User Mapping Strategies in Multi-Cloud Streaming: A Data-driven Approach

Guowei Zhu[1], Chou Mo[1], Zhi Wang[2], and Wenwu Zhu[1]
[1]Department of Computer Science and Technology, Tsinghua University
[2]Graduate School at Shenzhen, Tsinghua University
{zgw15@mails.,mc14@mails.,wangzhi@sz.,wwzhu@}tsinghua.edu.cn,

***Abstract*—Using content delivery networks (CDNs) for video distribution has become a *de facto* approach for today's video streaming, due to the easy usage and good scalability. Today, it has become a norm rather than an exception for video providers to hire multiple cloud CDNs for their video services in a pay-per-use manner, to not only serve users at different locations, but also reduce the operation costs. Given the multiple CDNs and their peering servers at many different locations, mapping a user to an edge CDN server has become a critical decision that can affect the quality of experience (QoE) of users. Conventional user mapping strategies are generally rule-based, e.g., assigning users to CDN servers according to only their locations or ISPs, which cannot guarantee any QoE. In this paper, we first propose to use a data-driven approach to study factors determining the streaming QoE in the multi-cloud CDN paradigm. Our findings suggest that the streaming QoE is affected by a combination of not only network factors but also user factors including their preference of video content. Then, we design a machine learning based predictive model to capture the QoE given the network conditions and user preference. Finally, we formulate the user mapping problem as an optimization problem and design algorithms to solve it: our algorithms identify users whose QoE are mostly affected by QoS and assign users to CDN servers so that the overall QoE can be maximized. Trace-driven experiments further verify the effectiveness of our design.**

***Keywords*—** CDN User Mapping, User Preference, QoE, Video Delivery

I. Introduction

Using the cloud infrastructure to deploy scalable content delivery networks (CDNs) for video streaming has become a norm rather than an exception in today's Internet. Compared to traditional CDNs, cloud-based CDNs give the video service providers the scalable service capacity in a pay-per-use manner [1]. To serve the increasing users from all over the world, it is promising for a video service provider to use *multiple* cloud providers to deploy their video streaming services, instead of using only a single cloud provider. A *multi-cloud* CDN paradigm can potentially improve the network performance (e.g., bandwidth) for users [2], since multiple clouds provide more *locations* where servers can be deployed, so that users can download video chunks from servers right next to them.

Though the multi-cloud CDN paradigm is a promising solution to improve the network performance for users, it is nontrivial to perform effective *user mapping* in this paradigm, determining *which user should be served by which cloud CDN server, to improve the quality of experience (QoE) for users in the online video streaming system.* As illustrated in Fig. 1, the multi-cloud paradigm provides cloud CDN servers at different locations hosted by different ISPs, effective user mapping can match users with the best CDN servers to meet their QoS requirement.

Conventional strategies to assign users to CDN servers are generally rule based [3], which cannot satisfy the QoE requirement in the video streaming context, due to the following reasons.

First, rule-based user assignment cannot reflect the dynamics of the network performance. In rule-based user mapping strategies, users' locations and hosting ISPs are usually considered for the request redirection. For example, Liu et al. [4] designed a coordinated video control plane that uses the location and ISP information to schedule users to download videos from different CDN servers. The limitation of such rule-based strategies is that they usually fail to capture the changing network conditions, and users can experience highly different network performance over time due to the varying background traffic patterns [5].

Second, QoE in video streaming is affected by a variety of factors, including non-QoS factors. Previous studies on understanding the QoE of users are mainly focused on the network performance, e.g., Dobrian et al. [6] confirmed the impact of QoS factors including buffering ratio and rendering quality; however, user characteristics and behaviors can also significantly affect their experience in a streaming session, e.g., the preference of a video content.

To address these problems, we propose a QoE-aware cloud CDN user mapping strategy. We first design a data-driven QoE model, which captures *both* the QoS factors and user factors that determine a user's quality of experience in a video streaming session. Then, we use the knowledge provided by the QoE model, to identify the "targeting users" that can mostly benefit from an improved CDN selection, i.e., users whose marginal quality of experience can be improved the most in the user mapping. Finally, we formulate the user mapping as an optimization problem to maximize the overall user experience in the multi-cloud CDN paradigm. In particular, our contributions can be summarized as follows.

▷ Using a dataset of a largest online video service provider recording over 20K users accessing about 22K videos in nearly 1 million sessions, we study the correlation between users' *engagement* (i.e., a session completion indication [6]) and both QoS factors including download speed and RTT experienced

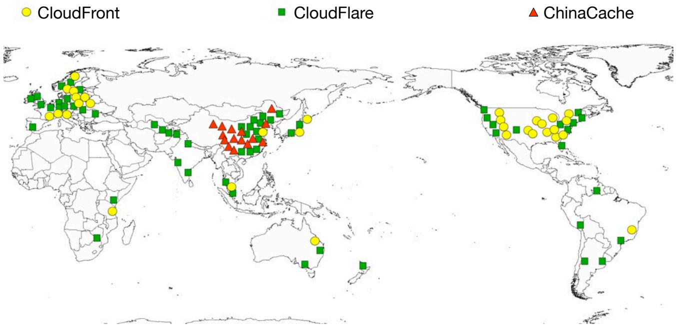

Fig. 1. Multi-cloud CDN for video delivery.

by users, and user factors including their preference of the video content. Based on our information gain studies, we observe that (1) the importance of these factors varies for different users, and (2) there are users whose QoE can be significantly improved by user mapping.

▷ We formulate the problem of user mapping as an optimization problem, and design a practical algorithm to solve it. In our algorithm, a marginal QoE improvement is calculated for candidate users. In each scheduling round, users with the highest marginal QoE improvement by user mapping will be scheduled to download from selected CDN servers.

▷ We use trace-driven experiments to verify the effectiveness of our design. In particular, we identify the impact of the implementation parameters including the user preference and QoS factors. Compared with conventional strategies including rule-based user mapping, our design can significantly improve the QoE of users up to 30%.

The rest of the paper is organized as follows. Section II describes related works. We present our measurement study of the correlation between QoS/user factors and users' QoE in Section III. We design the QoE model and formulate the user mapping problem in the multi-cloud CDN paradigm in Section IV, and evaluate its performance in Section V. Finally, we conclude the paper in Section VI.

II. Related Work

Hosting content across multiple CDNs can potentially improve the content availability and quality of service for content delivery [7]. Wang et al. [8] propose to jointly perform video transcoding and video delivery for adaptive streaming in an online manner to improve user experience. A problem in the video streaming scenario is how to improve the streaming quality using the multi-cloud CDN paradigm. The limitation of conventional rule-based strategies is that they fail to consider the complex factors that affect the quality of experience of users in video streaming, including both network factors and user factors.

Several efforts have been devoted to understanding QoE of users in video streaming. Song et al. [9] adopted regression analysis and presented a streaming quality assessment framework, Q-score, which consists of both offline learning and online computation. Mok et al. [10] proposed three relationship models between network conditions and quality of experience. Zhang et al. [11] proposed a video engagement metric, which is defined by the completion ratio of video sessions, to avoid subjective measurement of user experience. To the best of our knowledge, an important factor, user preference of video content, which affects the streaming QoE has not been well studied for user mapping in multi-cloud CDN. In our study, we propose to jointly use network and user factors to make user mapping decisions.

III. Characterizing QoS and User Factors Affecting QoE

In this section, we will study key factors that affect users' QoE in video streaming.

A. Methodology

We use a data-driven approach to study the factors affecting users' engagement. We first present the datasets used in our study. Then we give the calculation of streaming engagement.

1) Datasets: Our objective is to find out the key factors that affect streaming QoE, as well as the importance of these factors. To this end, we use a dataset of one of the largest online video service provider recording over 20K users accessing about 22K videos in nearly 1 million sessions in a whole month in 2013, we study the correlation between users' streaming engagement and QoS/user factors.

Log Format: The traces are recorded by the videos CDN servers. In particular, one trace item is recorded after a video chunk is delivered to a user. Each trace item has the following information: (1) Timestamp, which records the time when the chunk is delivered to the user; (2) User information, including the ID of the requesting user, the IP address of the user; (3) CDN server information, including the IP address of the CDN server; (4) QoS information, including the download time (i.e., time spent on downloading the chunk), download speed, round trip time (RTT) between the user and server; (5) Content information, including the ID of the video requested, video title, video duration and its category, spanning *movie* (3.0%), *news* (5.1%), *series* (55.8%), *sports* (3.2%), *fun* (5.0%), *documentary* (3.1%), *children* (24.4%) and *others* (0.4%). The percentage in brackets represents the content distribution of these videos.

2) Engagement Calculation: In our datasets, we quantify users engagement level based on the streaming completion ratio, which has been confirmed to be effective to model QoE in on-demand video systems [6]. For a particular video session, a session completion ratio can be calculated as the fraction of chunks watched by users over the number of chunks in the video. A larger completion ratio indicates that users have been more engaged in the streaming, instead of leaving sessions at the early stage. In our study, we use the session completion ratio as our engagement measurement index.

Next, we present the QoS and preference factors, which can potentially affect the streaming engagement.

3) QoS Factors: In our study, we calculate the following QoS factors: (1) Average bitrate, which captures the video quality (e.g., resolution), (2) Buffering ratio, which captures how much data is available in users' cache [12], (3) Number of rebuffers, which captures the number of player re-buffeers.

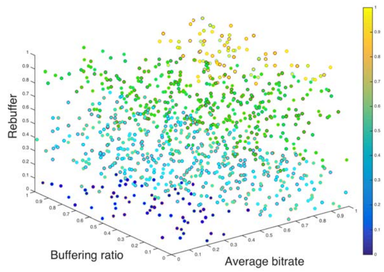

Fig. 2. Samples of QoS factors after normalized.

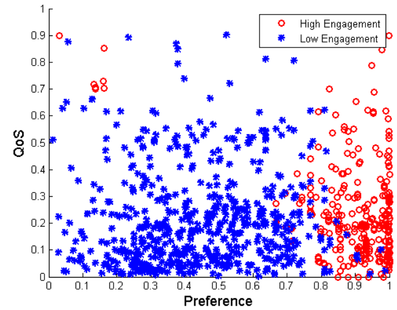

Fig. 3. Engagement is affected by both QoS and preference.

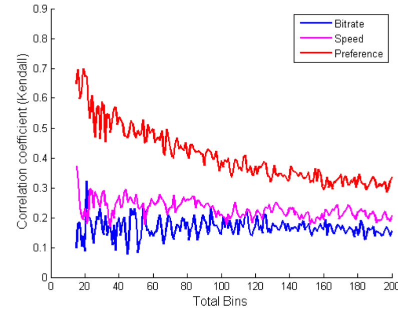

Fig. 4. Importance of different factors.

First, we make each QoS factor normalized. The normalization is calculated as follows:

▷ *Average bitrate*: As the efficiency of bitrate is positive, we normalize average bitrate as follows:

$$f_1(x_{uv}) = \frac{g_1(x_{uv}) - \min g_1(x)}{\max g_1(x) - \min g_1(x)} \quad (1)$$

Let $f_1(x_{uv})$ denote the normalized average bitrate and $g_1(x_{uv})$ denote average bitrate value, where x_{uv} is a video session, which represents a user u requests a video v. Average bitrate of all the video sessions is denoted by $g_1(x)$.

▷ *Buffering ratio*: Although buffering ratio also impacts QoE positively, when it increases to some point, its effectiveness is not obvious. So we normalize buffering ratio as follows:

$$f_2(x_{uv}) = \frac{\ln g_2(x_{uv}) - \min \ln g_2(x)}{\max \ln g_2(x) - \min \ln g_2(x)} \quad (2)$$

Let $f_2(x_{uv})$ denote the normalized buffering ratio and $g_2(x_{uv})$ denote buffering ratio value of a video session. Buffering ratio of all the video sessions is denoted by $g_2(x)$.

▷ *Rebuffer*: Intuitionally, when number of rebuffers is excessive, QoE falls sharply. So we normalize number of rebuffers as follows:

$$f_3(x_{uv}) = exp(-\lambda g_3(x_{uv})) \quad (3)$$

Let $f_3(x_{uv})$ denote the normalized number of rebuffers and $g_3(x_{uv})$ denote number of rebuffers, where x_{uv} is a video session. λ is the weighting parameter.

Then, we try to combine these normalized QoS factors into one unified QoS indicator. Since each QoS factor impacts QoE differently [13], we calculate the unified QoS index by adding QoS factors after weighted, where weighting parameters are calculated from trace fitting according the importance of different QoS factors for video streaming. The unified QoS index is calculated as follows:

$$\mathbf{Q}(x_{uv}) = 0.4 * f_1(x_{uv}) + 0.2 * f_2(x_{uv}) + 0.4 * f_3(x_{uv}) \quad (4)$$

The rationale of $\mathbf{Q}(x_{uv})$ is that a larger $\mathbf{Q}(x_{uv})$ indicates that user u has a better network performance for video streaming.

Finally, we randomly select 2000 video sessions, where QoS factors are normalized, and plot normalized QoS factors distribution in Fig. 2, and our observation is that QoS factors are well-distributed.

4) User Factors: Since engagement is affected by not only the QoS factors, but also user factors, in our study, we particularly investigate the impact of users' preference of video content on their engagement [14], [15]. Users' preference of contents is calculated as follows.

First, we build a content vector for each video. We use a L-dim vector to capture the content feature of a video v: $s_v = (s_{v1}, s_{v2}, \ldots, s_{vL})$. Each entry in this vector represents a content category, e.g., $s_{vk} = 1$ indicates that the video belongs to the k-th category. Based on the labels assigned to these videos by the content analysis (e.g., performing LDA on the descriptions of the videos), we are able to use these vectors to capture the content characteristics of these videos.

Then, we build a user vector for each user according to their historical behaviors. The L-dim user vector for user u is defined as follows: $s_u = (s_{u1}, s_{u2}, \ldots, s_{uL})$, where s_{uk} is the number of videos that are labeled with the kth category, and watched by user u in a recent time window T (e.g., 1 week). In particular, we define a $*$ operator as follows: if a is labeled k, $a * k = 1$, otherwise, $a * k = 0$.

Let $A(u)$ denote the set of videos watched by user u in the recent time window. The k-th entry is then calculated as follows: $s_{uk} = \beta_k \sum_{a \in A(u)} a * b_k$ where β_k is a weighting parameter that captures the importance of different content categories to measure users' preference [16].

Finally, we can measure the "similarity" between a content and a video, to infer the preference level of the user to the content. The similarity is calculated as the cosine distance between the content vector and user vector as follows: $\mathbf{P}_{uv} = \frac{s_u s_v}{|s_u||s_v|}$. A larger $\mathbf{P}_{uv}$ indicates that user u is more likely to like the video content v, since the video v is similar the the ones u has watched [17].

B. Measurement Results

Based on our traces collected, we are able to generate session samples to mine the correlation between engagement and QoS/user factors. In each sample, a whole video session is included, providing the session completion ratio, the bitrate, re-buffer status, and the preference $\mathbf{P}_{uv}$ of the user to the content.

1) Engagement is Affected by both QoS and Preference: We randomly select $2,000$ video sessions, and divide them into high-engagement sessions and low-engagement sessions, according to the session completion ratio, i.e., the session is regarded as a high (resp. low) engagement session if its completion ratio is larger (resp. smaller) than 0.5. Then we investigate how QoS and preference determine the engagement.

In Fig. 3, the two types of sessions are plotted against the unified QoS index (y-axis) and preference (x-axis). Our observations are as follows: (1) In our dataset, a large fraction of sessions are marked low engagement, i.e., about 80% of the selected samples are low-engagement sessions. (2) Different from previous studies [11], [12], we observe that preference has a significant impact on engagement: about 90% of the high engagement sessions have a preference value larger than 0.8.

2) Correlation Comparison: After confirming that the QoS and preference jointly affect users' engagement in video sessions, we further study the importance of different factors. We randomly divide the samples in the previous experiment into different number of groups, ranging from 10 to 200. Then, we calculate the Kendall rank correlation coefficient between the engagement levels and (1) bitrates, (2) download speed and (3) preference indices of sessions across the groups in each setting, respectively. As illustrated in Fig. 4, the three curves represent the Kendall rank correlation coefficient against the three factors, versus the number of groups divided.

Our observations are as follows: (1) The three factors all demonstrate positive correlation with the engagement, confirming that both QoS and preference factors affect the QoE of users; (2) We observe that the Kendall coefficient between engagement and preference is even larger than that the Kendall coefficient between engagement and the other two factors.

Our observations suggest that using the joint QoS and preference factors can help infer user QoE in video streaming. Motivated by this, in the condition that network resource (e.g., bandwidth) is limited and the same amount network resource can generate different amount of QoE satisfaction for different users, we propose to select target users whose QoE is highly affected by network performance instead of preference, and prioritize network resource to improve their engagement, so as to improve the overall QoE of all users in the system.

IV. QoE-Aware Cloud CDN User Mapping Strategy

In this section, we will design a QoE-aware cloud CDN user mapping strategy. We propose a heuristic algorithm to solve the problem.

We first present problem statement. Then we will design a data-driven QoE model, which captures both the QoS factors and user factors that determine a user's quality of experience in a video streaming session. Finally, we give a heuristic algorithm of user mapping.

A. Problem Statement

Engagement can be considered as a function of QoS factors and user preference. That is, we want to capture a relationship $\mathbf{E}_{uv} = F(g_1(x_{uv}), g_2(x_{uv}), g_3(x_{uv}), \mathbf{P}_{uv})$, where $\mathbf{E}_{uv}$ is the session engagement, which can be estimated by the session completion ratio, and each $g_i(x_{uv})$ represents i-th QoS factor, which can be monitored by online statistics function of today's CDNs.

Our objective is to improve the overall QoE of all users in the system. We use $X(u)$ to represent the CDN that serves user u in the mapping strategy X, where $g_i^{X(u)}(x_{uv})$ represents i-th QoS factor that user u gets in a video session. S represents the set of CDNs. We want to maximize the overall QoE of all users, in the condition that the capacity of CDN is within its limitation. Let $C(w)$ denote the capacity of CDN w and $BW(u)$ denote the bandwidth that user u obtains. Specifically, our objective can be formalized as follows:

$$\max \sum_{u} F(g_1^{X(u)}(x_{uv}), g_2^{X(u)}(x_{uv}), g_3^{X(u)}(x_{uv}), \mathbf{P}_{uv}) \quad (5)$$

subject to

$$\sum_{u, X(u)=w} BW(u) \leq C(w) \quad \forall w \in S \quad (6)$$

Next, we will design a data-driven QoE model, which captures both the QoS factors and user factors that determine a user's quality of experience in a video streaming session.

B. QoE Model

We apply classification tree to predict the classification of engagement and regression tree to predict the value of engagement. First, we present the feature vector used in QoE model.

1) Feature Vector: The feature vector is composed by QoS factors and user preference. Each feature vector has the following information: i) *QoS factors*: Average bitrate, download speed and video type; ii) *User preference*: The measurement approach is $\mathbf{P}_{uv} = \frac{s_u s_v}{|s_u||s_v|}$, which is given in Section III;

Next, we define class label of samples, which is a binary divided of engagement.

2) Class label: We apply decision tree to predict classification of engagement. In training dataset, if engagement value of a sample is less than 0.5, classification of sample belongs to low-engagement, otherwise, it belongs to high-engagement.

Finally, we use the decision tree model in matlab statistics tools, which can build model automatically by using $ID3$ algorithm.

C. Prediction Accuracy

We use test dataset to measure the accuracy of QoE models. In Table I, the accuracy rates of classification prediction and value prediction are calculated separately. We compare the effectiveness of our model, which uses the joint QoS factors and preference to infer QoE, with conventional model that only uses QoS factors. Our observations are as follows:

(1)In engagement classification prediction, improvements of accuracy rate are 7.63% and 5.94% for short videos and long videos. (2)In engagement value prediction, improvements of accuracy rate are 5.5% and 4.07% for short videos and long videos.

Next, we will design a heuristic algorithm to guide user mapping in video streaming sessions.

TABLE I
PREDICTION RESULTS

	Classification Prediction(%)		Value Prediction(%)	
	Short Videos	Long Videos	Short Videos	Long Videos
Our model	84.25%	69.06%	90.6%	82.32%
Conventional model	76.62%	63.12%	85.1%	78.25%
Improvement	7.63%	5.94%	5.5%	4.07%

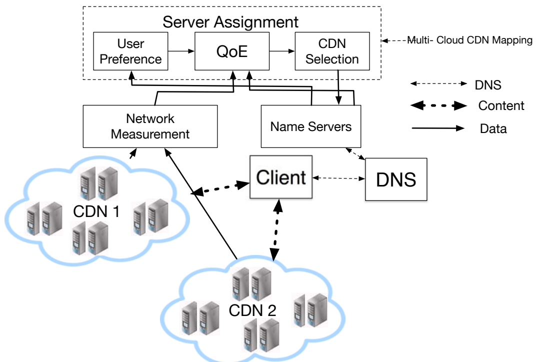

Fig. 5. User mapping architecture of multi-cloud CDN.

D. User Mapping Strategy

We first present the procedure of user mapping strategy. The details of this approach are summarized in Algorithm 1. As illustrated in Fig. 5, our user mapping algorithm works as follows:

(i) Let (u_i, v_{u_i}) denote that the user u_i requests the video v_{u_i}, u_i gets the server IP through DNS. We take the set R to represent all the users who request videos, and parameter K as input.

(ii) Different CDNs serve users randomly.

(iii) For each user u_i, $\mathbf{P}_{u_i v_{u_i}}$ and $\mathbf{E}_{u_i v_{u_i}}$ are calculated, which $\mathbf{P}_{u_i v_{u_i}}$ is denoted by $Pref(u_i, v_{u_i})$. Improvement of $\mathbf{E}_{u_i v_{u_i}}$ is calculated when u_i is redirected to next CDN. Let I_{u_i} denote improvement of $\mathbf{E}_{u_i v_{u_i}}$. Let $g_i'(x_{u_i v_{u_i}})$ denote the i-th QoS factor index of next CDN.

(iv) All the improvements are sorted in descending order.

(v) If the CDN, which user u will be allocated to, is not over loaded, CDN serving user u is adjusted. The bandwidth which user u_i obtains in $X(u_i)$ is denoted by $BW(X(u_i))$. We adjust the top K users mapping in the order of improvement, which is denoted by $Adjust(u_i)$, by using mechanisms including DNS/GSLB to perform the user redirection. And the parameter K, which is set according to the actual number of users, is used to reduce to cost of adjustment, such as 70% percent of the whole users.

This algorithm can actually be implemented in a fully distributed manner, so that each user can be redirected upon arrival.

Algorithm 1: Heuristic selection strategy of CDN.

CDN Selection (R, K) ;
for $(u_i, v_{u_i}) \in R$ **do**
  $X(u_i) \leftarrow Random(CDNs)$;
end
for $(u_i, v_{u_i}) \in R$ **do**
  $\mathbf{P}_{u_i v_{u_i}} \leftarrow Pref(u_i, v_{u_i})$;
  $\mathbf{E}_{u_i v_{u_i}} \leftarrow F(g_i(x_{u_i v_{u_i}}), \mathbf{P}_{u_i v_{u_i}})$;
  $I_{u_i} \leftarrow F(g_i'(x_{u_i v_{u_i}}), \mathbf{P}_{u_i v_{u_i}}) - \mathbf{E}_{u_i v_{u_i}}$;
end
Rank u in the descending order of I_{u_i} ;
for 1 *to* K **do**
  if $BW(X(u_i)) < C(X(u_i))$ **then**
    $X(u_i) \leftarrow Adjust(u_i)$;
  end
end

V. PERFORMANCE EVALUATION

In this section, we will verify the effectiveness of our design. We use the simulation environment to compare three different user mapping strategies.

A. Experiment Setup

In simulation environment, we generate a number of users, videos and CDNs. Each user-CDN pair has a QoS value in $[0, 1]$ assigned, which reflects the QoS the user can obtain from downloading video content from that CDN. We randomly assign the preferences of users to different videos, and users will choose videos to watch randomly following the distribution of their preferences to these videos. Each CDN has a fixed capacity. In each of our experiments, we generate 1000 sessions, in which users access these videos. We compare our design with the following strategies: (1) QoS-based scheme, which only considers QoS factors to map users to CDNs; (2) Random strategy, which randomly maps users to CDNs, serving as a baseline.

Next, we present the experiment results.

B. Experiment Results

1) Performance under Different Number of CDNs: First, we study the relationship between QoE satisfaction and number of CDNs. We set both number of users and number of videos as 100. We vary the number of CDNs between 2 and 15. As illustrated in Fig. 6, the curves illustrate the QoE indices for different strategies, versus the number of scheduling CDNs. Our observations are as follows: (1) Both our strategy and the QoS-based strategy outperform the ramdon baseline significantly; (2) Compared with the QoS-based strategy, our design achieves larger overall QoE gain when the number of CDNs is large. Given that today's video service providers are using more and more CDNs and cloud CDNs, our design can meet the requirement in such many-CDN delivery solution.

2) Performance under Different Number of CDNs: Next, we study the relationship between QoE satisfaction and the number of users. We set number of videos as 100 and number of CDNs as 10. We vary the number of users between 50 and 250. As illustrated in Fig. 7, the curves illustrate the QoE indices for different strategies, versus the number of streaming users. Our observations are as follows: Similarly, our design and the QoS-based scheme have close QoE satisfactory, which is much higher than the pure random scheme.

3) Impact of Preference on QoE: Finally, we study the impact of preference on QoE. We set both the number of

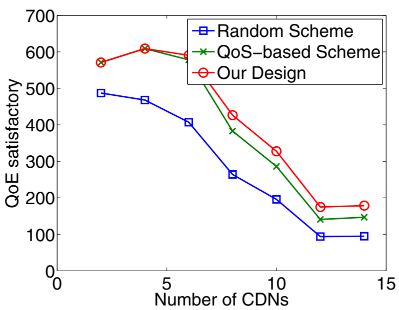

Fig. 6. QoE satisfaction versus the number of CDNs.

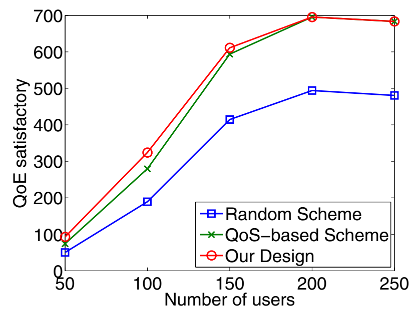

Fig. 7. QoE satisfaction verus the number of users.

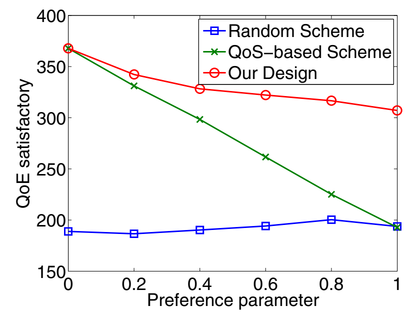

Fig. 8. QoE satisfactory versus the preference impact.

users and the number of videos as 100, and the number of CDNs as 10. Preference parameter (x-axis) is used in our simulation to adjust the impact of preference on QoE, i.e., a larger preference parameter indicates that QoE is more affected by preference instead of QoS. As illustrated in Fig. 8, the curves illustrate the QoE indices for different strategies, versus the preference parameter. We observe that when the preference parameter is 0, our strategy and the QoS-based scheme have similar QoE performance. However, when the value is large, our design can significantly outperform the conventional QoS-based strategy, suggesting that our design can work for today's highly personalized video services.

VI. Conclusions

It has become a norm for video service providers to use the multi-cloud based CDN paradigm in their video content distribution. Mapping different users to different cloud CDN servers is the key to provision satisfactory quality of experience. In this paper, we propose a joint QoS and user preference driven mapping strategy. First, we identify the important QoS factors including download speed, bitrate and caching capacity, and user preference level that jointly determine a user's engagement level in a video session. Second, we build a decision tree to predict the engagement of users given these factors; Third, based on the prediction, we design an optimization framework to improve users' quality of experience by CDN selection. We also verify the effectiveness of our design using simulation experiments. Compared with conventional strategies, our design can identify targeting users whose QoE is mostly affected by CDN selection, and improve the overall QoE of users.

Acknowledgement

This work is supported in part by the National Basic Research Program of China (973) under Grant No. 2015CB352300 and the National Natural Science Foundation of China under Grant No. 61402247.

References

[1] F. Chen, K. Guo, J. Lin, and T. L. Porta, "Intra-cloud lightning: Building cdns in the cloud," in *IEEE INFOCOM*, 2012.

[2] H. Liu, Y. Wang, Y. Yang, A. Tian, and H. Wang, "Optimizing Cost and Performance for Content Multihoming," in *ACM SIGCOMM*, 2012, pp. 371–382.

[3] J. Dilley, B. Maggs, J. Parikh, H. Prokop, R. Sitaraman, and B. Weihl, "Globally distributed content delivery," *Internet Computing, IEEE*, vol. 6, no. 5, pp. 50–58, 2002.

[4] X. Liu, F. Dobrian, H. Milner, J. Jiang, V. Sekar, I. Stoica, and H. Zhang, "A Case for a Coordinated Internet Video Control Plane," in *ACM SIGCOMM*, 2012.

[5] Z. Wang, W. Zhu, M. Chen, L. Sun, and S. Yang, "CPCDN: Content Delivery Powered by Context and User Intelligence," *IEEE Transactions on Multimedia*, vol. 17, no. 1, pp. 92–103, January 2015.

[6] 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," in *Proceedings of the ACM SIGCOMM 2011 Conference*, ser. SIGCOMM '11. New York, NY, USA: ACM, 2011, pp. 362–373.

[7] H. H. Liu, Y. Wang, Y. R. Yang, H. Wang, and C. Tian, "Optimizing cost and performance for content multihoming," in *Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication*. ACM, 2012, pp. 371–382.

[8] Z. Wang, L. Sun, C. Wu, W. Zhu, and S. Yang, "Joint Online Transcoding and Geo-distributed Delivery for Dynamic Adaptive Streaming," in *IEEE International Conference on Computer Communications (INFOCOM)*, 2014.

[9] H. H. Song, Z. Ge, A. Mahimkar, J. Wang, J. Yates, Y. Zhang, A. Basso, and M. Chen, "Q-score: proactive service quality assessment in a large iptv system," in *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. ACM, 2011, pp. 195–208.

[10] R. K. Mok, E. W. Chan, X. Luo, and R. K. Chang, "Inferring the qoe of http video streaming from user-viewing activities," in *Proceedings of the first ACM SIGCOMM workshop on Measurements up the stack*. ACM, 2011, pp. 31–36.

[11] 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.

[12] 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 *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4. ACM, 2013, pp. 339–350.

[13] 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*. ACM, 2012, pp. 97–108.

[14] P. J. Guo, J. Kim, and R. Rubin, "How video production affects student engagement: An empirical study of mooc videos," in *Proceedings of the first ACM conference on Learning@ scale conference*. ACM, 2014, pp. 41–50.

[15] C. Ilioudi, M. N. Giannakos, and K. Chorianopoulos, "Investigating differences among the commonly used video lecture styles," 2013.

[16] S. Y. Jung, J.-H. Hong, and T.-S. Kim, "A statistical model for user preference," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 17, no. 6, pp. 834–843, 2005.

[17] A. Huang, "Similarity measures for text document clustering," in *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand*, 2008, pp. 49–56.