Measurement-based Analysis of the Video Characteristics of Twitch.tv

Mark Claypool, Daniel Farrington, and Nicholas Muesch
Worcester Polytechnic Institute
Worcester, MA, 01609, USA
Email: claypool@cs.wpi.edu

***Abstract*—Technology advances have brought a growth in live streaming, where users can broadcast real-time videos of the games they are playing to interested viewers. While characteristics of traditional video on demand sites are relatively well understood, live streaming has received comparatively less attention. A better understanding of such systems is needed to plan for future systems and architectures. This paper uses a crawler and stream analyzer to characterize Twitch.tv, the most popular live streaming system for games. Analysis of the data confirms live stream volume from previous research, while providing novel frame rate and bitrate analysis.**

I. Introduction

The spread of inexpensive, yet powerful end-user devices has enabled the growth in user generated multimedia content, most notably through online photos and video on demand. Most recently, video sharing services, such as provided by YouTube [1] and Twitch.tv,[1] have enabled users to stream videos live to interested viewers. Twitch.tv caters to the niche of live streaming video games as they are being played.

In order to adequately plan for online infrastructures and network management to support live streaming, it is important to understand the characteristics of these emerging systems. For traditional video on demand systems (e.g., YouTube), there have been thorough studies characterizing the popularity of online videos [2] as well as the network characteristics of videos [3], but there have been relatively fewer studies for live streaming systems [4]–[7].

Our work complements these previous studies and fills gaps in the knowledge of live streaming content in several ways:

1) Our data is more recent, as of January 2015. Given the evolving nature of technologies and use for video streaming, frequent analysis is needed in order to represent and understand current behavior and ascertain trends.
2) Our analysis describes live streaming characteristics that have not been analyzed, including frame rates and frame resolutions. Such analysis moves beyond the number of broadcasters, viewers and popularity trends that have been previously analyzed and, to the best of our knowledge, is the first such analysis for live streaming.
3) Our results are compared to previous work, where applicable. While often overlooked by the computer science community, comparison with past results for reproducibility, one of the main principles of the scientific method, as well as to highlight differences is crucial for generalizing knowledge beyond the experience of the individual scientist.

In order to better understand the characteristics of live streaming systems, we analyze Twitch.tv, currently the most popular live streaming service. Every month on Twitch.tv about 1.5 million game players broadcast 11 million videos to 100 million viewers [8]. Twitch.tv already ranks fourth in the U.S. for peak Internet traffic, accounting for 1.8% of all traffic during peak hours [9]. We deploy a crawler to harvest and analyze Twitch.tv live streams for about one month in January 2015, and analyze the video characteristics of the most popular streams.

Analysis of the results shows the number of Twitch.tv streams has a pronounced correlation with the time of day and the day of the week based on U.S. time zones, confirming results from previous studies. The length of Twitch.tv videos on demand appear to be heavy-tailed, which may contribute to the self-similarity of Internet traffic, a result shown for other file types but novel analysis for live streaming systems. Additional novel analysis shows the dominant live stream resolutions are 720p and 1080p and about 40% of live stream videos have frame rates of 60 f/s, much higher than traditional video "full-motion" frame rates around 30 f/s.

The rest of the paper is organized as follows: Section II introduces work most related to this paper; Section III describes our methodology to gather and analyze Twitch.tv data; Section IV analyzes the results; and Section V summarizes our conclusions and presents possible future work.

II. Related Work

Our Twitch.tv data set is from January 2015. Our aggregate analysis includes the number of live streams versus time of day and day of week. Our individual streams analysis includes resolutions, frame rates and bitrates, as well as the lengths of videos on demand.

Mehdi et al. [4] analyze data from September 2011 to January 2012. As in our work, they analyze the number of live streams with time of day and day of week correlations. However, they do not analyze individual live stream characteristics. They have analysis of live stream durations, but not of video on demand lengths.

[1] http://www.twitch.tv/

Pires and Simon [5], [6] analyze Twitch.tv and YouTube data from January to April 2014. As in our work, they analyze the number of live streams with time of day and day of week correlations. The only video characteristic they analyze is bitrate.

Zhang and Liu [7] analyze Twitch.tv from October to November 2014. They do not analyze any of the same Twitch.tv data as we do, but do show the duration of live stream media is about 100 minutes.

III. Methodology

We built a Web crawler to automatically gather information from Twitch.tv. Our crawler is based on Scrapy,[2] an open source Web crawler built in Python, which we modified to obtain information from the Twitch.tv API. Twitch.tv uses a RESTful API, providing information about the resources hosted. The API consists of calls that are static URLs with slightly different names depending upon the information the client wants to receive, e.g., the channel name.

The crawler gathered the number of active live streams on Twitch.tv and the number of videos on demand, as well as the length for each video on demand.

In order to gather technical information about an active stream, we used a stream analyzer[3] that takes in the channel name of an active stream as input and provides video resolution, frame rate and bitrate as output.

The crawler ran every hour from January 1st to January 23rd, 2015. During this time, the analyzer was used on the ten most popular streams (based on number of viewers) twice a day, at 4 p.m. and 11 p.m.

Additional data gathered on the live streams included audio and video encoding characteristics, broadcast delay, creation, up and update times, and number of viewers, and on the VoDs included views and game name. Analysis of the additional data can be found in our full report [10].

IV. Results

Network traffic often shows time of day dependencies, with "awake" hours having more network traffic for user-based applications as well as day of week dependencies, and with weekend hours having more network traffic for user-based leisure applications. Figure 1 depicts the number of Twitch.tv live streams over 24-hours for a representative weekday and weekend. The x-axis is the time of day in hours in the U.S. Eastern Standard time zone, and the y-axis is the number of streams. The blue "star" trendline is the weekday and the red "plus" trendline is the weekend. There is a noticeable change in number of streams over the course of the day, with the minimum around 6 a.m. and a maximum of 2x to 3x the minimum around 4 p.m. The weekend has more streams than does the weekday at all hours, from about 10% more at midnight to about 50% more at 6 a.m.

At the end of 2011, Twitch.tv had about 800 live streams, with similar time of day and day of week correlations [4]. By early 2014, Twitch.tv had grown to about 6500 live streams, with similar time of day and day of week correlations [5], [6]. The number of streams we observe about one year later is approximately double.

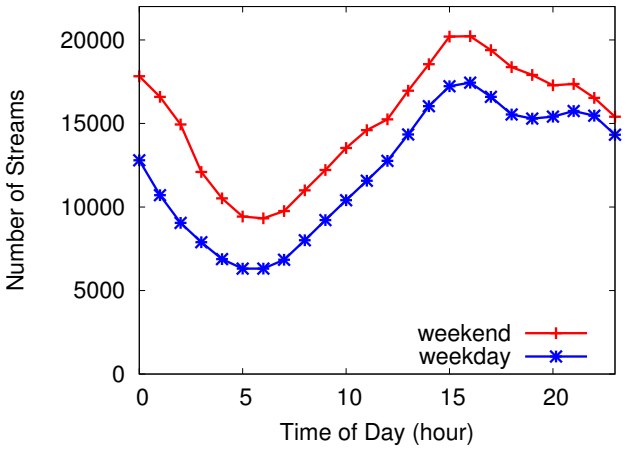

Fig. 1. Streams. Number of streams over time of day (in hours). All times are U.S. Eastern Standard Time.

While much of the emphasis is on live streaming, Twitch.tv also includes videos on demand (VoDs) that let users stream former live events. Figure 2 depicts the cumulative distribution of the VoD lengths in seconds, shown in its complement (a CCDF) with logscale axes to allow for clearer examination of the tail. The median VoD length is 497 seconds (about 8 minutes), slightly higher than, but not dissimilar to, video lengths on YouTube [3]. The longest VoD is almost 50 days long. As of 2011, Twitch.tv live streams tended to be longer than VoDs (median 45 minutes), with a much shorter maximum length (longest 20 hours) [4]. As of 2014, Twitch.tv live streams were longer (median of 100 minutes) [7].

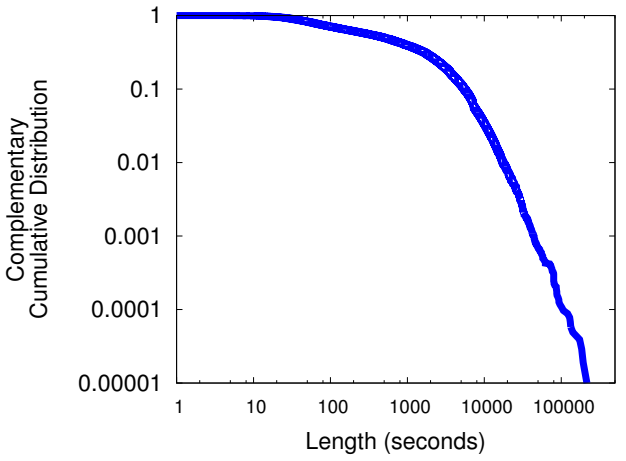

Fig. 2. Lengths. Complementary cumulative distribution (CCDF) of video lengths in seconds.

Previous studies [11] have suggested long-tailed distribution

[2]http://scrapy.org/
[3]https://r-1.ch/analyzer/

of transfer times may contribute to the self-similarity of Internet traffic. If the distribution of Twitch.tv video durations is long-tailed, then streaming media may contribute to Internet traffic self-similarity, especially as the popularity of live streaming media grows. The definitive test for a long-tailed distribution is that the steepness of the slope in the graph does not increase in the extreme tail. By visual inspection, the duration distribution in Figure 2 appears to be long-tailed.

Unlike in VoD systems where a server can take the time to transcode an uploaded video into multiple possible streaming resolutions, Twitch.tv live streams are streamed at the resolution captured by the broadcaster. Both the video experienced by the user as well as the network traffic is impacted by the resolution choice. Figure 3 depicts a histogram of the live stream resolutions. The x-axis is the screen resolution in pixels, sorted from smallest to largest, and the y-axis is the percentage of the live streams at the given resolution. By far, the two most dominant live stream resolutions are 1280×720 pixels and 1920×1080 pixels, corresponding to the HD resolutions for 720p and 1080p, respectively, with a small number at other resolutions. The mean bitrates for live streams at 1920×1080 is 2928 kb/s, about 2.5 times the mean live stream bitrate at 1280×720 of 1216 kb/s.

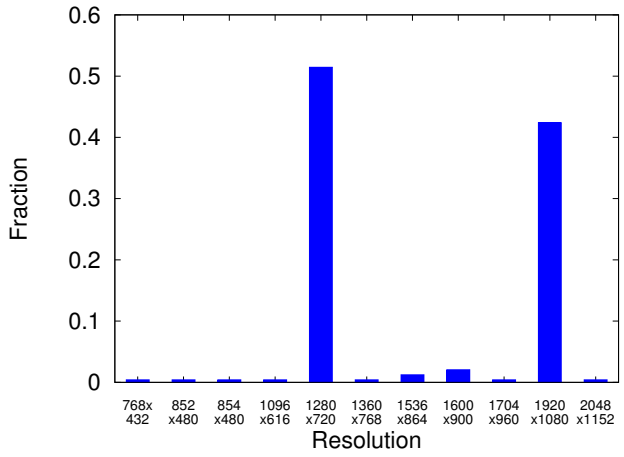

Fig. 3. Screen Resolutions. Histogram of live stream screen resolutions in pixels, ordered from lowest resolution to highest resolution.

As for resolution, the video experienced by the user and the network traffic is impacted by the encoded frame rate. Figure 4 depicts a cumulative distribution of the live stream frame rates. The x-axis is the frame rate in frames per second (f/s) and the y-axis is the cumulative distribution. About half of encoded frame rates are at the "full-motion" rates of 25 (5%) f/s or 30 (42%) f/s. However, unlike traditional video encoding, about 40% of live streams have an encoding rate of 60 f/s, likely because many PCs run games at frame rates of 60 f/s or higher.

Networks supporting live streaming must be able to provide the needed video bitrates since given the live, ephemeral nature of the broadcast, there is no opportunity to download the video

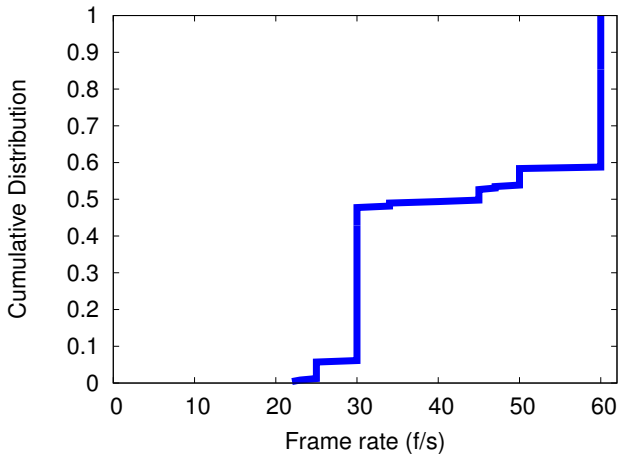

Fig. 4. Frame rates. Cumulative distribution of frame rates, in frames per second.

and watch it later. Figure 5 depicts a cumulative distribution of the live stream bitrates. The x-axis is the bitrate in kilobits per second (kb/s) and the y-axis is the cumulative distribution. The median bitrate is about 3000 kb/s with a fairly even distribution of bitrates from about 500 kb/s to a maximum of about 5000 kb/s. In early 2014, the median bitrate for Twitch.tv was 2 Mb/s for 1080p, and only slightly less for 720p [5], [6].

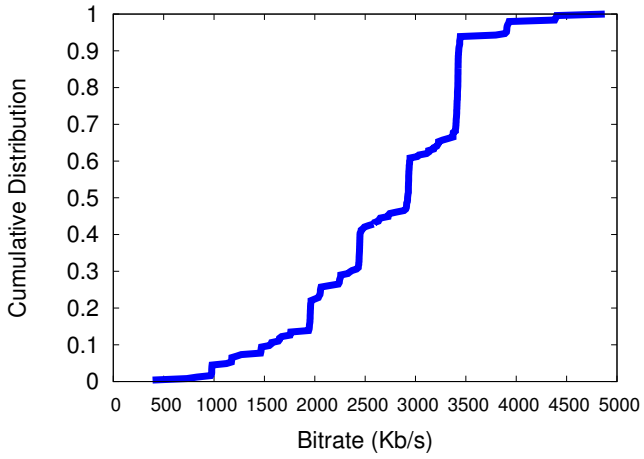

Fig. 5. Bitrates. Cumulative distribution of streaming bitrates, in kilobits per second.

V. Conclusion

A better understanding of emerging live streaming characteristics can help in capacity planning, management of streaming networks, the design of new live streaming systems, and in future network research. This work seeks to complement previous live streaming characterization by confirming select previous results and filling the gap in live streaming knowledge with video characteristics. To do this, we crawl Twitch.tv, the most popular system for live streaming game play, and analyze

the characteristics of the most popular live streams for a one-month period in January 2015.

Analysis of the results shows the number of streams correlates with time of day and day of week, with 6 a.m. on the weekday having the fewest streams (about 6300) and 4 p.m. on the weekend having the most (about 20,000). These correlations confirm previous research, but the recency of our data shows growth in number of streams. Twitch.tv video on demand lengths are heavy-tailed, which may contribute to the self-similarity of Internet traffic, a result shown previously for Web traffic but a novel finding for live streaming. Typical high definition 720p and 1080p resolutions make up almost 95% of all Twitch.tv live streams. While traditional "full-motion" rates around 30 f/s make up about 40% of all live streams, game frame rates of 60 f/s also make up 40% of live streams, this latter rate uncommon compared to traditional video.

Future work may include additional longitudinal analysis of Twitch.tv as the system continues to evolve. Work may also include in-depth analysis of stream behavior in response to congestion or other system bottlenecks. Other live streaming systems, such as from YouTube [1], can also be studied.

REFERENCES

[1] "YouTube Live Streaming Guide," Online at https://support.google.com/youtube/answer/2474026, accessed 06/10/15.

[2] 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 *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement (IMC)*, San Diego, CA, USA, 2007, pp. 1–14.

[3] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "YouTube Traffic Characterization: A View From the Edge," in *Proceedings of the ACM Internet Measurement Conference (IMC)*, San Diego, CA, USA, Oct. 2007.

[4] M. Kaytoue, A. Silva, L. Cerf, W. Meira, Jr., and C. Raïssi, "Watch Me Playing, I Am a Professional: A First Study on Video Game Live Streaming," in *Proceedings of the 21st International Conference Companion on World Wide Web*, Lyon, France, 2012, pp. 1181–1188.

[5] K. Pires and G. Simon, "DASH in Twitch: Adaptive Bitrate Streaming in Live Game Streaming Platforms," in *Proceedings of the Workshop on Design, Quality and Deployment of Adaptive Video Streaming (Video Next)*, Sydney, Australia, 2014, pp. 13–18.

[6] ——, "YouTube Live and Twitch: a Tour of User-generated Live Streaming Systems," in *Proceedings of the 6th ACM Multimedia Systems Conference*, Portland, OG, USA, 2015, pp. 225–230, (dataset paper).

[7] C. Zhang and J. Liu, "On Crowdsourced Interactive Live Streaming: A Twitch.Tv-based Measurement Study," in *Proceedings of the ACM Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, 2015, pp. 55–60.

[8] "Two Thousand MOREteen," Online at http://www.twitch.tv/year/2014, 2014.

[9] D. Fitzgerald and D. Wakabayashi, "Apple Quietly Builds New Networks," *Wall Street Journal*, Feb. 2014.

[10] D. Farrington and N. Muesch, "Analysis of the Charisteristics and Content of Twitch Live-Streaming," *Interactive Qualifying Project IQP-MLC-TT14*, May 2015, (Advisor Mark Claypool).

[11] M. Crovella and A. Bestavros, "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 835–846, 1997.