

Understanding Throughput Stability and Predictability to Enable Better Video Quality of Experience

Xiaoqi Yin, Junchen Jiang, Vyas Sekar, Bruno Sinopoli
Carnegie Mellon University

Video consumption is undergoing a dramatic shift from traditional media (e.g., broadcast, cable, satellite TV) to the Internet. As Internet video goes mainstream, users' expectations for quality have dramatically increased as users expect higher delivered bitrate, fewer buffering events, and low start-up delays [4, 11, 12, 16]. Poor quality-of-experience (QoE) can be a huge cause for concern for ISPs, content providers, and content delivery networks as evidenced by media reports on the “growing pains” of the Internet video ecosystem (e.g., [2, 5]).

This growth in Internet video is driven by the shift from customized video protocols (e.g., RTMP [6]) to HTTP-based [21] adaptive streaming protocols (e.g., [3, 7, 8, 19]). Given the complex ecosystem and presence of diverse bottlenecks, the *bitrate adaptation logic* in the client video player is critical to ensure a rich user experience [8]. This algorithm dynamically (at the granularity of seconds) adjusts the requested video bitrate based on the network conditions and the occupancy of its buffer. Since the video is already chunked and encoded at different bitrate levels, the goal of an adaptive video player is to choose the bitrate level for future chunks to deliver the highest possible QoE; e.g., maximizing bitrate while minimizing the likelihood of buffering and avoiding too many bitrate switches. As recent efforts have pointed out, this problem is challenging because the aforementioned goals are opposing ones giving rise to a natural tradeoff and because the adaptation algorithm has to be robust to varying network conditions as well as interactions with lower-layer control loops (e.g., [9, 14, 15, 22]).

In our recent research, we presented an attempt to formally characterize the problem by using rigorous mathematical analysis and design. To this end, we formulate the adaptive video streaming over HTTP problem through the “lens” of control theory. This provides a general framework to objectively compare different *classes of algorithms* and analyze their sensitivity to operating conditions rather than evaluate specific artifacts. We developed a control-theoretic model to capture the key dynamical phenomena involved in video adaptive streaming over HTTP with formally stated QoE objectives [23] and demonstrated the feasibility of implementing it in industry reference platforms such as DASH.js [1] through a model-

predictive control (MPC) approach.

A key component of our proposed MPC-based video adaptive player algorithm and any video adaptation player is a reasonably accurate estimate of the available bandwidth between the client and the server for the next several epochs. While there is a rich history of Internet measurement to understand various properties of Internet paths including studies on understanding routing inefficiency, route stability, inter-arrival characteristics of flows, and TCP available bandwidth measurements [20, 24, 25], they do not specifically target the requirements of video workloads. There are three key distinguishing characteristics of video workloads in the context of control-theoretic formulation:

- First, a key metric that impacts the control logic is not the available bandwidth or the capacity, but the *stability* of the bandwidth. In fact, there has been surprisingly little work on understanding bandwidth stability—the three closest related works we are aware of focusing on the stability of Internet path properties focus largely on non-video workloads and are quite dated [10, 13, 24].¹
- Second, the specific target servers are not general end-hosts but CDN servers hosting videos.
- Third, during the initial starting window when the rate estimates are unreliable, we also need some estimates of the *expected* bandwidth between the client and the video server based on long-term forecasts rather than current measurements.

We argue that we need community efforts along two thrusts to better understand and improve video quality of experience:

- First, we need large-scale wide area measurements to analyze the stability and predictability of bandwidth to video servers. The selection of the initial bitrates can be simplified if the available bandwidth of the client is predictable on long-term timescales (e.g., days or weeks). For example, if we can reuse historical data of the client's throughput measurements or

¹These are from late 90s and early 2000s; i.e., largely predating the widespread deployment of high-speed broadband, CDNs, and the growth of Internet video.

share throughput measurements (e.g., clients sharing the same /24 prefix), then the initial bitrate selection logic can be simplified and avoid the “slow start” like effects. While the long-term throughput estimates help inform the initial bitrate selection, the real-time adaptation logic depends more critically on the stability of the bandwidth on shorter timescales (e.g., on the scale of few seconds or minutes). Again, if we know that the short-term variability is low or that the bandwidth tends to be bimodal, then the adaptation logic may be able to better react to the changes. For instance, if we know that the bandwidth is stable on the scale of few minutes but highly variable on the scale of seconds, then we can use larger buffers or use better dampening logic to avoid frequent bitrate switches.

- Second, we need to develop robust estimation functions that can serve as accurate bandwidth predictors. As prior work has shown, the throughput observed by a player for each chunk is not a reliable estimate of the available capacity due to a range of reasons; e.g., weird interactions with TCP [14] or timing effects on when the chunk was actually scheduled [15]. Thus, we need more systematic mechanisms to build a robust bandwidth estimator from per-chunk estimates. Our prior work suggests two guidelines to build a more robust bandwidth estimator [15]. First, instead of using the instantaneous throughput, we use a smoothed value computed over the last several chunks. Second, we want this smoothing to be robust to outliers. For example, using the arithmetic mean is biased by outliers if one chunk sees a very high or low throughput. Thus, we will investigate other estimation functions (e.g., harmonic mean) that are more robust to outliers. Other metrics such as chunk retrieval delay and estimation processes like stochastic prediction have also been suggested in prior work but not evaluated at scale (e.g., [17, 18]).

References

- [1] Dash-Industry-Forum, `dash.js`. <https://github.com/Dash-Industry-Forum/dash.js/wiki>.
- [2] HBO Go Crashes During ‘True Detective’ Finale. <http://www.hollywoodreporter.com/live-feed/hbo-go-crashes-true-detective-687087>.
- [3] Adobe http dynamic streaming. www.adobe.com/products/hds-dynamic-streaming.html.
- [4] Driving Engagement for Online Video. <http://goo.gl/p05Cj>.
- [5] Internet TV Was the Big Loser on Oscar Night. <http://www.theatlantic.com/entertainment/archive/2014/03/internet-tv-was-the-big-loser-on-oscar-night/284173/>.
- [6] Real-time messaging protocol. www.adobe.com/devnet/rtmp.html.
- [7] Smoothstreaming experience. <http://www.iis.net/media/experiencesmoothstreaming>.
- [8] I. Sodagar. The MPEG-DASH Standard for Multimedia Streaming Over the Internet. *IEEE Multimedia*, 2011.
- [9] S. Akhshabi, L. Anantakrishnan, C. Dovrolis, and A. C. Begen. What Happens when HTTP Adaptive Streaming Players Compete for Bandwidth? In *Proc. NOSSDAV*, 2012.
- [10] H. Balakrishnan, M. Stemm, S. Seshan, and R. H. Katz. Analyzing Stability in WideArea Network Performance. In *Proc. ACM SIGMETRICS*, 1997.
- [11] F. Dobrian, A. Awan, I. Stoica, V. Sekar, A. Ganjam, D. Joseph, J. Zhan, and H. Zhang. Understanding the impact of video quality on user engagement. In *Sigcomm*, 2010.
- [12] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. A. Joseph, A. Ganjam, J. Zhan, and H. Zhang. Understanding the impact of video quality on user engagement. In *Proc. SIGCOMM*, 2011.
- [13] Q. He, C. Dovrolis, and M. Ammar. On the predictability of large transfer TCP throughput. In *Proc. ACM SIGCOMM*, 2005.
- [14] T.-Y. Huang, N. Handigol, B. Heller, N. McKeown, and R. Johari. Confused, Timid, and Unstable: Picking a Video Streaming Rate is Hard. In *Proc. IMC*, 2012.
- [15] J. Jiang, V. Sekar, and H. Zhang. Improving Fairness, Efficiency, and Stability in HTTP-based Adaptive Video Streaming with FESTIVE. In *Proc. CoNext*, 2012.
- [16] S. S. Krishnan and R. K. Sitaraman. Video Stream Quality Impacts Viewer Behavior: Inferring Causality using Quasi-Experimental Designs. In *Proc. IMC*, 2012.
- [17] C. Liu, I. Bouazizi, and M. Gabbouj. Rate adaptation for adaptive http streaming. *Proc. ACM MMSys*, 2011.
- [18] K. Miller, E. Quacchio, G. Gennari, and A. Wolisz. Adaptation Algorithm for Adaptive Streaming over HTTP. In *Proc. Packet Video Workshop*, 2012.
- [19] R. Pantos. Http live streaming. 2011.
- [20] V. Paxson. End-to-end routing behavior in the Internet. *IEEE/ACM ToN*, 5(5):601–615, 1997.
- [21] L. Popa, A. Ghodsi, and I. Stoica. HTTP as the narrow waist of the future internet. In *Proc. HotNets*, 2010.
- [22] S. Akhshabi, A. Begen, C. Dovrolis. An Experimental Evaluation of Rate Adaptation Algorithms in Adaptive Streaming over HTTP. In *Proc. MMSys*, 2011.
- [23] X. Yin, A. J. V. Sekar, and B. Sinopoli. A Control-Theoretic Approach for Dynamic Adaptive Video Streaming over HTTP. In *Proc. ACM SIGCOMM*, 2015.
- [24] Y. Zhang and N. Duffield. On the constancy of Internet path properties. In *IMW*, 2001.
- [25] Y. Zhang, V. Paxson, S. Shenker, and L. Breslau. The stationarity of Internet path properties: Routing, loss, and throughput. Technical report, ACIRI Technical report, 2000.