

Abstract

In the age of low-cost, high-end smartphones, smart televisions (TVs), cheap data plans, and COVID-19 pandemic, video streaming platforms over the Internet have emerged as an essential service for online meetings as well as for entertainment. Most online video streaming services are adaptive in nature, wherein adaptive bitrate (ABR) algorithms are used for maintaining high Quality of Experience (QoE), along with improving the data usage and reducing energy consumption. From the conceptualization of ABR streaming algorithms in 2009, several researchers have proposed different versions of these algorithms and architectural changes to facilitate their implementations on various devices. Nevertheless, there are many open rooms still available and pertinent for improving video streaming systems to boost their scalability to support millions of devices while supporting better on-device resource utilization. This thesis investigates the fundamentals behind different ABR video streaming services while supporting optimizations and improvements in their performance for large-scale usage.

Video currently constitutes the most considerable fraction of the global Internet and mobile traffic. To support the continuous increase in the number of video subscribers, every player on the Internet must adopt improved technologies. For example, during the COVID-19 pandemic, various over-the-top (OTT) streaming providers had decided to drop their top-qualities to reduce the amount of Internet traffic to support the higher number of work-from-home users. This raises the doubt whether the existing ABR algorithms alone can single-handedly handle the excess bandwidth requirements. Although the answer remains elusive, the doubt establishes the importance of the domain. It motivates one to dig deeper to understand how the existing systems work and the specific measures to be adopted for its improvement.

YouTube is one of the major video streaming providers and the pioneer of web-based video streaming service. As per their recent press release, YouTube serves nearly a billion videos to 30 million users every day. Hence, a deep dive into how YouTube serves video to users adaptively is expected to yield answers to several intuitions. As the literature lacks an in-depth analysis of YouTube's adaptive streaming service, in this work, we have first performed thorough experiments to understand the same. We use our homegrown tools to get a request-response pair during video playback in a browser for 500 YouTube videos and have subsequently analyzed them. Our analysis has revealed that YouTube uses an opportunistic up-scale and conservative downscale of video quality based on the network quality and adapts the segment length before changing the rate. We have also observed that YouTube sometimes tries to download the same segment at different qualities, which can cause data wastage. However, our calculations also show that the ensuing data wastage is negligible.

Currently, Google uses Quick Internet UDP Connection (QUIC) as the default transport protocol for all its services from the Chrome browser. So, in the second work, we have carried out experiments to determine its impact on video streaming services, especially on the ABR algorithms. The experiment has two parts. First, we have tested over YouTube and found that QUIC provides better video quality with lower rebuffering than TCP. In the second part, we have compared popular ABR algorithms' performance over QUIC and TCP and found that these ABR algorithms are not well suited for QUIC. On investigating further deep, we have realized that the problem lies in the userspace implementation of QUIC. In this part of the work, we have also explored the impact of video streaming on device's energy efficiency under mobility, particularly in smartphones. We have studied the power consumption of various commodity smartphones while playing YouTube videos. The experiment included observations in different mobility conditions and in different geographic locations. The extensive study has revealed that YouTube does not give any importance to the instantaneous throughput or the radio state of the phones, which may cause a severe impact on power consumption.

With these observations, in the third part of the work we have designed EnDASH - a power smart ABR streaming algorithm that reduces the power consumption during video downloads by tuning the segment download schedule to various device-related parameters, such as cellular radio, battery state, and device speed, etc. EnDASH divides the entire playback session into slots. At the beginning of each slot, it first estimates the average throughput and then selects the maximum playback buffer-length based on the predicted average throughput as well as the observations of the last download slot. By choosing the buffer length to be a function of the channel throughput, EnDASH ensures that power consumption is reduced. The predicted buffer length is subsequently used by EnDASH to decide the video quality of the next video segment. EnDASH has three different learning engines - a) Throughput estimation engine, b) Buffer length prediction engine, and c) Quality Selection engine. The throughput estimation engine is based on Random Forest learning. The buffer length prediction engine and the bitrate selection engine use an actor-critic based recurrent neural network. It is found that EnDASH offers improved energy efficiency than existing ABR algorithm with minor compromise on the QoE.

In the thesis's final contribution, we have developed FLiDASH, a federated adaptive live streaming system that exploits players' locality by forming coalitions. We posit that live streamers of mega-events are often localized, and in most of the cases are connected via a local network. However, current setup does not exploit the local network. Instead, every player acts individually and downloads their own segment. FLiDASH utilizes the local network information to find out players who can communicate without involving the Internet. Of these players the ones having similar quality targets form a coalition. Every player in a coalition contributes to the QoE's improvement by downloading the best quality video that it can download.

We use the moving leader concept, where the leader is responsible for downloading a segment and deciding the leader for the next segment. These two algorithms work in a way so that a player with the lowest download speed gets to download the furthest segment in the long run. Our evaluation reveals that FLiDASH can improve the QoE for a coalition while reducing the server and the network load. We also find that every player gains QoE and average quality using FLiDASH than using any other ABR algorithms.

In summary, this thesis contributes to the literature from two different aspects. First, it analyzes the existing streaming mechanisms and ABR algorithms to check and explore their performance over today's Internet. Second, we propose two advancements over the current systems – (a) from the context of energy efficiency over a smartphone and (b) for improving the QoE and reducing the network overhead during live streaming. Although this thesis opens up and explores new research in video streaming, there are still many unexplored areas, which can be taken up for extending the research and developments in this direction.

Keywords: Video Streaming; Adaptive Bitrate Streaming; OTT Media Services; Live Streaming