Understanding Internet Video Viewing Behavior In the Wild

Athula Balachandran
Carnegie Mellon University
abalacha@cs.cmu.edu

Vyas Sekar
Stony Brook University
vyas@cs.stonybrook.edu

Aditya Akella
University of Wisconisin Madison
akella@cs.wisconsin.edu

Srinivasan Seshan
Carnegie Mellon University
srini@cs.cmu.edu

ABSTRACT

Over the past few years video viewership over the Internet has risen dramatically and market predictions suggest that video will account for more than 50% of the traffic over the Internet in the next few years. Unfortunately, there has been signs that the Content Delivery Network (CDN) infrastructure is being stressed with the increasing video viewership load. Our goal in this paper is to provide a first step towards a principled understanding of how the content delivery infrastructure must be designed and provisioned to handle the increasing workload by analyzing video viewing behaviors and patterns in the wild. We analyze various viewing behaviors using a dataset consisting of over 30 million video sessions spanning two months of viewership from two large Internet video providers. In these preliminary results, we observe viewing patterns that have significant impact on the design of the video delivery infrastructure.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Performance Attributes

General Terms

Experimentation, Measurement, Performance

Keywords

Internet video, User behavior

1. INTRODUCTION

Video viewership over the Internet has increased dramatically over the past few years. Recent reports suggest that video constitutes 51% of the traffic on the Internet in 2011 and it is estimated to be as much as 56% in the next few years [2]. User expectations for quality of video streamed over the Internet has also increased placing more onus on various content providers (e.g., Hulu, Netflix) to efficiently distribute the content to its customers.

Content providers primarily rely on Content Distribution Networks(CDNs) that have presence across different geographical locations to increase their global availability with low costs and low latency. Unfortunately, there has been signs that the CDN infrastructure is being stressed with the increasing amount of video workload [8]. This raises the important question of how the delivery infrastructure should be designed, provisioned and managed to meet the growing demands for high quality video delivery.

SIGMETRICS'13, June 17-21, 2013, Pittsburgh, PA, USA.
ACM 978-1-4503-1900-3/13/06.

The goal of this study is to take a first step towards this by trying to uncover and understand video access patterns in the wild that have implications for designing a video delivery system. In this work, we investigate various video access patterns using a large dataset consisting of client-side measurements provided by `conviva.com`. A longer-term goal is to characterize these viewing patterns and use them to analyze different infrastructure designs that have received significant industry traction recently [9, 5, 1, 10].

Our preliminary evaluation indicates several interesting video access patterns like time-of-day effects, regional interest in specific content and partial interest. In our ongoing work, we plan to understand these further by characterizing them and explore how these observations can be effectively used to tailor new delivery infrastructure designs.

2. DATA

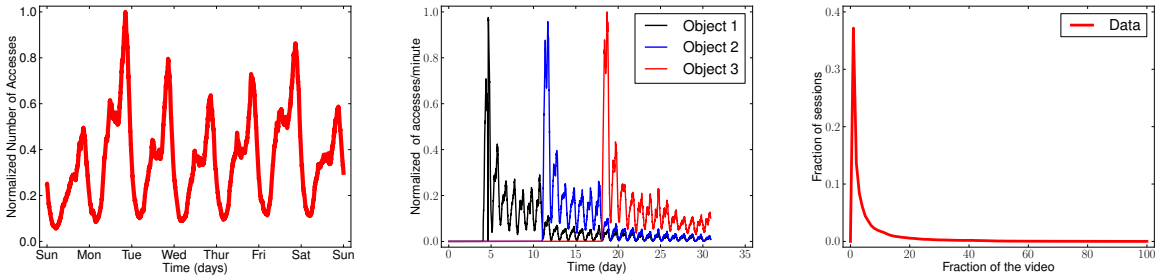
For our study, we used data from two popular United States-based video content providers that appear consistently in the Top 500 sites in terms of overall popularity. The data was collected by `conviva.com` using their client-resident instrumentation library that runs in video players. The library listens to various events from the player as well as polls for additional statistics, and reports these to a data aggregation service that runs in the data center. The data was collected over two months and consists of information about more than 30 million video viewing sessions. The sessions can be classified based on the type of video that was viewed as:
(1) **Video on Demand (VOD):** VOD objects are between 35 minutes and 60 minutes long and consist of television series episodes, news shows and reality shows.
(2) **Live:** These objects are primarily sporting events that are broadcast while the event is happening.

We collect information about each session including: (1) clientID (a unique identifier assigned to the player and stored in the Flash cookie to be used by subsequent views), (2) geographical location (country, state and city of the user), (3) the Internet Service Provider of the user, (4) start time and duration of the session, (5) performance metrics (such as average bitrate of the video, estimated bandwidth at the client), and (6) other information such as name of the video and its actual duration.

3. RESULTS

Our preliminary analysis revealed several interesting video access patterns. Unsurprisingly, there is significant difference between VOD and live content in terms of user viewing behavior. Hence, we report the viewing patterns for VOD and live separately.

Some of the interesting observations in the case of VOD included:



(a) Diurnal pattern for VOD (b) Longitudinal decay for VOD (c) Partial User Interest for Live

- **Time-of-day effects:** We observed strong diurnal patterns in the demand for VOD objects as shown in Figure 1a. Zooming into the data and observing accesses from different regions (e.g., US East Coast, US West Coast, US Mid-West) separately, we observed that the load typically peaks around 8pm local time for each region with a lull in the night (not shown due to lack of space). The strong predictable diurnal characteristcs points to high synchrony in viewing behavior. Similarly, the temporal shift in peak loads can be leveraged to handle peak loads in one region by using spare capacity from a different region.
- **Temporal decay in demand:** Figure 1b shows the temporal variation in demand for three sample video objects. We see that the peak demand for VOD objects typically occured on the day of release and then decayed in the subsequent days. We observed this pattern for other video objects as well. The strong diurnal patterns that are observed from this time series plot again points to high synchrony in viewing behavior on a per-object basis. Also, the temporal decay aspect has important implications to provisioning the infrastructure.

Live objects consisted of sporting events, some of which had regional sports teams. A few interesting observations in the case of live included:

- **Regional interest:** Typically, for VOD objects, we observed significant population-induced difference in load across different regions (eg., US East Coast, US West Coast, Mid-West). But for several live events that have regional biases like a local team playing a match, we obsered significantly skewed access rates from regions that exhibit low loads in the typical case. This is an important factor that has to be considered while provisioning the delivery infrastructure to handle unexpected high loads.
- **Partial Interest:** We also observed partial interest in the video objects. Users watched only part of the content during a session as shown in Figure 1c. We also observed that several users joined and quit multiple times during an event. Considering that our dataset consists of sporting events, we assume that this might be caused by users checking current score of the matches frequently.

4. RELATED WORK

There is a large body of related work that examines various aspects of video viewing behavior on the Internet. Our work is also similar in spirit. However considering the fact that Internet video has really taken off in the past few years, we revisit several of these behaviors. For example, there are several studies that have looked at content popularity in the context of content prefetching and caching [4, 7]. We build on these works by focusing on temporal evolution of demand for VOD objects. Similarly there has been past work on user behavior including partial views [3, 6]. Our analysis reconfirms several of these observations. In our ongoing work we are looking at characterizing these behaviors more systematically and placing their implications in the context of delivery infrastructure design.

5. CONCLUSION AND FUTURE WORK

Internet video has really taken off in the past few years. Given this growth, we revisit the question of user access patterns in the wild. Our preliminary analysis reveals several patterns that have implications to designing and provisioning the infrastructure. In future work, we intend to analyze the data in greater detail to uncover other interesting video access patterns and characterize them. We also want to put these observations in the context of new delivery infrastucture designs (e.g., Federated CDNs and hybrid P2P-CDN architectures) and analyze them. The eventual goal is to understand how we can tailor and optimize these delivery infrastructure designs for the video viewing workloads.

Acknowledgements

The authors thank Conviva Inc. for making the video viewing data available for this study.

6. REFERENCES

[1] Akamai NetSession. `http://www.akamai.com/client/`.
[2] Cisco forecast. `http://blogs.cisco.com/sp/comments/cisco\_visual\_networking\_index\_forecast\_annual\_update/`.
[3] L. Plissonneau and E. Biersack. A Longitudinal View of HTTP Video Streaming Performance. In *Proc. MMSys*, 2012.
[4] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon. I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System. In *Proc. IMC*, 2007.
[5] D. Rayburn. Telcos and carriers forming new federated cdn group called ocx (operator carrier exchange). June 2011. StreamingMediaBlog.com.
[6] A. Finamore, M. Mellia, M. Munafo, R. Torres, and S. G. Rao. Youtube everywhere: Impact of device and infrastructure synergies on user experience. In *Proc. IMC*, 2011.
[7] H. Yu, D. Zheng, B. Y. Zhao, and W. Zheng. Understanding User Behavior in Large-Scale Video-on-Demand Systems. In *Proc. Eurosys*, 2006.
[8] X. Liu, F. Dobrian, H. Milner, J. Jiang, V. Sekar, I. Stoica, and H. Zhang. A Case for a Coordinated Internet-Scale Video Control Plane. In *Proc. SIGCOMM*, 2012.
[9] R. Powell. The federated cdn cometh. May 2011. TelecomRamblings.com.
[10] H. Yin, X. Liu, T. Zhan, V. Sekar, F. Qiu, C. Lin, H. Zhang, and B. Li. Design and Deployment of a Hybrid CDN-P2P System for Live Video Streaming: Experiences with LiveSky. In *Proc. ACM Multimedia*, 2008.