Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017)

A First Look at User Switching Behaviors Over Multiple Video Content Providers

Huan Yan, Tzu-Heng Lin
Department of Electronic Engineering,
Tsinghua University, China

Gang Wang
Department of Computer Science,
Virginia Tech, Blacksburg, VA

Yong Li
Department of Electronic Engineering,
Tsinghua University, China

Haitao Zheng
Department of Computer Science,
UC Santa Barbara, Santa Barbara, CA

Depeng Jin
Department of Electronic Engineering,
Tsinghua University, China

Ben Y. Zhao
Department of Computer Science,
UC Santa Barbara, Santa Barbara, CA

Abstract

Watching videos from multiple content providers (CP) has become prevalent. For individual CPs, understanding user video consumption patterns among CPs is critical for improving on-site user experience and CP's opportunity of success. In this paper, based on a two-month dataset recording 9 million users' 269 million video viewing requests over 6 most popular video CPs in China, we provide a first look at users' video consumption and switching behaviors across different CPs. We find that rarely do active video watchers stick to one single CP. When users switch from one CP to another, they are more likely to switch to larger CPs with more videos.

Introduction

Video watching has become one of the most popular online activities and led to an enormous market with various video content providers (CP). On one hand, the development of the broadband services greatly improves the user experience of streaming videos (*e.g.*, Netflix), the traffic of which has already accounted for more than half of the peak time downstream bandwidth in North America (Karamshuk et al. 2015). On the other hand, the wide spread of smart devices and the fast growth of social networking also lead to more user uploaded videos (*e.g.*, YouTube) to be shared over the Internet (Scellato et al. 2011).

In recent years, various video content providers have formed a giant ecosystem, where all of them seek to attract more users to maximize revenue. To succeed or even survive in this ecosystem, each CP strives to provide the best services, *i.e.*, by putting significant efforts to design video recommendation mechanisms or build faster content delivery infrastructures to improve user video viewing experience.

One of the prerequisites underlying the solutions to improve user experience is to thoroughly understand users' needs with their consumption patterns. In the context of video consumption, data-driven user behavior analysis is popular. However, most previous analyses are based on data collected from a single content provider or focused on one specific context (Karamshuk et al. 2015; Li et al. 2015; 2014; 2012; Dobrian et al. 2011; Gopalakrishnan et al. 2011; Yin et al. 2009; Cha et al. 2008; Huang et al. 2008; Yu et al. 2006). Given the broad differences of the video contents and the marketing strategies of different content providers, it is critical to study user video consumption patterns of different providers in the same picture.

In this paper, we seek to provide an initial view of user switching behavior in video consumption across different content providers. We do so by analyzing a large-scale dataset that covers user video watching logs at 6 most popular video content providers in China including Youku, IQiyi, Sohu, Kankan, LeTV and Tencent Video. Our dataset is obtained from one of the major ISPs in China. The dataset contains the complete 269 million video viewing request records from 9 million users in the city of Shanghai (China) from November 1 to December 31 in 2015. We analyze the data seeking to understand how users spread their attention across different content providers, and how users "switch" from one provider to another for video consumption.

To the best of our knowledge, our study is the first to systematically analyze user video consumption across different CPs and the corresponding provider switching behaviors. The results from our month-scale dataset indicate that user video consumption is highly dependent on the content types and CP popularity. In addition, the unique characteristics of CPs, such as their close connection to online social networks are also playing an important role. Our high-level findings can be summarized as the follows:

- First, for regular video watchers, users are likely to switch across multiple content providers and spread their attention evenly. This suggests the evidence for multiple video content providers to co-exist.
- Second, users are more likely to watch the next video within the same content provider, rather than switch providers directly. When users do switch providers, they are more likely to switch to larger (more popular) ones who have richer video content.

Our analysis serves as a first step towards understanding the complex user behaviors within the ecosystem of video content providers. In this paper, we first describe our dataset and metrics. Then, we discuss user switching behaviors across multiple CPs. Finally, we conclude the paper.

Dataset and Metrics

Dataset. Our dataset is obtained from a major ISP net-

Category	# Views (10^6)	# Videos (10^6)
TV Series	112.6 (41.8%)	0.7 (8.4%)
Show	39.4 (14.7%)	1.0 (11.2%)
Movie	24.2 (9.0%)	0.2 (2.7%)
Cartoon	8.7 (3.2%)	0.2 (2.8%)
News	8.8 (3.3%)	0.3 (4.4%)
UGV	4.7 (1.7%)	0.4 (4.3%)
Others	70.6 (26.3%)	6.0 (66.2%)

Table 1: # of Videos and views per category. The numbers are displayed in millions.

work in China, and consists the complete video viewing logs of 9,517,339 users in Shanghai City from November 1 to December 31 in 2015. The dataset covers 6 largest video content providers in China: Youku (YK), IQiyi (IQI), Sohu (SH), Kankan (KK), LeTV (LE), and Tencent Video (TC). We collect video viewing requests to these video content providers through deep packet inspection appliances at the gateways of the ISP. Each viewing request is characterized by user ID, timestamp and request URL. The user ID is produced by the ISP based on device-level information to map to a user device (not IP). To obtain the detailed information about the video (*e.g.,* video category), we use a web crawler to fetch the video URLs.

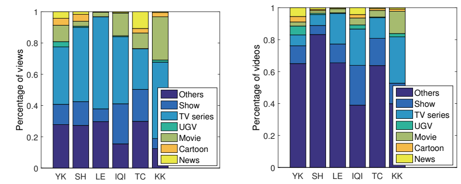
In total, our dataset contains 269 million viewing requests from 9,517,339 users who watched over 9,055,188 videos from the six content providers. Note that this ISP network has an 85% of market share for the broadband access in China, and thus our data can provide a highly comprehensive view of these users' video viewing behaviors at major content providers.

Ethics. Our study seeks to provide a better understanding of user video watching behaviors across content providers. The high-level goal is to help content providers to improve service quality for better user experience. Our dataset is collected via our collaboration with ISP, and the data does not contain personally identifiable information. The "user ID" field has been anonymized (as a bit string) and does not contain any user meta data.

Metrics. Our goal is to understand user video consumption across different CPs. We seek to answer two lines of questions. First, how do users spread their attention across different providers? Do they stick to one site or more likely to spread their views evenly? Second, how often do users switch from one provider to another? When switching, which provider/video categories are users more likely to switch to?

In the following, we describe our analysis to answer the above questions. We design a series of metrics to evaluate user video viewing and switching.

We first introduce the key notations for users' video viewing. We denote v_i^k as the total views in CP k from user i. The total number of views $V_i = \sum_{k=1}^{K} v_i^k$ $(1 \leq i \leq M)$ where M is the total number of users and K is the total number of CPs. The length of viewing sequence is N_i. Finally, we denote $s_i^{k,k'}$ as the total number of times when user i switches from CP k to CP k'.



(a) Views (b) Videos

Figure 1: Distribution of videos and views by categories per content provider.

Content Provider	YK	SH	LE	IQI	TC	KK
# Views (10^6)	116	55	34	29	25	10
# Users (10^6)	5.4	3.9	3.2	3.6	3.7	1.5
# Videos (10^6)	6.6	1.2	0.3	0.3	0.4	0.1
P2P service	N	N	N	N	N	Y
Social networks	N	N	N	N	Y	N

Table 2: Statistics of the 6 content providers. The numbers are displayed in millions.

To understand the overall user video viewing and switching, we define three metrics to drive our analysis:

- **View Entropy** measures how evenly do users' views distributed among different CPs.

$$W_i = -\sum_{k=1}^{K} \frac{v_i^k}{V_i} log_2 \frac{v_i^k}{V_i}, \quad (1)$$

where a higher value indicates more uniformly distributed views among CPs.

- **Cross-site Switching Frequency** measures how frequently a user switches between different CPs.

$$F_i = \frac{\sum_{k=1}^{K} \sum_{k'=1, k' \neq k}^{K} s_i^{k,k'}}{N_i - 1} \quad (2)$$

Its value ranges from 0 to 1. In particular, $F_i = 0$ indicates the user only watch videos in a single CP.

- **CP Switching Probability** measures how likely a user switches from one CP to another. For the probability of user i to switch from CP k to k':

$$P_{k,k'} = \frac{\sum_{i=1}^{M} s_i^{k,k'}}{\sum_{k'=1}^{K} \sum_{i=1}^{M} s_i^{k,k'}} \quad (3)$$

Switching Behavior Analysis

In this section, we present our analysis results following the proposed metrics. We analyze the overall user video viewing and switching behaviors by examining these users in a single population.

Preliminary Analysis. Videos are can be roughly classified into 6 major categories including "TV series", "Show", "Movie", "Cartoon", "News", "User-generated videos (UGV)" based on resolved URLs. Some videos have defunct URLs or have no category information, and we put them under "Others." Table 1 shows the number of videos and views in each category. The most popular category is

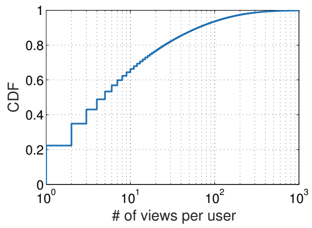
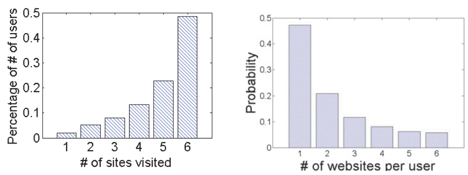


Figure 2: Number of total video views per user.



(a) Users with 100+ Views (b) All Users

Figure 3: Number of CPs visited per user.

TV Series which has attracted 41.8% of the total views with only 8.4% of total videos. Note that the "Others" category, even though takes more than half of the videos, only attracts 20% of the views. Thus it shouldn't impact our later conclusions.

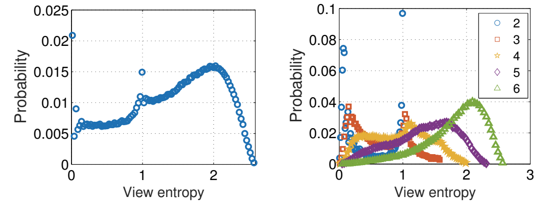
Despite the common video categories shared by these content providers (CP), they each have different emphasis and unique features. YK is the most popular site with the most videos (6.6 million), views (116 million) and users (5.4 million) as shown in Table 2. YK is also known for its rich categories of videos. In contrast, the other 5 smaller CPs are more specialized in providing certain types of videos. For instance, SH, LE and IQI are well known for providing the latest movies, dramas and variety shows. This is reflected in Figure 1 as these three categories ("Movies", "TV series" and "Shows") are the most viewed categories.

TC's unique feature is the connection to China's largest mobile social network (Tencent QQ). TC also serves as a major news portal, and news stories are pushed to end-users through the social network. The impact is clear: even though "News" videos only take less than 1% of all TC videos (Figure 1(b)), it has successfully drawn more than 10% of total views in TC (Figure 1(b)). A counter example is IQI (with no social network). IQI also provides "News" (5%) but draws less than 1% views.

KK started its business by providing P2P downloading services, and then expanded as a video streaming service. Other than "TV series" that all CPs are commonly providing, KK's is specialized at "Movie" content.

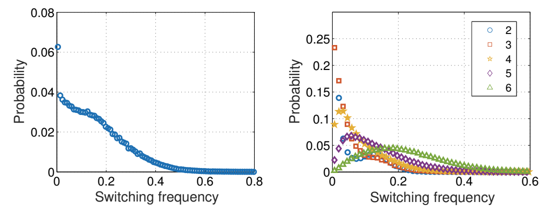
Viewing and Switching Behaviors

Data Preprocessing. To produce statistically meaningful result, we focus on users that have sufficient number of views in our dataset. As shown in Figure 2, 20% users have only one video watching event, and 90% users have less than 100. To this end, our analysis primarily focuses on users who have 100+ views over the two-month period (if not otherwise stated). This leaves us 612,760 users who have contributed 160 million video views.



(a) Overall Distribution (b) Breakdown by # of CPs

Figure 4: View entropy of users: a breakdown by the total number of sites a user visited.



(a) Overall Distribution (b) Breakdown by # of CPs

Figure 5: Switching frequency of users: a breakdown by the total number of sites a user visited.

User View Preferences (Entropy). First, we examine how often users visit multiple CPs and how evenly their views spread among CPs. Figure 3(a) shows selected users (with 100+ views) are broadly using multiple CPs to watch videos. Nearly 50% of them visit all 6 CPs and only 2% would stick to one single CP. This suggests for regular video watchers, they have the need to use multiple CPs to access video content. In contrast, for occasional video watchers, they visit less CPs or even just one (Figure 3(b)).

Figure 4(a) shows the distribution of user *view entropy* metric. This distribution is skewed towards right, indicating users tend to spread their views relatively evenly among different CPs. Figure 4(b) shows that the more CPs that users visit, the more likely their views spread evenly among different CPs. These results further confirm watching videos across multiple CPs is prevalent, and thus it makes sense for many CPs to co-exist in the ecosystem.

Cross-site Switching Frequency. Figure 5(a) shows the distribution of user's *switching frequency* across content providers. We find most users' switching frequency is below 0.5, indicating switching videos across sites is not as frequent as switching within the same site. This suggests that, even though users commonly watch videos from different CPs (Figure 4), they would watch multiple videos from one CP before switching to another CP. This suggests that even with multiple CPs co-existing, it is still possible for individual CPs to engage their users for consecutive video watching within the site. Figure 5(b) shows cross-site switching is more common among users who use more CPs.

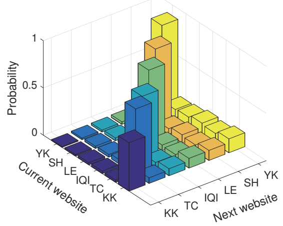


Figure 6: Switching probability between two CPs.

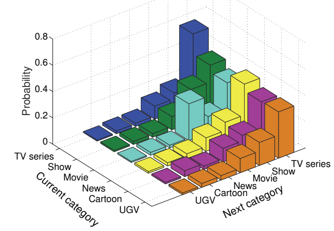


Figure 7: Video categories of cross-site switching probability.

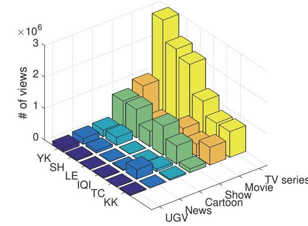


Figure 8: Video categories that users switch to on the target CP.

CP Switching Probability. A critical question for content providers is, when users leave their site to a new site, where do users go and why they leave. We analyze the *switching probability* across CPs in Figure 6. Consistent with earlier observations, it is more often for users to switch video within the same site than cross-site (the bars are highest along the diagonal). For cross-site switching, we observe that YK is the CP that users are most likely to switch to, followed by SH, LE, IQI, TC and KK. This order is relatively consistent with the size of the site (Table 2). This indicates users are more likely to switch to larger site with richer choice of videos.

In addition, we seek to understand whether the category of videos plays a role in the cross-site switching behaviors. As shown in Figure 7, the most dominating trend is that users would switch to more popular video categories such as "TV" or "Show" during the cross-site switching. On top of that, we also observe some users would switch site for the same category of the videos (the bars along diagonal are slightly higher than the nearby bars, *e.g.*, "Movie" and "Cartoon").

Finally, we examine what categories users switch to when switching to a particular site. As shown in Figure 8, for "TV", "Movie" and "Show", the largest site YK still has a dominating advantage over the other sites. However, we do observe that smaller CPs' unique features help them to draw user views during user migration. For instance: TC, with the help of its social network to push "News" videos to users, has receive more views on "News" after user switching compared to other sites. SH is currently doing well in "Cartoon" category; Despite the smaller total video collection in IQI and KK, they follow right after YK on "Movie" views from switching due to their emphasis on Movie content.

Based on our analysis, we make two key observations. First, there is no single CP can fully meet all users needs, and thus multiple CPs can co-exist. Second, when users do switch providers, they are more likely to switch to larger providers (e.g., YK) for the popular videos (e.g., TV series).

Conclusion

To the best of our knowledge, this is the first to systematically analyze user video consumption and switching behaviors across different content providers. We concentrate on uncovering the overall patterns of how users switch from one CP to another (and possible reasons). Our analysis serves a first step towards understanding the complex user behaviors in ecosystem of numerous content providers. Our findings also provide key insights for CPs to improve their services and better engage users. As future work, we plan to further investigate key factors that trigger user migration over video content providers. In addition, we are interested in modeling the dynamic user behavior changes over time.

References

Cha, M.; Rodriguez, P.; Crowcroft, J.; Moon, S.; and Amatriain, X. 2008. Watching television over an ip network. In *Proceedings of IMC*, 71–84.

Dobrian, F.; Sekar, V.; Awan, A.; Stoica, I.; Joseph, D.; Ganjam, A.; Zhan, J.; and Zhang, H. 2011. Understanding the impact of video quality on user engagement. *ACM SIGCOMM Computer Communication Review* 41(4):362–373.

Gopalakrishnan, V.; Jana, R.; Ramakrishnan, K.; Swayne, D. F.; and Vaishampayan, V. A. 2011. Understanding couch potatoes: measurement and modeling of interactive usage of iptv at large scale. In *Proceedings of IMC*, 225–242.

Huang, Y.; Fu, T. Z.; Chiu, D.-M.; Lui, J.; and Huang, C. 2008. Challenges, design and analysis of a large-scale p2p-vod system. In *ACM SIGCOMM Computer Communication Review*, 375–388.

Karamshuk, D.; Sastry, N.; Secker, A.; and Chandaria, J. 2015. On factors affecting the usage and adoption of a nation-wide tv streaming service. In *Proceedings of INFOCOM*, 837–845.

Li, Z.; Lin, J.; Akodjenou, M.-I.; Xie, G.; Kaafar, M. A.; Jin, Y.; and Peng, G. 2012. Watching videos from everywhere: a study of the pptv mobile vod system. In *Proceedings of IMC*, 185–198.

Li, Z.; Xie, G.; Lin, J.; Jin, Y.; Kaafar, M.-A.; et al. 2014. On the geographic patterns of a large-scale mobile video-on-demand system. In *Proceedings of INFOCOM*, 397–405.

Li, Z.; Xie, G.; Kaafar, M. A.; and Salamatian, K. 2015. User behavior characterization of a large-scale mobile live streaming system. In *Proceedings of WWW*, 307–313.

Scellato, S.; Mascolo, C.; Musolesi, M.; and Crowcroft, J. 2011. Track globally, deliver locally: improving content delivery networks by tracking geographic social cascades. In *Proceedings of WWW*, 457–466.

Yin, H.; Liu, X.; Qiu, F.; Xia, N.; Lin, C.; Zhang, H.; Sekar, V.; and Min, G. 2009. Inside the bird's nest: measurements of large-scale live vod from the 2008 olympics. In *Proceedings of IMC*, 442–455.

Yu, H.; Zheng, D.; Zhao, B. Y.; and Zheng, W. 2006. Understanding user behavior in large-scale video-on-demand systems. In *ACM SIGOPS Operating Systems Review*, 333–344.