User Behavior Characterization of a Large-scale Mobile Live Streaming System

Zhenyu Li†, Mohamed Ali Kaafar⋆, Kavé Salamatian‡, Gaogang Xie†

†ICT-CAS, China, ⋆National ICT Australia, ‡Université de Savoie, France

{zyli, xie}@ict.ac.cn, Dali.Kaafar@nicta.com.au, kave.salamatian@univ-savoie.fr

ABSTRACT

Streaming live content to mobile terminals has become prevalent. While there are extensive measurement studies of non-mobile live streaming (and in particular P2P live streaming) and video-on-demand (both mobile and non-mobile), user behavior in *mobile live streaming* systems is yet to be explored. This paper relies on over 4 million access logs collected from the PPTV live streaming system to study the viewing behavior and user activity pattern, with emphasis on the discrepancies that might exist when users access the live streaming system catalog from mobile and non-mobile terminals. We observe high rates of abandoned viewing sessions for mobile users and identify different reasons of that behavior for 3G- and WiFi-based views. We further examine the structure of abandoned sessions due to connection performance issues from the perspectives of time of day and mobile device types. To understand the user pattern, we analyze user activity distribution, user geographical distribution as well as user arrival/departure rates.

Categories and Subject Descriptors

C.2.4 [**Computer Applications**]: Distributed applications

Keywords

Mobile live streaming; viewing behavior; user activity

1. INTRODUCTION

It has been recently reported that in 2013 mobile video traffic account for more than 50% of total mobile traffic [1]. This share is expected to increase to 2/3 by 2018. This makes the understanding of watching behavior of video content, both in Video-on-Demand (VoD) and live streaming systems, a major issue for content providers, delivery networks and all actors in wireless Internet. In particular, the past year has witnessed the emergence of live streaming, which involves streaming of highly popular events like sport, artistic, cultural or political events, as well as highly popular TV programs.

Live streaming is vastly different from VoD as it is event-driven and real-time streaming based, and as such one cannot easily extend findings of VoD studies to event-based streaming. Likewise, even though live streaming systems on wired links (non-mobile), and in particular Peer-to-Peer (P2P) have been intensively studied in the past years [5, 14, 17], the mobile nature of communication and users may result in different behavior pattern. On the other hand, while several studies have looked at user behavior in mobile VoD and mobile Internet TV systems [8, 9, 10], these were not on event-based streaming. Our object in this paper is to investigate whether live content consumption behavior is different between mobile and non-mobile users. We will also study how mobile users react to connectivity and performance issues that are more frequent in mobile platforms. In particular, we aim to study the effect of time and the mobile device type on viewing problem.

Our work is based on observations from the PPTV mobile live streaming service[1]. PPTV is providing a video reception software platform that can run both on mobile phone or tablet under iOS or Android operating systems, or on traditional computers (both laptop or desktop). PPTV considers people accessing video on their phones or tablets as mobile users, while users watching videos over traditional computers are coined as "non-mobile". Even though WiFi users might not be strictly considered as mobile, we resort to the use of PPTV terminology: mobile users for people accessing videos on their phones or tablets and non-mobile for people accessing it on traditional computers. We nevertheless consider difference between "mobile" users connected to Internet via 3G and "mobile" users connected through WiFi.

For this purpose, we have gathered two-weeks of access logs from both mobile and non-mobile users viewing PPTV live streaming. This dataset allows us to observe the main discrepancies of users behavior that might exist when accessing live content from mobile and non-mobile terminals. We focus on the viewing duration, viewing abandonment rate and user activity. In addition, we characterize poor-performance viewing sessions and analyze user behavior when experiencing such viewing issues. The findings could help content providers, content delivery networks as well as mobile live streaming APP designers for system design and optimization. To sum up, we make two main contributions.

First, we perform an in-depth analysis of user viewing behavior in the PPTV live streaming system. We observe

[1]previously known as PPLive

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2743054.

a high abandonment rate of mobile views (especially for 3G views) and identify that 3G views and WiFi views are abandoned for different reasons: 3G views are abandoned mostly because of poor wireless connection performance, while loss of interests and connection problem contribute almost equally to the WiFi views abandonment. We also identify that the ratio of mobile sessions that suffer from poor connection performance varies over time of day and channels. In particular, we perform a QED (Quasi-Experiment Design) analysis, which reveals that the mobile device type has a notable impact on viewing abandonment rate.

Second, we make a comprehensive comparison of the viewing pattern between mobile and non-mobile users. We find that although user activity for both mobile and non-mobile live streaming follows the Pareto principle ("a minor proportion of causes generate a major proportion of effects"), the activity distribution for mobile one is more skewed due to the diversity of device types and connection performance. In comparison with non-mobile users, mobile users also show a more uniform geographical distribution for individual channels. Besides, mobile live streaming has much higher user dynamics as mobile users when disconnected from the streaming often try to immediately reconnect.

The remainder of this paper is as follows. Section 2 describes the dataset in use, while Section 3 characterizes the temporal access trend and watching view duration. In Section 4, we examine the user activity. Section 5 surveys related work and Section 6 concludes the paper.

2. DATASET DESCRIPTION

Our study is based on access logs of the popular live streaming service PPlive [2]. In order to have a comparative study of user behavior, we collected two datasets. The first dataset, referred as `mobile` dataset, consists of 4,887,195 access logs from April 1st to 14th, 2013 of mobile devices (smartphones and tablets) running PPTV's mobile app. In this dataset, 12.8% of the accesses are connections through cellular network (*i.e.* 3G)[2], while the rest represents WiFi connections. The second dataset, referred as `non-mobile` dataset, consists of 4,519,512 access logs during the same observation period, from traditional computers (laptops or desktops) using the PPTV desktop software client.

Each log contains the view starting time (in seconds, formatted as GMT+8), the duration of the view (in seconds, excluding pause, buffering and joining delay), the geographic location (at provincial level) where the view was from, the channel unique ID and the viewer unique ID. The geographic locations include 34 provincial regions. In total, there are 82 unique channels in PPLive, 68 of which are available for mobile live streaming, mainly sport events and TV programs. Regardless of the channel type, all programs are non-stop streaming channels. A mobile access log contains two additional fields: the connection type (*i.e.* WiFi or 3G) and terminal type (*e.g.* iPhone, iPad). We find that more than 99.9% mobile access logs are coming from four types of terminals: iPhone, iPad, Android Phone (aPhone for short) and Android Pad (aPad for short).

PPTV uses a hybrid Content Delivery Network (CDN) architecture consisting of dedicated delivery servers and a P2P structure for live streaming content. The non-mobile clients (regardless of whether they are connected via WiFi or wired connection) and mobile devices using WiFi connection might be involved in the P2P transferring. However, our datasets do not contain the statistics of P2P chunk transferring.

[2]LTE technology is not yet widely deployed in China at the time of the logs.

3. USER VIEWING BEHAVIOR

In this section, we first examine the viewing duration distribution, and then proceed to an in-depth analysis on the identified "problematic" mobile viewing sessions.

3.1 Viewing Duration

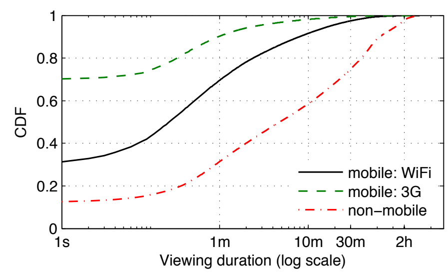


Figure 1: Viewing duration for views from non-mobile and mobile terminals using WiFi and 3G connection

Figure 1 depicts a semi-log cumulative distribution of the viewing duration. In order to examine the impact of mobile connection type on the viewing time, we divide the mobile views into 3G and WiFi connections, and show them separately. The figure shows that mobile views last much shorter than non-mobile ones. For example, while up to 40% of non-mobile views last more than 10 minutes, only 10% of WiFi views and 2% of 3G views last longer than 10 minutes. Notably, about 70% of 3G views and 30% of WiFi views were abandoned before playback. Such a short viewing duration for mobile views makes the design of P2P live streaming more challenging because of high churns. While shorter viewing time has been also observed in mobile VoD [9], we particularly notice that the number of abandoned views before playback in mobile live streaming is much higher. A possible reason is that while VoD content can be optimally replicated on the CDN servers close to users, such optimization is not applicable to live streaming content due to the nature of real-time content.

The shorter duration of mobile views can be explained by several factors. First, most of mobile devices do not well support multitasking, meaning that when users switch between apps to answer a call or to browse a web page, the mobile streaming is stopped. Moreover, mobile users are more careful about their mobile terminal usage both from battery perspective and from traffic volume usage especially over 3G where a cap is generally applied on data traffic. Finally, the discomfort of a small screen might be among the causes that explain the shorter viewing time of mobile users.

In order to optimize video quality [6], content providers (such as PPTV and alike) could benefit from a deeper understanding of the causes of abandoning viewing sessions before playback, especially the ones that are caused by connectivity issues. Next, we look into what we call *problematic sessions*, *i.e.*, viewing sessions aborted before playback because of connection performance problem.

3.2 Identifying Problematic Sessions

Aborting a video session can be broadly caused by either loss of interest or poor connectivity. Unfortunately, our dataset does not provide statistics of connection performance. We therefore resort to a heuristic to detect problematic sessions. We postulate that when a user is really interested in viewing a live streaming channel and it has been impacted by poor connectivity, he would retry multiple times to connect to the channel during a short time window. We thus mark a session v abandoned before playback as a *problematic session* if there is at least another request from the same user on the same channel within a time window of T minutes after v's request.

The time window value T should be set close to the users' patience toward startup delay: T_d, the time the user waits for before sending further requests to the video server. Krishnan *et al.* [7] have shown that users with wireless connections are more patient than others toward startup delay and about 30% of mobile views are not abandoned even with $T_d = 1$ minute. We therefore set $T = 2$ minutes in order to account for the patience of mobile users [3]. We acknowledge that such an identification method might underestimate the number of sessions abandoned due to connection performance, as a session abandoned due to seriously poor connection and thus not followed by another request within T would be identified as a non-problematic session. However, it provides us a pragmatic way for identifying problematic sessions using only information available in our dataset.

Table 1: Aborted sessions statistics

	sessions	abandoned sess.	problem. sess.
WiFi (mob.)	4,261,100	1,300,533	558,775
3G (mob.)	626,095	437,728	329,831
non-mobile	4,519,512	556,634	120,836

Table 1 summarizes the numbers of sessions, abandoned sessions and the identified problematic sessions. We observe that more than 43% of the abandoned WiFi views are problematic, while this number increases to 75% for 3G views. This is expected as 3G connections are more likely to suffer from low performance and high network delays. In other words, the dominant reason for 3G viewers to abandon sessions is connection performance problem, while WiFi users might abort sessions because of loss of interests or connection problems in almost equal share. We also observe that about 20% of 0-duration viewing sessions from non-mobile clients are also problematic, possibly due to network and content servers issues.

3.3 User's Patience when Suffering Problematic Sessions

One of the aspects that content providers are interested in is the patience of users when suffering problematic sessions. To study this, we measure the probability of giving up when suffering x continuous problematic sessions as $G(x)/(G(x)+S(x))$, where $G(x)$ is the number of users that give up a live streaming session after experiencing less than x continuous problematic sessions, and $S(x)$ is the number of users that still try to connect to the system after experiencing at least x continuous problematic sessions.

[3]It is noteworthy that the threshold $T = 2$ might not be an ideal choice. One can tailor this value if he has access to more statistics, such as connection performance statistics and startup delays that mobile users could tolerate.

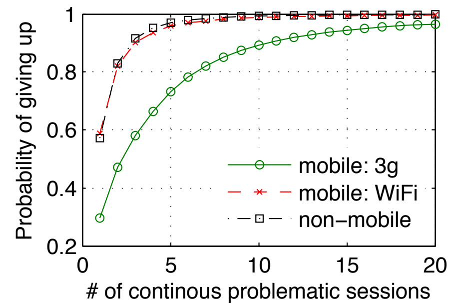


Figure 2: probability of giving up when suffering x continuous problematic sessions

Figure 2 plots the probability of giving up sessions when varying x. Mobile users with WiFi connection have shown comparable user patience than non-mobile users when suffering continuous problematic sessions: the giving up probability is as high as 0.8 when suffering 2 continuous problematic sessions and almost all users would give up after suffering 5 continuous problematic sessions. The mobile users with 3G connection on the other hand seem to be significantly more patient: users would give up with less than 0.5 probability when suffering 2 continuous problematic sessions and in some cases users could even tolerate up to 10 continuous problematic sessions. It seems that 3G users are much aware of the possible poor performance of 3G network. Another notable observation from Figure 2 is that the probability of giving up grows sharply when $x \leq 5$, implying a higher marginal benefit for content providers to reduce the number of continuous problematic sessions.

3.4 Problematic Sessions Characteristics

Understanding the temporal and device dependent characteristics of problematic sessions is important for content providers and ISPs. We define the *problematic session ratio* of a channel (all channels, resp.) in a given time interval J as the proportion of problematic sessions in all sessions of the channel (all channels, resp.) during the time period J. The period length J is set to 10 minutes in our analysis and we restrict our analysis to mobile live streaming (as from table 1 non-mobile streaming has a limited proportion of problematic sessions). In order to be exhaustive, the analysis should be done by stratifying the mobile users in between WiFi and 3G users. Because of lack of space we cannot do this complete analysis here. We therefore present the analysis without differentiating between 3G and WiFi mobile users. However this differentiation is considered when we analyze the impact of device type in Section 3.4.3.

3.4.1 Temporal Effect

We first investigate whether the problematic session ratio over all channels relates to the time of day. Figure 3 plots the average, median, minimum and maximum ratio during the day with a 10-minute observation granularity. The right y-axis shows the number of views in 10-minute intervals. The average and median ratios fluctuate between 0.15 to 0.25. We observe that the problematic session ratio increases from 5AM and reaches its peak at 6AM, and then stabilizes at around 0.2 during the rest of day. This is surprising as we

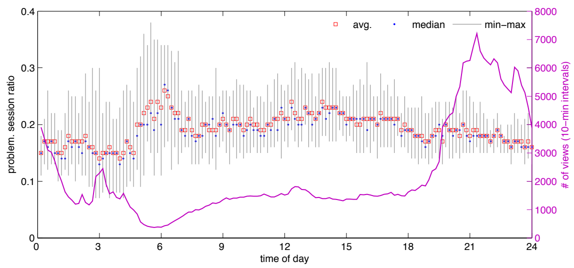


Figure 3: Temporal effect on problematic session ratio

expected a higher problematic session ratio during 8PM to 11PM when the total viewing workload reaches its peak of day. One possible reason is that during peak time there are more users connected so that the P2P component of PPlive live streaming [5] becomes more efficient as it is easier to find available peers to help in the streaming.

On the other side, the increase in the problematic session ratio from 5AM to 6AM is most likely due to the increase of viewing shares from overseas countries. We find that the proportion of views from overseas countries during this period of time is 2 times larger than other time periods, which can be explained by the time zone difference between China and US where a large proportion of overseas users are located. In fact, overseas countries suffer a higher problematic session ratio because neither dedicated CDN servers nor P2P are deployed in these regions.

3.4.2 Channel Popularity Effect

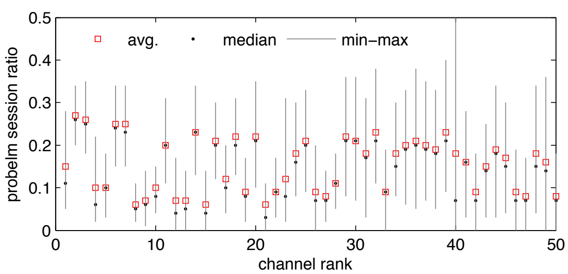


Figure 4: Channel popularity effect on problematic session ratio

Figure 4 depicts the problematic session ratio of each channel versus its popularity (measured in number of views) rank, with a decreasing rank ordering. From the figure it is hard to get into a clear relationship between popularity and problematic sessions ratio. Nonetheless, by grouping channels into two categories: channels with an average ratio around 0.2 and channels with average ration around 0.1, we can have a better view. When we look at channels with a ratio around 0.1, we observe that most of these channels are sport streaming channels, *e.g.* 11 channels among the top 20 popular channels are with an average ratio around 0.1, and 10 of these 11 channels are sport streaming channels with only 1 channel being TV program. This observation is consistent with what was observed in Figure 3, as sport events are often happening in early morning around 3AM or in the evening, during which the problematic session ratio is low. Another possible reason is that sport channels often attract a larger number of users which benefit the P2P streaming.

3.4.3 Mobile Device Type Impact

Table 2: Device impact on problematic session ratio

device type	% of views	avg. problematic ratio
iPad	18.6%	0.07
aPad	11.1%	0.11
iPhone	15.5%	0.14
aPhone	54.7%	0.23

Table 3 summarizes the problematic session ratio information derived over the four main types of device that account for 99.9% of the total views. It can be seen that more than 50% of mobile streaming views are coming from Android phones. Moreover, Android phones and tablets are more likely to experience problematic sessions than iOS based devices. In order to further investigate this observation and to rule out any bias, we implemented a non-parametric factorial analysis using a Quasi Experimental Design (QED) [7]. This approach is used to assess the objective impact of a cause variable on an outcome variable by excluding the possible effect of other covariates. In QED, each uniformly sampled individual u is compared with an individual v randomly selected from those that have identical covariates with u but the cause variable. And thus, any outcome difference between these two individuals can be attributed to the cause variable we are tracking. In our context, the cause variable we want to assess is the device type, and the covariate factors include time of day, connection type, as well as location. We first group views based on the combination of device type, geolocation, channel ID and connection type and we do a QED analysis of the problematic session ratio on each resulting group with a 10-minute granularity.

Let T be the set of problematic session ratios obtained with device type k, location g, channel c, connection type w and starting hour h. Each element $u \in T$ is matched with a problematic session ratio v uniformly and randomly selected from the ratios computed at the same location, same channel, same connection type and close starting hour $h \pm 2$, but with *another device type* k'. Each matched pair is labeled with "1" if $u > v$, "-1" if $u < v$, and "0" if $u = v$. Finally, we average all the outcome differences over the matched pairs. A positive average outcome indicates that views from device type k are more likely to be aborted than views from device type k'.

In order to see if the cause variable significantly impacts the outcome, we make a statistical test with null hypothesis H_0: the device type has no impact on the problematic session ratio. Under such a hypothesis, the number of matched pair labeled as "1" (denoted as X) should follow an n-trial binomial distribution with success probability $p = 1/2$, where n is the number of matched pairs. When n is large enough the distribution of X converges to a normal distribution with a mean $n/2$ and variance $n/4$. If H_0 holds, the probability (*i.e.* p-value) of x positive values is at most $P(|X - n/2| \geq |x - n/2|)$. If the p-value is very small (*i.e.* $p < 0.001$), we can safely reject the null hypothesis and consider the impact of the cause variable statistically significant. We show the results of applying QED and this test on the different couple of device categories in Table 3.

The results show that regardless of connection types, location and time of the day, the likelihood that an aPhone (resp. aPad) user experiences a higher problematic session ratio than an iPhone (resp. iPad) user is greater than the likeli-

Table 3: QED results on the impact of device type

k_0 vs. k_1	avg. outcome	p-value
aPhone vs. iPhone	0.35	10^{-16}
aPhone vs. aPad	0.04	0.74
iPhone vs. iPad	0.19	10^{-7}
aPad vs. iPad	0.31	10^{-16}

hood that opposite holds by a margin of 35% (resp. 31%). Besides, while iPhone views are more likely to be problematic than iPad ones, we do not observe similar results in comparing Android phones and pads as the null hypothesis cannot be rejected. One possibility is that Android devices have more diverse capacity (*e.g.* CPU/memory, screen size) than iOS devices. Another potential factor is the difference of video download approaches and buffer management strategies used by Android and iOS devices [12]. For example, iOS devices send more HTTP requests to download video content than Android devices [12]. We leave the exact reason behind the observation as future work as it is out the scope of this paper.

4. UNDERSTANDING USER PATTERNS

In live streaming systems, users are fairly limited in interaction with the streaming object. It thus becomes important to understand how users react in such a context.

4.1 User Activity

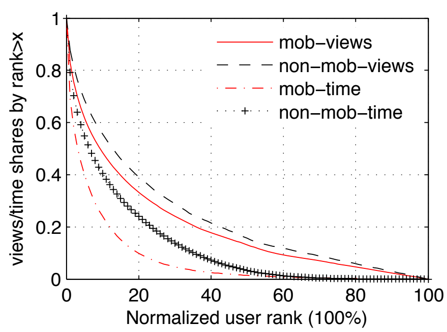


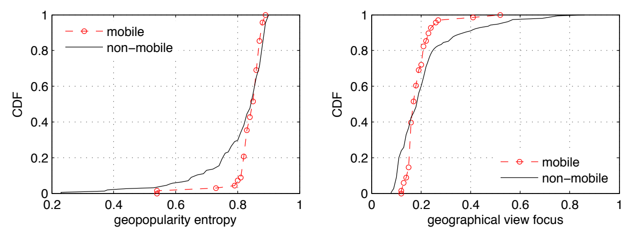
Figure 5: User activity rank distribution

User activity can be measured by either the number of views that successfully start or the total viewing time during the observation period. Figure 5 examines the user activity distribution by ranking users according to the number of views and the total viewing time. In total, we observe 717,578 unique mobile users and 1,166,011 non-mobile users. We normalize user ranks and compute the normalized aggregated views/time of the least x-th active users. For mobile live streaming, the ranking distribution based on viewing time is much skewed than that based on the number of views. The reason should be that users with larger screen terminals and better connection performance tend to have longer viewing duration than other, although they might make nearly the same number of views.

Compared with non-mobile streaming, mobile streaming shows a more skewed user ranking distribution. Nevertheless, for both mobile and non-mobile live streaming, the ranking distributions based on viewing time follow the Pareto principle: top 20% of users make more than 80% of aggregated viewing time. The biased user activity distribution implies that the top active users are highly stick to the system and they are possible candidates for value-added services (*e.g.* HD streaming).

4.2 User Geographical Distribution

One of the challenges in designing a live streaming content delivery system is where the content of a particular channel should be broadcasted: globally or locally in a few regions. We answer this question by analyzing the user geographical distribution for individual channels, *i.e.* how the viewers of individual channels are distributed across provincial locations. We use the *viewer geographical entropy* of a channel to measure the uniformity of its viewers over locations and the *viewer geographical focus* to measure the intensity of its viewers in the top popular location. More specifically, the entropy of a channel k is defined as $e_k = (\sum_{i=1}^{n} p_{ki} \log p_{ki})/\log n$, where $p_{k,i}$ is the fraction of channel k's viewers in location i and $n = 34$ is the number of locations. A smaller (resp. larger) entropy indicates a more biased (resp. uniform) geographical distribution. The viewer focus of a channel k is the fraction of users in the most popular location t of channel k and defined as $v_k = p_{k,t}$.



(a) Entropy (b) Focus

Figure 6: Channel geographical popularity distribution

Figure 6 plots the viewer entropy and focus distributions. We observe that all mobile channels except one have a view entropy higher than 0.8, and for 80% of mobile channels, the most popular location contributes only 20% of viewers. In other words, viewers of individual mobile channels tend to be distributed evenly across all locations. Looking at the non-mobile channels, we observe that several channels exhibit an effect of geographical concentration in a few locations as they have a low entropy and high focus values. A further analysis reveals that these channels are in Cantonese language and are not available in mobile live streaming. As such, they attract viewers only from a limited number of Cantonese-spoken geolocations.

The uniform user geographical distribution for individual channels implies that one should be careful when selecting peers in the P2P delivery, since a random peer selection strategy would select peers located in different regions (or AS), which could greatly undermine the performance.

4.3 Arrival and Departure Rates

Live event streaming is characterized by the fact that the schedule of views is not decided by the viewer but rather by the event itself. This means that the arrival process of viewers for such event is likely to be biased at least by a synchronization at the beginning of the streamed event. Nevertheless all people do not join a stream program precisely at its starting time, some users might join during the event (or even before). Similarly, one can expect to have a concentration of people leaving a streaming session at the end

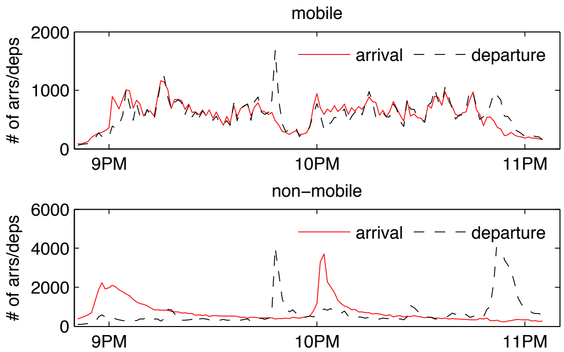


Figure 7: Arrival and departure rate during an event

of the event, while some people might leave it earlier. This means that the arrival process of a streaming video is expected to vary over time and an assumption of a constant distribution will not be reasonable as the arrival process is non-stationary.

In order to validate these points, we show in Figure 7 the number of arrivals (*i.e.* non-abandoned views) and departures per minute for both mobile and client streaming of a single soccer game event which started at 9PM. We observe notable difference between the mobile and non-mobile curves. The non-mobile case shows a gradual increase around the time of the beginning of the game and a decrease of interest that is shown by a decrease in the arrival rate. At the beginning of the second half time (about 5 minutes before the 10PM), there is another increase followed by a decrease. However the mobile arrival curve is more spiky. One can understand the reason by observing that spikes that happen during the game are concomitant with large number of departures. This suggests that the spikes are caused by a significant number of mobile users being disconnected from the streaming and immediately trying to reconnect.

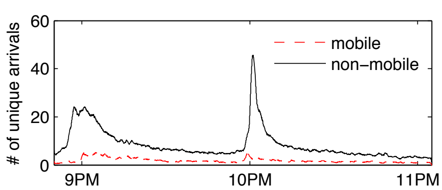


Figure 8: Unique arrivals at 1 second granularity

In order to have a better view of the real user arrival process, we considered only the first time a user is connecting to the streaming as an arrival and we show in Figure 8 this arrival at 1 second granularity for both mobile and non-mobile cases. It can be seen that the arrival follows a modulated exponential process, which models the arrival rate $\lambda(t)$ over time as :

$$\lambda(t) = \sum_{e \in E} \alpha_e \beta_e^{t-T_e} u(t - T_e) + n(t)$$

where E is a set of events, for example the beginning of the game or the half-time, happening at time T_e, α_e represents the relative importance of the model and β_e is the forgetting rate of the event e, the function $u(t-T_e)$ represents the step function stepping up at time $t = T$ and $n(t)$ is a gaussian noise. This model is appealing because of its relative simplicity and also because of the fact that it assumes that the effect of event e happening at time T_e decreases exponentially with time. The same type of model can be used in order to describe the arrival of users after a disconnection (see Figure 7) by simply adding the disconnection event to the set of events. This type of model is widely used in the context of system identification in signal processing and as such there is complete methodologies for estimating the different parameters [18]. However, because of lack of space we do not go further in analyzing this model and we push this back to future work. Notably, we observe very similar graphs for other live-streaming events that we analyzed.

5. RELATED WORK

The practical live streaming systems have been examined in [16][15][5][17]. Veloso *et al.* [16] characterized a live streaming media at 3 granular levels: clients, sessions and transfers. They found that live streaming workload is heavily driven by the nature of the content. Sripanidkulcha *et al.* [15] found the heavy-tailed distribution of viewing duration by analyzing a live streaming workload. In [5], the authors analyzed a popular P2P IPTV by examining high-level statistics of user behavior (*e.g.* daily access pattern) and the system behavior (*i.e.* peer selection). Vieira *et al.* [17] presented a set of crawled logs for the SopCast P2P streaming system. Our work differentiates from these studies in that we focus on user behavior in a mobile live streaming system.

Mobile users become interested in viewing videos with smart terminals. Li *et al.* [9] measured the user behavior and video popularity in a mobile VoD system. The authors in [13] examined how the cellular network dynamics impact user download size. The mobile video popularity and its impact on P2P delivery were studied in [11]. Balachandran *et al.* [3] developed a QoE prediction model for Internet video system and took type of video and connection type as potential factors. Liu *et al.* compared the video download performance of Android and iOS devices [12]. Finamore *et al.* [4] examined the differences of traffic patterns when using PCs with landline connections or mobile terminals with WiFi connections to watch YouTube videos. These works are largely complementary to ours.

6. CONCLUSION

This paper has characterized the user behavior in the PPTV mobile live streaming system from various perspectives. We particularly examined the structure of abandoned session problem and found that the problematic session ratio varies over time of day, channels as well as geolocations. Notably, mobile device type could have a huge effect on the probability of being a problematic session. We have also compared user patterns between mobile and non-mobile live streaming and identified the mobile streaming exhibits much higher user churns, more skewed user activity distribution and less biased user interest on channels.

Acknowledgment

This work was supported by the National Basic Research Program of China No. 2012CB315801, the NSFC No. 61133015 and 61272473, the National High-tech R&D Program of China No. 2013AA013501, and CAS YZ201229. Zhenyu Li was also partly supported by NICTA Australia and CNNIC DNSLab.

7. REFERENCES

[1] Cisco visual networking index: Global mobile data traffic forecast update, 2013-2018. Technical report, Cisco, 2014.

[2] Pptv. `http://www.pptv.com`, 2015.

[3] 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.

[4] A. Finamore, M. Mellia, M. M. Munafò, R. Torres, and S. G. Rao. Youtube everywhere: Impact of device and infrastructure synergies on user experience. In *Proceedings of the ACM IMC*, 2011.

[5] X. Hei, C. Liang, J. Liang, Y. Liu, and K. Ross. A measurement study of a large-scale p2p iptv system. *Multimedia, IEEE Transactions on*, 9, 2007.

[6] J. Jiang, V. Sekar, and H. Zhang. Improving fairness, efficiency, and stability in http-based adaptive video streaming with festive. In *Proceedings of the ACM CoNEXT*, 2012.

[7] S. S. Krishnan and R. K. Sitaraman. Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs. In *Proceedings of the ACM IMC*, 2012.

[8] Y. Li, Y. Zhang, and R. Yuan. Measurement and analysis of a large scale commercial mobile internet tv system. In *Proceedings of the ACM IMC*, 2011.

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

[10] Z. Li, G. Xie, J. Lin, Y. Jin, M.-A. Kaafar, and K. Salamatian. On the geographic patterns of a large-scale mobile video-on-demand system. In *Proceedings of the IEEE INFOCOM*, 2014.

[11] J. Lin, Z. Li, G. Xie, Y. Sun, K. Salamatian, and W. Wang. Mobile video popularity distributions and the potential of peer-assisted video delivery. *Communications Magazine, IEEE*, November 2013.

[12] Y. Liu, F. Li, L. Guo, B. Shen, and S. Chen. A comparative study of android and ios for accessing internet streaming services. In *Proceedings of PAM*, 2013.

[13] Z. Shafiq, M., J. Erman, L. Ji, A. Liu, X., J. Pang, and J. Wang. Understanding the impact of network dynamics on mobile video user engagement. In *Proceedings of the ACM Sigmetrics*, 2014.

[14] T. Silverston and O. Fourmaux. Measuring p2p iptv systems. In *Proceedings of the NOSSDAV*, 2007.

[15] K. Sripanidkulchai, B. Maggs, and H. Zhang. An analysis of live streaming workloads on the internet. In *Proceedings of the ACM IMC*, 2004.

[16] E. Veloso, V. Almeida, W. Meira, A. Bestavros, and S. Jin. A hierarchical characterization of a live streaming media workload. In *Proceedings of the ACM IMC*, 2002.

[17] A. B. Vieira, A. P. C. da Silva, F. Henrique, G. Goncalves, and P. de Carvalho Gomes. Sopcast p2p live streaming: Live session traces and analysis. In *Proceedings of the ACM Multimedia Systems Conference*, 2013.

[18] E. Walter and L. Pronzato. *Identification of parametric models from experimental data.* Springer-Verlag, 1997.