we demonstrate that the proposed model reaches an improvement in accuracy by 9.99% over the baseline model which only considers QoS factors. The proposed model has potential for designing QoE-oriented scheduling strategies in various network scenarios, especially in the fog computing context.

1. Introduction

The past two decades have witnessed the growth and popularity of Video-on-Demand (VoD) applications on both PCs and mobile devices, and the trend is moving from basic video offering toward better quality of user experiences (QoE) in both industry and academia [1]. In video streaming services, the traditional Mean Opinion Score (MOS) is now replaced by *user engagement*, which more directly impacts the return for investment from stakeholders of a network ecosystem [2, 3].

The requirement of a good QoE for optimizing returns has led to a rapid development of network services and the emergence of new technology, including fog computing which could achieve real-time processing and feedback of high-volume video streaming and scalability of service on low-bandwidth output data [4].

As the first step to optimize user engagement, there is an urgent need for a general engagement model that could provide insight for diverse network environments, especially for the up-to-date fog computing context. There are several studies on understanding and modeling engagement in prior works where the quality of service (QoS) together with some objective context factors, such as location, device, and temporal attributes, have always been regarded as the fundamental factors related to engagement. The relationship of user engagement and QoS, either at the application level (e.g., *startup delay, buffer freque*ncy, and *bitrate*) or at the network level (e.g., *throughout, signal strength*) has been widely explored in related works.

However, most of the related works ignored another important factor-subjective human factors, such as users' personalized interest in a specific video. In fact, the new era of customization raises the aesthetic and personal needs

of users, especially in mobile video services where users typically present predictable features and service demand [5]. Accordingly, user engagement, as a reflection of "the degree of delight or annoyance of the user of an application or service [6]", cannot be satisfied only by high-quality delivery. Thus, sessions with the same buffer frequency sometimes have different user engagement due to the divergence of user interest in the video content. However, most of the prior works ignore or exclude the impact of the subjective factors as it is difficult to quantify them. Some works consider video popularity as a subjective human factor in QoE models [2, 7, 8], but video popularity could only roughly describe users' average preferences but without any personalization. Only a few works [9–11] at a macrolevel evaluate subjective human factors for each viewing session from psychological and cognitive perspectives. In these works, the subjective factors are obtained by extensive experiments and surveys with a very large population of subjects; however, such experiments and survey are expensive to conduct and not suitable for streaming applications in VoD systems.

An engagement model based on both *QoS* and *user interest* is not only necessary for accurately understanding and predicting user engagement, but also beneficial for optimizing system resource allocation and for providing better personalized services. On the one hand, finding out how the subjective factor impacts user engagement could help designers to deploy appropriate bandwidth resources for optimizing user engagement. On the other hand, recommendation system (RS) could make a tradeoff decision between the QoS factors and the human factors to recommend to users the videos that they are interested in and also that have good QoS and finally that they can enjoy for a longer time. Hence, it is critical to understand the relationship between engagement and the human factors as well as QoS in order to shed light on how best to allocate resources and customize services.

A challenge prior to building such a model is how to quantify the degree of a user's interest in a video. In prior recommender systems, user interest is measured either to be explicit ratings by user study or roughly to be implicit ratings, e.g., user engagement. However, the former measurement method is accuracy but is impossible to be collected in time-sensitive applications. Instead, the later one is inaccurate, as user engagement is sometimes not the reaction to his/her pure interest in the video but also impacted by other factors, e.g., quality issues like *startup delay*.

Another challenge is how to characterize the relationships between user engagement and the two factors in a unified model to provide insight for practical applications. Intuitively, the two factors impact user engagement but not independently, which is beyond the scope of a linear regression model. For example, user interest not only impacts user engagement itself but also impacts users' patience with the QoS problem. Machine learning (ML) algorithms, e.g., decision trees and Naive Bayes [2, 12], can characterize such a dependent relationship but not in a concise formula and therefore could not provide clear insight for practical applications.

We devote this paper to respond to these two challenges. We propose an *Extraction-Inference (E-I)* algorithm to estimate the user interest from easily obtained user behaviors. Through a measurement on a real-world VoD system, we analyze the relationship of user engagement with both QoS and user interest. Based on our empirical analysis on a large-scale dataset, we build a *QoS and user Interest based Engagement (QI-E) regression model*. Our empirical evaluation shows that the incorporation of human factors brings an improvement of 9.99% in prediction accuracy over the baseline model based on only QoS factors. Finally, we discuss the application potential of the proposed model and the future work.

2. Related Work

2.1. MOS versus Engagement Metrics. User engagement instead of Mean Opinion Score (MOS) is widely employed as QoE metric in streaming applications. As a standard metric in the ITU-T recommendations, MOS is a numeric value from 1 to 5 (i.e., poor to excellent) obtained through user studies or surveys [13]. There are many analyses on its influential factors in the domain of web services, E-commerce, multimedia, and so on [7, 13, 14]. However, since conducting a survey or user study is expensive, time- consuming, and without repeatability, MOS cannot be directly used in video streaming applications [13], especially in the fog computing context where the real-time response (real-time sensing and data processing) is required [5]. These limitations motivate the development of objective metrics, such as user engagement which quantifies a user's behavior as the reaction to the level of QoE. In video applications, user engagement is usually measured as session length [8, 15], abandonment rate [12, 15], number of visits [2], or skip rate [12]. These user behavior metrics are more directly relevant to increasing opportunities for advertising and upselling, leading to greater revenues [8].

2.2. QoS Parameters. QoS parameters are often studied as primary factors related to QoE since they can be controlled by the platform at least partially [9]. The specific QoS metrics vary across different domains [7], including the metrics on the application level, e.g., startup delay, buffer frequency, buffer ratio, and bitrate [2, 8, 16] and also including the metrics on the network level, e.g., flow throughput, flow duration, handover rate, and signal strength [12, 15]. Due to uncertainty of network conditions, QoS metrics on the application level capture the quality perceived by the users more closely than those on the network level.

Sometimes QoS metrics compete with or conflict each other and require a tradeoff in system design. For example, prior works [17, 18] point out the competing relationship between the initial time and the buffer event. Moreover, the tradeoff between the bitrate and the buffer event is often studied in bitrate adaptation schemes [17].

There are also several techniques proposed to adjust QoS parameters at the client, at the server, or in the network [19]. Those techniques include bitrate adaptation [20], prefetching [21, 22], transport protocol selection [14], and cache deployment [23].
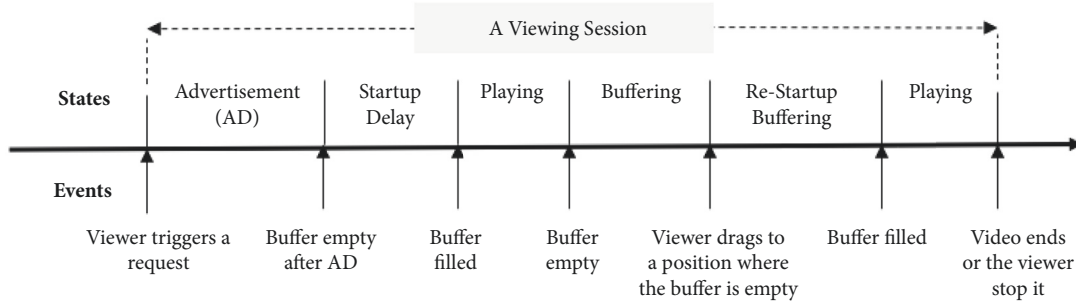
FIGURE 1: Illustration of a typical video streaming session.

2.3. Other Influential Factors. Recently, it has been realized that QoS factors alone cannot determine users' satisfaction and other potential "confounding" factors need to be explored [7]. The considered factors can be classified into context information such as connectivity and temporal effects [2], content attribute such as types and popularity [2], user attribute such as location [24], device, and gender [8].

However, user interests are often ignored. A few works addressed human factors (similar with user interest defined in this paper in essence) [9, 10] from psychological and cognitive perspectives and some attempted to integrate all kinds of factors including human ones [11, 25]. However, conducting such works is expensive, requiring a long period and the participation of both technicians and psychologists [9]. Identifying human subjective factors which are suitable for real-time streaming applications is still an open problem.

In terms of understanding the impact of human factors (e.g., user interest) on system design, one study [26] exploits individual interest to optimize the storage resource allocation in CDN. However, this study does not consider quality factors and hence it cannot help making a tradeoff decision on user interest and QoS.

2.4. Predicting User Interest. Predicting a user's interest in a specific item is the target problem addressed in a personalized recommendation system (RS), which is a hot topic in the face of information overload over the last decade [27, 28]. As one popular kind of recommendation algorithms, Collaborative Filtering (CF) algorithms, such as K-Nearest Neighbor, Bayesian belief nets, and Matrix Factorization, have been widely studied [27, 29–34]. As CF algorithms do not need data referring to item content, they are applicable to the video systems where explicit content descriptions of items (i.e., videos) are difficult to be obtained.

In addition, how to obtain and quantify user interest is also under study [35]. Most works ask users for explicit ratings after they purchase items, watch videos, or browse website. This method is accurate but not practical in time-sensitive applications like online VoD systems. Instead, user behaviors, such as the time spent on a page, scrolling and clicks on web pages [36, 37], time spent on a video [38, 39], and purchases in the past [40], are used as implicit ratings in some applications.

User behaviors are sometimes capable of serving as reliable implicit ratings as proved in [41], but in VoD systems they are still quite noisy [42]. For example, as we address in this paper, user behavior depends on not only user interest but also the QoS during the watching time. As far as we know, such noise in implicit rating has not yet been considered in existing VoD recommendation systems.

3. Problem and Definition

This section defines the scope of the problem we focus on and then provides the definition of the metrics of the factors considered in the target model.

3.1. Problem Statement. The main objective in this paper is to propose a practical engagement model based on the objective QoS factors and the subjective individual interest factor in VoD systems. For clarity, we omit the impact of other confounding factors, such as the type of video, device, and temporal attributes.

Thus, the objective is to build a model expressed as

$$E = f(\mathbf{Q}, r) \tag{1}$$

where r, E, and $Q^k \in \mathbf{Q}$, respectively, denote user interest level, user engagement, and the value of the k-th QoS metric in a session for a user w. To make the model clear for practical applications, the dependency function $f(\bullet)$ should be in a clear and concise form.

3.2. Metrics Definition. This section defines the metrics of the factors in our model. To clarify the definitions, we first introduce the terminologies of a viewing session by Figure 1.

A typical viewing session starts when a user initiates a request and ends when he/she finishes viewing the content or changes to another one or closes the client agent. At the very beginning, most users may experience a period of advertisement and sometimes a startup delay. During viewing, users may suffer from a frozen delay due to congestion or speed limit in network. Also, users can trigger a pause or drag the process bar to a new position. If the user drags to a new position where the content has not been prefetched already, he/she may experience a restartup delay.

According to the definition of a viewing session, we now define the metrics of three factors.

3.2.1. The Metric of Engagement. We define user engagement as the *valid watching ratio* in this paper, which measures how much the video has really played. It is computed as the ratio of the playback time to the video length. Here, the playback time is referred to as the time of a session excluding startup delay, restartup delay, buffering, pause, and advertisement time. In general, the range of engagement value is restricted within 0% to 120%.

3.2.2. The Metric of QoS. In this paper, we consider the application-level QoS metrics which capture delivery-related effects on the client-side. Specifically, we focus on the following metrics.

Startup Delay. It is the time before a video starts playing and immediately after the user initiates a request, exclusive of the time taken by advertisements. It is measured in seconds.

Buffer Frequency. It is the ratio of the number of buffering events to the total time of buffering and playing. It is measured in the number of times per minute.

Buffering Ratio. It is the ratio of the time spent on buffering or restartup buffering occurring in a session to the total time of buffering and playing. It is measured in percentage.

As an extension of the study in [2], we also consider another new quality metric, *average buffer length*.

Average Buffer Length. It is the average time the user has to spend in buffering once a buffer event occurs. It is computed as the ratio of the *buffering ratio* to the *buffering frequency*. It is a new metric we propose to complement the buffer frequency.

We do not discriminate between buffering and restartup buffering event and do not discuss another usually-used quality metric, *bitrate*, due to data absence. But this omission does not lessen the value of our study because the point here is not to study the relationship between quality metrics, but to address the competing or conflicting relationships between quality and user interest. If necessary or once the data is available, the model can be extended.

Compared with network-level QoS metrics, the metrics at application level are more generally applicable in diverse network contexts. They can manage the application-level QoS metrics through respective technology, such as service selection in content delivery network (CDN) [2], prepushing scheme in P2P network [43, 44], and caching technique in fog computing [4, 45, 46], and then optimize user engagement based on the application-level QoS metrics involved in the QoE model.

3.2.3. The Metric of User Interest. User interest is a user's subjective sense of concerning with and curiosity about the content of a specific video. We regard user interest to be a nondimensional parameter between 1 to 10. As discussed earlier in this paper, it is difficult to ask users for explicit ratings and directly using implicit ratings measured by user
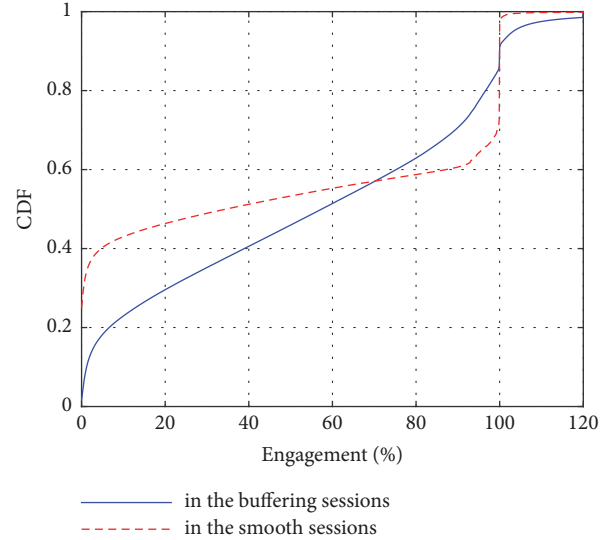

FIGURE 2: The distributions of user engagement in the buffering sessions and that in the smooth sessions. For clarity, we do not present the distribution of the engagement larger than 120% that are in the tiny minority.

behaviors may lead to inaccuracies. Thus, we propose an inference algorithm to obtain user interest, which will be depicted in Section 4.

3.3. Dataset. To build a data-driven model, we collect a large-scale dataset from the client-end of PPTV [47], a typical commercial P2P streaming system in China. The anonymous user logs range from 23th March to 28th March in 2011 with 75 million requests covering over 130,000 unique videos and 6 million users. The logs record viewing sessions of all the behavior-related, quality-related information for each user-item pair.

To filter the impact of other confounding factors (e.g., device, temporal attribute and the type of video), we only consider the sessions collected from client-end, during the periods of 7 p.m. to 12 p.m. and related to the videos of movie type in our analysis. But the method proposed here is not limited to these contexts and can be extended to other context, e.g., mobile device or other video types.

4. User Interest Inference

In this section, through the analysis of our dataset, we first show that user engagement is impacted not only by user interest but also by QoS. This means mapping the engagement time into user interest level is inaccurate. We then develop an *Extraction-Inference (E-I)* interest estimation algorithm and evaluate it on our dataset.

4.1. Measurement and Analysis. We compare the distribution of user engagement in the sessions with/without buffering events or startup delay (called *buffering sessions/smooth sessions,* respectively).

As the plots given in Figure 2 show, there are significant gaps, which show that the engagement alone fails to describe a

Step I (extraction). Extract users' interest from their engagement records in the **Smooth** sessions where users did not experience any quality problems including buffering event or start-up delay.
Step II (inference). Based on the extracted interest records in Step I, infer users' interest in other sessions.

ALGORITHM 1: **E**xtraction and **I**nference (E-I) interest estimation algorithm.

TABLE 1: Inference methods of Baseline II.

Gmean	the global average of all users' interest levels in the training sessions.
Uavg/Iavg	the average of the interest level of the active user/the active video.
UIavg	the geometric mean of the active user's and the active video's average interest levels.
KNN	the average of the interest of the active user's K nearest neighbors who have the most similar interest in other videos with her.

user's interest accurately, as user engagement is also impacted by quality problems. As shown, the buffering event makes the distribution of the engagement more even in the entire range. Without experiencing any buffering events, most users either finish watching the entire video (accounting for 32% of the sessions) or abandon the session quite early before 5% of the entire video is viewed (for 38%); on the contrary, in the buffering sessions, these two extreme cases account for below 22% together.

The gap of the distributions in Figure 2 is attributed to the quality problems that, on the one hand, impair users' watching experience and reduce the probability of the long engagement time and, on the other hand, indicate that the users still have a certain interest in the videos rather than abandoning the sessions at the beginning even before any buffering events.

4.2. Extraction-Inference (E-I) Algorithm. We propose a heuristic***Extraction and Inference (E-I) interest estimation*** algorithm based on the following two assumptions: (1) given the QoS and the seeking state, a user's engagement time only depends on her interest level in the video and (2) a user's preference remains consistent during a short period, as widely accepted in recommendation systems. We develop this algorithm in two steps as shown in Algorithm 1.

First, in the selected sessions, QoS and the seeking state are determined; i.e., all the values of the QoS metrics and the seeking state are equal to zero. Accordingly, the engagement time in these sessions is decided by users' interest. We uniformly map user engagement in smooth sessions into 10 bins referred to the user's interest implicit rating (where 1 represents poor and 10 excellent).

Based on the collected users' interest in the selected sessions, we next use Matrix Factorization (MF) [33, 34], a typical Collaborative Filtering (CF) algorithm [35–38], to infer their interest in other sessions. Compared with some other typical CF algorithms, e.g., KNN algorithm [35, 36], the MF algorithm is better at dealing with the data sparsity [36] and, in our experiments, the data used for training is quite sparse. The selected sessions (used for both training and testing) only account for 23% of the sessions in our dataset.

The MF algorithm supposes that users' interest can be explained by characterizing both the users and the videos to a joint latent factor space. In this space, users and videos should be represented to be M-dimensional latent factor vectors. For a user u, provided with her vector p_u and a video v's vector q_v, her interest in this video $\hat{r}_{u,v}$ can be predicted to be the product of these two vectors in addition to the global average interest μ, the user's bias b_u, and the video's bias b_v. That is, $\hat{r}_{u,v} = \mu + b_u + b_v + p_u^T q_v$. In practice, the latent factor vectors, p_u and q_v, and the biases, b_u and b_v, are learned by the stochastic gradient descent (SGD) algorithm [14] looping through the training dataset to minimize a utility function; i.e., $\min_{b_*, q_* p_*} \sum_{\tilde{r}_{u,v} \in \tilde{I}} (\tilde{r}_{u,v} - \mu - b_u - b_v - p_u^T q_v)^2 + \lambda(\|p_u\|^2 + \|q_v\|^2 + \|b_u\|^2 + \|b_v\|^2)$. Here $\tilde{I}$ denotes the set of the historical interest records.

4.3. Evaluation. We exploit the data in smooth sessions for training and evaluation. Via 10-fold cross validation, we fix the parameters, λ and M, in the E-I algorithm to be 0.05 and 30, respectively.

For comparisons, we propose two types of baseline methods. The first type (namely, Baseline I) modifies Step I of the E-I algorithm. It directly maps user engagement into user interest in all the sessions (regardless of the quality problems). The second type (namely, Baseline II) replaces the inference algorithm, the MF algorithm, with some statistical methods and another CF algorithm, K-Nearest Neighbors (KNN), as defined in Table 1.

We evaluate the engagement prediction accuracy (measured by Root Mean Square Error (RMSE) [39]) of the algorithms on the "selected" sessions of the test dataset where the engagement records do represent users' pure interest.

According to the results listed in Table 2, E-I algorithm improves the engagement prediction accuracy by 18% on Baseline I, 27% on Gmean, 7.26% on Uavg, 8.9% on Iavg, 6.3% on UIavg, and 4.8% on KNN. The positive results of our proposed E-I algorithm confirm that it is necessary to eliminate the effect of the other relevant factors when we extract user interest from their engagement records.

5. Engagement Predictive Model

In this section, we measure the relationships of user engagement with user interest and QoS, respectively, and then propose the engagement predictive model.

TABLE 2: Estimation accuracy of user interest inference models and the improvement achieved by E-I algorithm on the baseline methods. The evaluation is conducted in the smooth sessions.

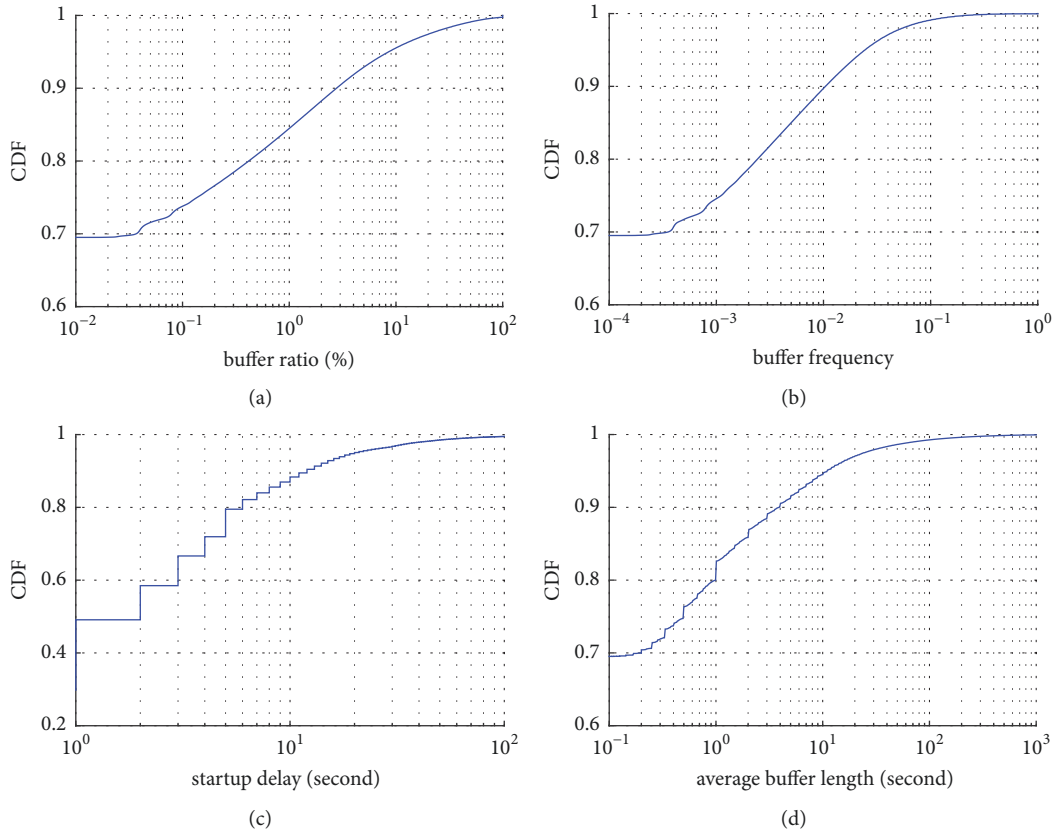| Methods	Baseline I	Baseline II					E-I algorithm
		Gmean	Uavg	Iavg	UIavg	KNN	
RMSE	3.63	4.07	3.19	3.25	3.16	3.11	2.96
Improvement (%)	18.4	27.2	7.2	8.9	6.3	4.8	- -

(a)

(b)

(c)

(d)

FIGURE 3: Cumulative distribution plots for QoS metrics.

5.1. User Engagement versus QoS Metrics

5.1.1. QoS Distribution.
We first look at the distributions of the values of the various QoS metrics. As shown in Figure 3, we find that the system has a good quality generally although the quality problems are not trivial in some sessions. 43.5% of the sessions in our dataset have not endured any quality problems. Specifically, 70% of the sessions have not endured a buffer event and 50% have not experienced *startup delay*. Still, there are some sessions having endured quite poor quality situations. For example, 5% of the sessions endure a buffering ratio over 10%. 5% of the sessions have more than 5 buffering events in their 100-second playing time.

The generally good quality situation restrains the usage range of the traditional engagement models that consider QoS only. These models could not differentiate the sessions without quality problems, although the QoS metrics should be taken into account as relevant factors as shown earlier.

5.1.2. Correlation Analysis.
Next, we examine the expectations of user engagement conditional on various quality metrics. Given a quality value $Q = q$, the conditional expectation of user engagement is calculated to be $\mathbb{E}(E \mid Q = q) = \sum_{e \in D(E)} e P(E = e \mid Q = q)$, where $P(E = e \mid Q = q)$ is the conditional probability and $D(E)$ is the range of the engagement value as defined in Section 3.2.

The plots of the conditional expectations are shown in Figure 4. in the dominant range of *buffer frequency* ([0, 18%] as shown in Figure 3(b)), as an example, the engagement generally decreases log-linearly as the quality becomes poorer as expected, as shown in Figure 4(b), which means that the decrease rate slows down with a larger buffer frequency. This result supports the intuition that users are less sensitive to a worse quality situation once they have suffered a bad one. The relationship between user engagement and the buffering ratio shows the similar characteristics. But with a larger average buffering length or startup delay, as shown in Figures 4(d)
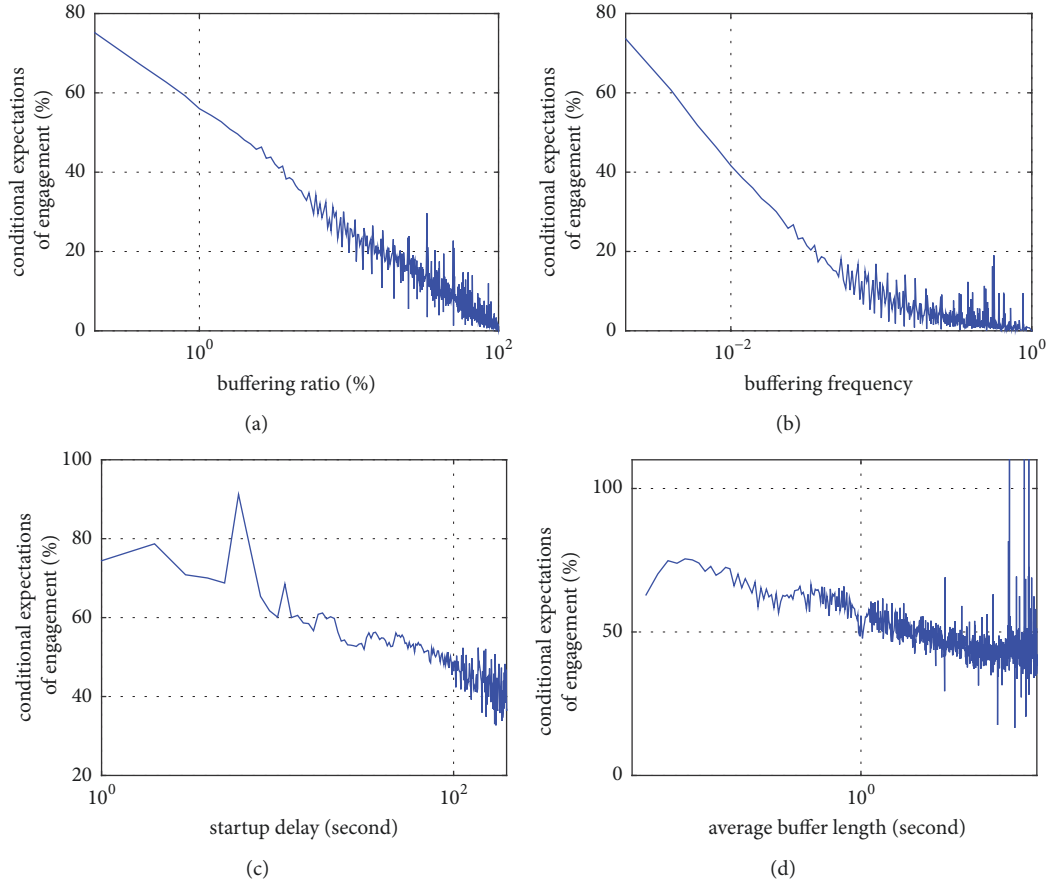
FIGURE 4: Expectations of user engagement against diverse QoS metrics. (a) Buffering ratio, (b) buffering frequency, (c) startup delay, and (d) average buffer length.

TABLE 3: Correlation coefficients of user engagement with QoS and user interest, respectively. The values of the QoS metrics are measured in logarithmic scale. *The correlation coefficient of user engagement with user interest will be introduced in Section 5.2, while those with QoS metrics are introduced in Section 5.1.

Correlation metric	Buffer ratio	Buffer frequency	Average buffer length	Startup delay	Interest*
Pearson	-0.2363	-0.3332	-0.0985	-0.1195	0.3374
Spearman	-0.2273	-0.3435	-0.1034	-0.1447	0.3595

and 4(c), user engagement decreases not as smoothly, which means a weaker correlation between user engagement and these two metrics.

Furthermore, we quantify the correlation coefficients of the *logarithmical* values of the QoS metrics with the conditional expectations of engagement. To alleviate value 0 in the logarithm operators, all the QoS values are increased by 1; i.e., $Q' = \ln(Q + 1)$, where Q is the original QoS value and Q' is the logarithmical one. We employ the correlation metrics Pearson correlation coefficient t [36] and Spearman rank coefficient [39]. The first metric could identify the linear relationship with some Gaussian noise while the second one emphasizes the monotonicity between variables. As shown in Table 3, the results confirm that, among all the QoS metrics, buffer frequency and restartup buffer ratio show the strongest log-linear correlations with engagement, which means the largest weight should be assigned to these two metrics in the engagement predictive model.

5.2. User Engagement versus User Interest. We now examine the relationship of user engagement and user interest. As shown in Figure 5(a), user engagement increases linearly with an increasing interest level, except for that at the head ([1, 2]) and the tail ([9.7, 10]). Fortunately, the ranges of the head and tail only account for 0.9% of the sessions. Thus, the linear characteristic still dominates the relationship.

Moreover, the correlation coefficient between user engagement and user interest is stronger than that with most of the QoS metrics, as listed in Table 1. The results confirm that individual interest has a roughly linear dependence on the engagement.
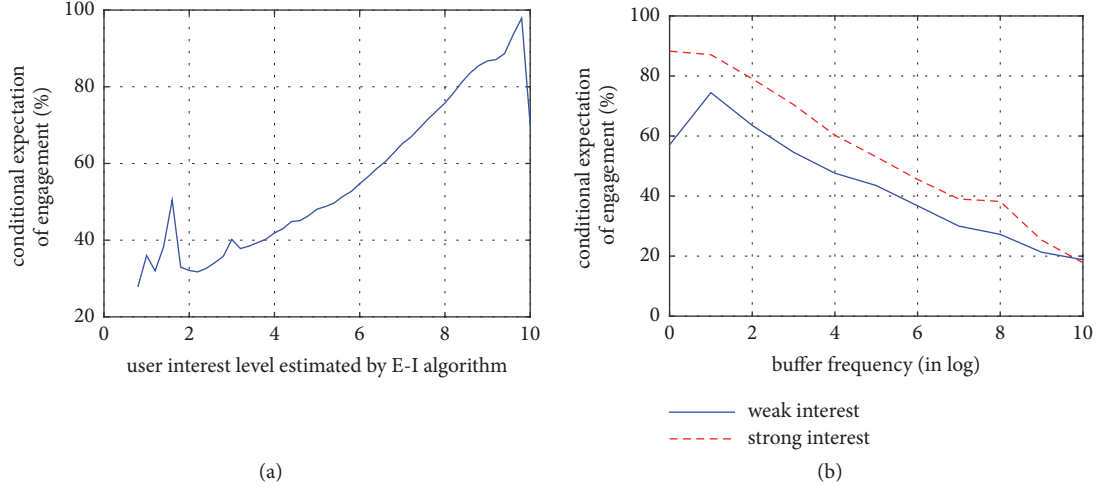
(a)

weak interest
strong interest

(b)

FIGURE 5: The expectations of user engagement conditional on (a) the user interest level and those on (b) the buffering frequency in the sessions where users are, respectively, strongly and weakly interested in the videos.

5.3. User Engagement versus QoS versus User Interest. Intuitively, QoS metrics and user interest may not independently impact user engagement. For example, users may have different tolerance to the quality issues when they have divergent interest degree towards the videos.

To confirm this intuition, we compare the engagement-QoS relationship under the strong-interest level and that under the weak-interest level, respectively, as shown in Figure 5(b). We use buffer frequency here as an example of the QoS metrics. In statistics, we split the logarithmical value of buffer frequency into 10 bins and denote the binned buffer frequency with a score between 0 and 10. The interval is set to be 1 score for statistic.

The boundary between the strong- and the weak-interest level is set to be the median of all the users' interest levels.

As shown, when the users are more interested in the video, the engagement decreases more quickly against an increasing buffer frequency. In other words, users with a stronger interest tend to be more sensitive to the quality problem. The significant difference between these two cases indicates the existence of the multiplicative effect of user interest and the QoS on user engagement.

6. Model Building and Evaluation

6.1. Engagement Model. According to the measurement results above, we propose a **QoS and user Interest based Engagement (QI-E) regression model**. In this model, for user u with an interest rating R with video v, we predict his/her engagement E for this video to be

$$E = f\left(\mathbf{Q}', R\right) = \boldsymbol{\theta}_q^T \mathbf{Q}' + \theta_r R + r \boldsymbol{\theta}_{qr}^T \mathbf{Q}' + \theta_0 \qquad (2)$$

where $\mathbf{Q}' = [Q'_1, \cdots, Q'_k, \cdots Q'_K]$ is a QoS vector where the element Q'_k is the logarithmical value of the kth QoS metric; i.e., $Q'_k = \ln(Q_k + 1)$, where Q_k is the original QoS value. All the QoS metrics are involved in the vector, and then

$K = 4$. $\theta_* \in \boldsymbol{\theta}$ are the weight parameters to be learned by experiments.

Based on this model, the expectation of user engagement conditional on the kth QoS metric with a logarithmical value of $Q'^{(k)} = q$ can be derived to be

$$\mathbb{E}\left(E \mid Q'^{(k)} = q\right) = (\mathbb{E}(R) + 1)q + c\mathbb{E}(R) + c \qquad (3)$$

where $\mathbb{E}(R)$ is the expectation of user interest rating. $c = \sum_{\bar{k}=1, \neq k}^{K} \mathbb{E}(Q'^{(\bar{k})}) + 1$ and $Q'^{(\bar{k})}$ is the logarithmical value of the $\bar{k}$th QoS metric but not the kth one. Both of them are constants. For clarity, the parameters $\theta_* \in \boldsymbol{\theta}$ are omitted in the derivation.

From (3), we observe that the conditional expectation of user engagement $\mathbb{E}(E \mid Q'^{(k)} = q)$ is linearly proportional to the logarithmical QoS value q, which is consistent with the measurement result in Section 5.1. Furthermore, with a larger interest rating R, there is a larger slope, $\mathbb{E}(R) + 1$, in the linear relationship, which is consistent with the measurement result in Section 5.3. Similarly, through a deviation of user engagement conditional on interest rating $R = r$, it is easy to show that the model is consistent with the measurement result in Section 5.2.

6.2. Model Evaluation. We randomly select 80% of the whole dataset for training and the rest for testing. Via 10-fold cross validation, we fix the parameters θ_* in the QI-E model.

For comparison, we propose three groups of linear regression models as baselines. In the first group, denoted by BS-1, the models consider separate QoS metrics, respectively. In the second baseline model, denoted by BS-2, we take into account all the QoS metrics and the third one, denoted by BS-3, considers user interest additionally.

In Table 4(a), we first evaluate the first group baselines models. Among them, the model considering *buffer frequency* has the smallest *RMSE of 3.1553* as expected as this metric has the largest correlation coefficient with engagement as shown in Table 3.

TABLE 4: Performance of the engagement predictive models.

(a) Models BS-1 using each single QoS metric as input feature

Input Feature	Buffer ratio	Average buffer length	Buffer frequency
RMSE	3.1727	3.1871	3.1553

(b) Models using quality factors and human factors

Model	Input Feature	RMSE	Improvement
BS-2	All the QoS metrics	3.0926	- -
BS-3	QoS, interest	2.9846	3.5%
QI-E	QoS, interest (multiplicative item)	2.8588	7.6%

In Table 4(b), we evaluate the proposed QI-E model compared with the model that consider all the QoS metrics. As the results shown, when user interest is considered, there is an improvement of 3.5%. When the multiplicative effect of the two kinds of factors is considered in the QI-E model, the improvement climbs up to 7.6%.

The positive results confirm the effectiveness of our proposed engagement predictive model and demonstrate the necessity of understanding how the QoS factors and user interest impact user engagement. Although the experiment is conducted on a dataset from an application on PC-clients, our method and results are easy to be extended to the context of mobile clients.

7. Summary and Discussion

7.1. Summary. In this paper we have shown that, in order to optimize user engagement in VoD streaming systems directly, an effective model of engagement incorporating both user interest and perceptual quality factors in an explicit function is needed. To this end, we have proposed an Extraction-Inference (E-I) algorithm to estimate the user interest from easily obtained user behaviors. Furthermore, we have built a QoS and user Interest based Engagement (QI-E) regression model based on an experimental analysis over a large-scale dataset. This model offers an improvement in accuracy by 9.99% over the baseline model considering only QoS factors. The positive results demonstrate that user interest as well as QoS plays an important role in user engagement prediction.

7.2. Discussion for Implications. Our research on understanding and modeling user engagement can be applied to most of the up-to-date network environments, including fog computing. For example, it could help designers to make tradeoffs between QoS factors under diverse user interest through *CDN selection, streaming decisions,* and so on [48, 49].

Especially in the fog computing context, as provided with fully explored localized user features and service demand, user interest in this case is more predictable and users' personalized requirement could be better satisfied [4, 46]. For a goal of global optimization, based on our model, designers could provide the users with higher interest with a priority to the bandwidth optimization under a limited bandwidth condition. As another example, in an engagement-oriented recommendation system, designers can make a tradeoff between videos that satisfy user interest and that provide a better QoS.

In the future, we will further extend the model to consider more factors. For example, some other kinds of user behaviors like drag can help understand the user's patience. Moreover, we can design scheduling schemes specifically for the fog computing context.

Data Availability

The dataset is supported by PPTV, a commercial enterprise. The data is not for public.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] B. Li, Z. Wang, J. Liu, and W. Zhu, "Two decades of internet video streaming: A retrospective view," in *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, vol. 9(1s), p. 33, 2013.

[2] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "Developing a predictive model of quality of experience for internet video," in *Proceedings of the ACM SIGCOMM 2013 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, SIGCOMM 2013*, pp. 339–350, China, August 2013.

[3] M. Watson, "Http adaptive streaming in practice," in *Proceedings of the In of the ACM Multimedia Systems Conference (MMSys)Keynote*, San Jose, CA, USA, 2011.

[4] S. Yi, C. Li, and Q. Li, "A survey of fog computing: concepts, applications and issues," in *Proceedings of the Workshop on Mobile Big Data (Mobidata '15)*, pp. 37–42, ACM, Hangzhou, China, June 2015.

[5] V. Pande, C. Marlecha, and S. Kayte, "A review- fog computing and its role in the internet of things," *International Journal of Engineering Research and Application*, vol. 6, no. 10, pp. 2248–96227, 2016.

[6] Recommendation ITU-T P.10/G.100 Amendment 2, Std., Jul. 2008.

[7] V. A. Siris, K. Balampekos, and M. K. Marina, "Mobile quality of experience: Recent advances and challenges," in *Proceedings of the 2014 IEEE International Conference on Pervasive Computing and Communication Workshops, PERCOM WORKSHOPS 2014*, pp. 425–430, Hungary, March 2014.

[8] S. S. Krishnan and R. K. Sitaraman, "Video stream quality impacts viewer behavior: Inferring causality using quasi-experimental designs," *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 2001–2014, 2013.

[9] K. U. R. Laghari, N. Crespi, B. Molina, and C. E. Palau, "QoE aware service delivery in distributed environment," in *Proceedings of the 25th IEEE International Conference on Advanced Information Networking and Applications Workshops, WAINA 2011*, pp. 837–842, Singapore, March 2011.

[10] W. Wu, A. Arefin, R. Rivas, K. Nahrstedt, R. Sheppard, and Z. Yang, "Quality of experience in distributed interactive multimedia environments: Toward a theoretical framework," in *Proceedings of the 17th ACM International Conference on Multimedia, MM'09, with Co-located Workshops and Symposiums*, pp. 481–490, China, October 2009.

[11] K. U. R. Laghari, K. Connelly, and N. Crespi, "Toward total quality of experience: a QoE model in a communication ecosystem," *IEEE Communications Magazine*, vol. 50, no. 4, pp. 58–65, 2012.

[12] M. Z. Shafiq, J. Erman, L. Ji, A. X. Liu, J. Pang, and J. Wang, "Understanding the impact of network dynamics on mobile video user engagement," *ACM SIGMETRICS Performance Evaluation Review*, vol. 42, no. 1, pp. 367–379, 2014.

[13] M. Alreshoodi and J. Woods, "Survey on QoE\QoS correlation models for multimedia services," *International Journal of Distributed and Parallel Systems*, vol. 4, no. 3, 2013.

[14] T. Hoßfeld, R. Schatz, and U. R. Krieger, "Qoe of Youtube video streaming for current internet transport protocols," in *Measurement, Modelling, and Evaluation of Computing Systems and Dependability and Fault Tolerance*, pp. 136–150, Springer International Publishing, 2014.

[15] A. Balachandran, V. Aggarwal, E. Halepovic et al., "Modeling web quality-of-experience on cellular networks," in *Proceedings of the 20th ACM Annual International Conference on Mobile Computing and Networking, MobiCom 2014*, pp. 213–224, USA, September 2014.

[16] F. Dobrian, V. Sekar, and A. Awan, "Understanding the impact of video quality on user engagement," in *Proceedings of the ACM SIGCOMM Conference (SIGCOMM '11)*, vol. 41, pp. 362–373, August 2011.

[17] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, "A survey on quality of experience of HTTP adaptive streaming," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 469–492, 2015.

[18] Y. Qi and M. Dai, "The effect of frame freezing and frame skipping on video quality," in *Proceedings of the Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP'06. International Conference on IEEE*, pp. 423–426, 2006.

[19] J. Jiang, V. Sekar, and H. Zhang, "Improving fairness, efficiency, and stability in http-based adaptive video streaming with festive," in *Proceedings of the 8th International Conference on Emerging Networking Experiments And Technologies*, pp. 97–108, 2012.

[20] L. De Cicco, S. Mascolo, and V. Palmisano, "Feedback control for adaptive live video streaming," in *Proceedings of the of the second annual ACM conference on Multimedia systems*, pp. 145–156, 2011.

[21] Y. Xu, E. Altman, R. El-Azouzi, S. E. Elayoubi, and M. Haddad, "QoE analysis of media streaming in wireless data networks," in *NETWORKING*, pp. 343–354, Springer Berlin Heidelberg, 2012.

[22] S. Khemmarat, R. Zhou, D. K. Krishnappa, L. Gao, and M. Zink, "Watching user generated videos with prefetching," *Signal Processing: Image Communication*, vol. 27, no. 4, pp. 343–359, 2012.

[23] S. H. Shen and A. Akella, "An information-aware qoe-centric mobile video cache," in *Proceedings of the of the 19th annual international conference on Mobile computing networking*, pp. 401–412, 2013.

[24] A. Gember, A. Akella, J. Pang, A. Varshavsky, and R. Caceres, "Obtaining in-context measurements of cellular network performance," in *Proceedings of the 2012 ACM Internet Measurement Conference, IMC 2012*, pp. 287–300, USA, November 2012.

[25] A. Floris, L. Atzori, and G. Ginesu, "Addressing un-interoperability issues in QoE models: Is a layered modelling effective?" in *Proceedings of the 2014 IEEE International Conference on Communications Workshops, ICC 2014*, pp. 563–568, Australia, June 2014.

[26] M. Verhoeyen, J. D. Vriendt, and D. D. Vleeschauwer, "Optimizing for video storage networking with recommender systems," *Bell Labs Technical Journal*, vol. 16, no. 4, pp. 97–113, 2012.

[27] K. Verbert, N. Manouselis, X. Ochoa et al., "Context-aware recommender systems for learning: a survey and future challenges," *IEEE Transactions on Learning Technologies*, vol. 5, no. 4, pp. 318–335, 2012.

[28] A. S. Lampropoulos and G. A. Tsihrintzis, "A survey of approaches to designing recommender systems," in *Proceedings of the Multimedia Services in Intelligent Environments*, pp. 7–30, Springer International Publishing, 2013.

[29] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, 2005.

[30] L. Si and R. Jin, "Flexible mixture model for collaborative filtering," *ICML*, vol. 3, pp. 704–711, 2003.

[31] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in Artificial Intelligence*, vol. 2009, 19 pages, 2009.

[32] N. Wang, T. Yao, J. Wang, and D. Y. Yeung, "A probabilistic approach to robust matrix factorization," in *Computer Vision–ECCV 2012*, pp. 126–139, Springer Berlin Heidelberg, 2012.

[33] R. Mittelman and E. L. Miller, "Fast Gauss Transforms based on a High Order Singular Value Decomposition for Nonlinear Filtering," in *Proceedings of the 2007 IEEE/SP 14th Workshop on Statistical Signal Processing*, pp. 94–98, Madison, WI, USA, August 2007.

[34] Y. Koren, "The bellkor solution to the netflix grand prize," *Netflix Prize Documentation*, 2009.

[35] M. Montaner, B. López, and J. L. de la Rosa, "A taxonomy of recommender agents on the internet," *Artificial Intelligence Review*, vol. 19, no. 4, pp. 285–330, 2003.

[36] W. Woerndl, A. Helminger, and V. Prinz, "Experiences from implementing collaborative filtering in a web 2," in *Proceedings of the International Workshop on Adaptation and Personalization for Web*, vol. 2, pp. 120–129, 2009.

[37] M. Claypool, P. Le, M. Wased, and D. Brown, "Implicit interest metrics," in *Proceedings of the 6th international conference on Intelligent user interfaces*, pp. 33–40, 2001.

[38] D. Weiß, J. Scheuerer, M. Wenleder, A. Erk, M. Gülbahar, and C. Linnhoff-Popien, "A user profile-based personalization system for digital multimedia content," in *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts, DIMEA 2008*, pp. 281–288, Greece, September 2008.

[39] J. Davidson, B. Liebald, J. Liu et al., "The YouTube video recommendation system," in *Proceedings of the Fourth ACM Conference on Recommender Systems*, pp. 293–296, 2010.

[40] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian personalized ranking from implicit feedback," in *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pp. 452–461, AUAI Press, 2009.

[41] D. O'Sullivan, B. Smyth, and D. Wilson, "Explicit vs implicit profiling-a case-study in electronic programme guides," in *IJCAI*, 2003.

[42] Y. Hu, Y. Koren, and C. Volinsky, "Collaborative filtering for implicit feedback datasets," in *Proceedings of the Data Mining, ICDM'08. Eighth IEEE International Conference on IEEE*, pp. 263–272, 2008.

[43] F. Lin and X. Lü, "QoS guaranteed pre-pushing scheme in peer-assisted streaming network," *China Communications*, vol. 11, no. 14, pp. 111–117, 2014.

[44] F. Lin, X. Zhou, X. Lü, and W. Song, "Novel pre-pushing scheme for peer-assisted streaming network based on multi-leader multi-follower stackelberg model," *Wireless Personal Communications*, vol. 80, no. 1, pp. 289–301, 2015.

[45] J. Su, F. Lin, X. Zhou, and X. Lu, "Steiner tree based optimal resource caching scheme in fog computing," *China Communications*, vol. 12, no. 8, Article ID 7224698, pp. 161–168, 2015.

[46] T. H. Luan, L. Gao, Z. Li et al., "Fog computing: focusing on mobile users at the edge," *Computer Science*, 2015.

[47] http://www.pptv.com/.

[48] H. Patrick, *"How Much Online Video Quality is Enough?" white paper, Senior Director, Product Marketing at Citrix*, Japan, 2011.

[49] D. D. Vleeschauwer and K. Laevens, "Performance of caching algorithms for IPTV on-demand services," *IEEE Transactions on Broadcasting*, vol. 55, no. 2, pp. 491–501, 2009.