

Complex Interactions in the Content Distribution Ecosystem and Quality-of-Experience: A White Paper for Internet QoE Workshop

Zhi-Li Zhang (zhzhang@cs.umn.edu)
Dept. of Computer Science & Engineering, University of Minnesota

1 Introduction

Since their advent in the early 90s, web applications and services have evolved from the simple client-server model, where content (e.g., a file) is delivered *point-to-point* from a single web server to a single client, to a vast and complex system consisting of numerous servers (some located close to users and others in large back-end data centers), network elements, and software components owned and controlled by a variety of entities such as content providers (CPs), content distribution networks (CDNs), and Internet Service Providers (ISPs). These three entities along with the users that they serve form the content distribution ecosystem. The content distribution ecosystem is a dominant force driving the evolution of the Internet as CDNs carry most of the world's web pages, videos, social networks, software downloads, e-commerce sites, software-as-a-service (SaaS) portals, and other web applications. Each entity in the ecosystem has its own business interest, and attempts to optimize the performance of its components more often *individually*, but sometimes in cooperation with others.

Online video streaming services in particular heighten the importance of understanding the complex interactions present in content distribution, where content providers often employ multiple CDNs to stream videos to a large number of geographically dispersed users. The enormous burden that video traffic places on the Internet infrastructure [8] creates tensions, but also opportunities. Interactions, whether competitive or cooperative, among content providers, CDNs and ISPs are inevitable, and will likely evolve rapidly and become far more complex. For example, they have fostered several important trends, e.g., increasing cooperation among ISPs and CDNs that are variously called Network CDNs, Telco-CDNs, Carrier CDNs, or Managed CDNs, while at the same time content providers compete to attract more users via (specialized) streaming apps or devices (e.g., chromecast) on the end systems.

The *complex interactions* among various entities in today's large-scale content distribution ecosystem makes it a challenging task to ensure high quality-of-experience (QoE) for end users. This is in spite that providing a high quality of QoE for their users is also a key goal for any content or application service provider including online video streaming service providers such as NetFlix, web search services such as Google and Microsoft Bing, social network service providers such as Facebook, and so forth. This is because these service providers rely on either online ads, user subscriptions, or pay-per-view / pay-per-click for their revenues. All three monetization models depend on engaging the user by enticing them to watch more videos, conducting more searches or staying on social networking sites longer, and to come back to their website more often.

2 Measurement Studies of Large-scale Online Services and QoE

We have conducted extensive measurement studies [2, 3, 5, 4, 6, 7] of several large-scale online service providers including YouTube, Netflix, Hulu, Google and Bing to uncover and understand the complex service delivery architectures, various entities involved and the existing mechanisms employed by service providers to improve service performance and QoE for users. To illustrate the complexity, let's examine the seemingly simple act of a user watching a movie from the popular video content provider Netflix. Fig. 1 depicts the high-level architecture of the Netflix video delivery service [5]. Apart from essential operations such as user registration and payment services that are hosted on its own servers, Netflix leverages the Amazon cloud service (e.g., AWS S3) for user sign-in, authorization, digital right management (DRM), video storage and so forth [9, 10]. It utilizes three major CDNs – Akamai, Limelight, and Level3 – for streaming videos to users via the DASH (Dynamic Adaptive Streaming over HTTP) protocol. Each Netflix video is encoded at several different quality levels and is divided into small “chunks” or “fragments” that are typically 1-10 seconds long. Netflix uses a deterministic algorithm to determine the CDN from which to stream video content to a user: the CDN preference order is specified in a *manifest* file sent to the user (rather, the Netflix Silverlight player plugin) in response to her request for a video. During playback, the Netflix player requests one video chunk at a time from the most preferred CDN. With each download, the player measures the received throughput; if network performance degrades, it switches to a lower video encoding rate for the next chunk. The player stays attached to the same CDN even if other CDNs can offer better video quality though one could imagine more fine-grained optimizations where the CDN can itself be switched midstream [5].

Video QoE is typically measured along four axes [11, 1]: *availability* that measures how often the video plays without failures, *start-up time*, which is determined by when the first few video chunks are delivered in order for the video playback to begin, *re-buffering* ratio which refers to the fraction of time the video playback is frozen so as to re-fill the video play-back buffer, and video presentation quality measured by the *average bitrate* at which the video is rendered on the user's screen. Thus, the overall QoE of a video playback session requires the maintenance of a *sustained* video delivery rate to maintain the minimum quality-level a user desires. As such, the user's QoE hinges on a plethora of interacting entities, system components, protocols and network elements, and the decisions they make in the delivery process. These decisions include the CDN selected by the content provider, the (edge) video cache servers selected by the CDN, the geographical and network location of these servers, the underlying ISPs, the available bandwidth between users and video cache servers, the rate adaptation algorithm at the content provider's media player, on-the-fly transcoding middleboxes deployed by ISPs that can step down the bitrates of video chunks in response to link congestion, TCP congestion control algorithms, and other video users that may share the same network path or video servers.

Analogous to video, other content services such as web search [7, 6], social networking, software downloads, and e-commerce each have their *application-specific* QoE metrics that need to be optimized and they too involve complex interactions between multiple entities, software components, and protocols. For instance, the primary QoE metric for web search is *search response time* (SRT), which is defined as the time between sending the query to receiving the complete response page at the browser. To illustrate the impact of complex interactions among various entities on the overall user QoE, in [7] we measured the SRT of Bing for a period of a week. While SRT varies from hour to hour and day to day,

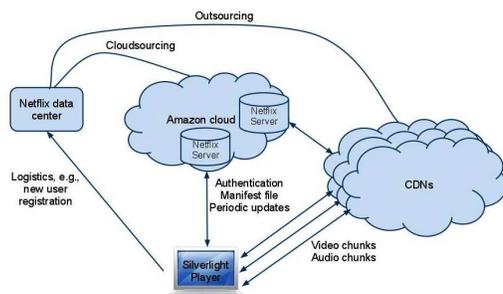


Figure 1: Netflix architecture.

it also exhibits counterintuitive behavior such as *SRT being actually higher during off-peak hours and weekends*. As shown in [7], this apparently *paradoxical* observation can be attributed to the confluence of multiple factors such as slower residential network connections where most search queries come from during off-peak hours and weekends, the types of queries and content delivered in response to these queries which tend to contain far more rich media content (e.g., images and videos), and so forth. As with video QoE, SRT is also known to have an impact on user behavior and studies show that lower SRT's will prompt more search queries[7], leading to more ads being shown, and greater monetization for the search engine provider.

3 QoE Research Collaboration

Our current research demonstrates that understanding and optimizing QoE require a fundamental understanding of the complex interactions that underlie the content distribution ecosystem. However, due to the vast complexity in large-scale online service delivery architecture and the various entities involved, addressing the challenges in improving user QoS and/or ensuring high QoE to users will require multi-faceted collaborative efforts in a number of different domains, ranging from compute and store systems, networks, applications, end devices and user perception and behavior. We are particularly interested in studying the system and network architectural issues in large-scale online services, the complex interactions arising thereof and their impacts on user quality-of-experience. We hope to collaborate with other researchers and contribute to the research in QoE.

References

- [1] *Akamai Streaming: When Performance Matters*, 2004. http://www.akamai.com/dl/whitepapers/Akamai_Streaming_Performance_Whitepaper.pdf.
- [2] Vijay Kumar Adhikari, Yingying Chen, Sourabh Jain, and Zhi-Li Zhang. Where do you "tube"? uncovering youtube server selection strategy. In *Proc. IEEE ICCCN'11*, July 2011.
- [3] Vijay Kumar Adhikari, Yingying Chen, Sourabh Jain, and Zhi-Li Zhang. Vivisecting YouTube: An active measurement study. In *Proc. of IEEE INFOCOM 2012*, March 2012.
- [4] Vijay Kumar Adhikari, Yang Guo, Fang Hao, Volker Hilt, and Zhi-Li Zhang. A tale of three CDNs: An active measurement study of Hulu and its CDNs. In *Proc. of Global Internet Symposium*, March 2012.
- [5] Vijay Kumar Adhikari, Yang Guo, Fang Hao, Matteo Varvello, Volker Hilt, Moritz Steiner, and Zhi-Li Zhang. Unreeling Netflix: Understanding and improving multi-CDN movie delivery. In *Proc. IEEE INFOCOM 2012*, March 2012.
- [6] Yingying Chen, Vijay Kumar Adhikari, Sourabh Jain, and Zhi-Li Zhang. Characterizing roles of front-end servers in end-to-end performance of dynamic content distribution. In *Proc. USENIX/ACM SIGCOMM Internet Measurement Confernece (IMC'11)*, November 2011.
- [7] Yingying Chen, Ratul Mahajan, Baskar Sridharan, and Zhi-Li Zhang. A providers perspective on search response time. In *Proc. ACM SIGCOMM'13*, August 2013.
- [8] Cisco. Visual Networking Index. <http://bit.ly/KXDUaX>.
- [9] Adrian Cockroft. Netflix Cloud Architecture. Velocity conference, 2011.
- [10] Adrian Cockroft, Cory Hicks, and Greg Orzell. Lessons Netflix Learned from the AWS Outage. Netflix Techblog, 2011.
- [11] Ramesh K Sitaraman and Reid W Barton. Method and apparatus for measuring stream availability, quality and performance, February 11 2003. US Patent App. 20,040/136,327.