MEDIA STREAMING WITH IBM CLOUD MEDIA STREAMING

ABSTRACT

Video streaming, in various forms of video on demand (VOD), live, and 360 degree streaming, has grown dramatically during the past few years. In comparison to traditional cable broadcasters whose contents can only be watched on TVs, video streaming is ubiquitous and viewers can flexibly watch the video contents on various devices, ranging from smart-phones to laptops and large TV screens. Such ubiquity and flexibility are enabled by interweaving multiple technologies, such as video compression, cloud computing, content delivery networks, and several other technologies. As video streaming gains more popularity and dominates the Internet traffic, it is essential to understand the way it operates and the interplay of different technologies involved in it. Accordingly, the first goal of this paper is to unveil sophisticated processes to deliver a raw captured video to viewers' devices. In particular, we elaborate on the video encoding, transcoding, packaging, encryption, and delivery processes. We survey recent efforts in academia and industry to enhance these processes. As video streaming industry is increasingly becoming reliant on cloud computing, the second goal of this survey is to explore and survey the ways cloud services are utilized to enable video streaming services. The third goal of the study is to position the undertaken research works in cloud-based video streaming and identify challenges that need to be obviated in future to advance cloud-based video streaming industry to a more flexible and user-centric service.

KEYWORDS

Video Streaming; Video Transcoding; Video Packaging; Delivery Network; Digital Management Right (DRM); Cloud Computing.

1.1 INTRODUCTION

1.1.1 Overview

The idea of receiving a stream of video contents dates back to the invention of television in the early years of the 20th century. However, the medium on which people receive and watch video contents has substantially changed during the past decade—from conventional televisions to streaming on a wide variety of devices (*e.g.*, laptops, desktops, and tablets) via Internet. Adoption of the Internet-based video streaming is skyrocketing to the extent that it has dominated the whole Internet traffic. A report by Global Internet

Phenomena shows that video streaming has already accounted for more than 60% of the whole Internet traffic [109]. The number of Netflix[1] subscribers has already surpassed cable-TV subscribers in the U.S. [128].

Nowadays, many Internet-based applications function based on video streaming. Such applications include user-generated video contents (*e.g.*, those in YouTube[2], Vimeo[3]), live streaming and personal broadcasting through social networks (*e.g.*, UStream[4] and Facebook Live[5]), over the top (OTT) streaming (*e.g.*, Netflix and Amazon Prime[6]), e-learning systems [120] (*e.g.*, Udemy[7]), live game streaming platform (*e.g.*, Twitch[8]), video chat and conferencing systems [87], natural disaster management and security systems that operate based on video surveillance [30], and network-based broadcasting channels (*e.g.*, news and other TV channels) [67].

As video streaming services grow in popularity, they demand more computing services for streaming. The uprising popularity and adoption of streaming has coincided with the prevalence of cloud computing technology. Cloud providers offer a wide range of computing services and enable users to outsource their computing demands. Cloud providers relieve video streaming providers from the burden and implications of maintaining and upgrading expensive computing infrastructure [84]. Currently, video stream providers are extensively reliant on cloud services for most or all of their computing demands [53]. The marriage of video streaming and cloud services has given birth to a set of new challenges, techniques, and technologies in the computing industry.

Although numerous research works have been undertaken on cloud-based video streaming, to our knowledge, there is no comprehensive survey that shed lights on challenges, techniques, and technologies in cloud-based video streaming. As such, the essence of this study is to *first*, shed light on the sophisticated processes required for Internet-based video streaming; *second*, provide a holistic view on the ways cloud services can aid video stream providers; *third*, provide a comprehensive survey on the research studies that were undertaken in the intersection of video streaming and cloud computing; and *fourth* discuss the future of cloud-based video streaming technology and identify possible avenues that require further research efforts from industry and academia.

Accordingly, in this study, we first explain the way video streaming works and elaborate on each process involved in it. Then, in Section 1.3, we provide a holistic view of the challenges and demands of the current video streaming industry. Next, in Section 1.4, we discuss how cloud services can fulfill the demands of video streaming, and the survey the research works undertaken for that purpose. In the end, in Section 1.5, we discuss the emerging research areas in the intersection of video streaming and cloud computing.

1.1.2 Cloud Computing for Video Streaming

To provide a high Quality of Experience (QoE) for numerous viewers scattered worldwide with diverse display devices and network characteristics, video stream providers

---

1. https://www.netflix.com
2. https://www.youtube.com
3. https://www.vimeo.com
4. https://video.ibm.com
5. https://www.facebook.com
6. https://www.amazon.com
7. https://www.udemy.com
8. https://www.twitch.tv

commonly pre-process (*e.g.*, pre-transcode and pre-package) and store the video contents of multiple versions [83]. As such, viewers with different display devices can readily find the version matches their devices. An alternative approach to pre-processing videos is to process them in a lazy (*i.e.*, on-demand) manner [83, 84]. In particular, this approach can be useful for videos that are rarely accessed. Recent analytical studies on the statistical patterns of accessing video streams (*e.g.*, [28]) reveal that the videos of a repository are not uniformly accessed. In fact, streaming videos follows a long-tail pattern [97]. That means, many video streams are rarely accessed and only a small portion (approximately 5%) of the videos (generally referred to as *hot video streams*) are frequently accessed. Both pre-processing and on-demand approaches need the video streaming provider to provide an enormous computing facility.

Maintaining and upgrading in-house infrastructure to fulfill the computational, storage, and networking demands of video stream providers is costly. Besides, it is technically far from the mainstream business of stream providers, which is video content production and publishing. Alternatively, cloud service providers, such as Amazon Cloud (AWS), Google Cloud, and Microsoft Azure, can satisfy these demands by offering efficient services with a high availability [82]. Video stream providers have become extensively reliant on cloud services for most or all of their computing demands. For instance, Netflix has outsourced the entirety of its computational demands to Amazon cloud [101].

In spite of numerous advantages, deploying cloud services has presented new challenges to video stream providers. In particular, as cloud providers charge their users in a pay-as-you-go manner for their services [111], the challenge for stream providers is to minimize their cloud expenditure, while offering a certain level of QoE for their viewers.

Numerous research works have been conducted to overcome the challenges of video streaming using cloud services. For instance, researchers have studied the cost of deploying cloud services to transcode videos [34, 82], the cost-benefit of various video segmentation models [65, 83], applying customized scheduling methods for video streaming tasks [88, 34], and resource (Virtual Machine) provisioning methods for video streaming [83, 64]. Nonetheless, these studies mostly concentrate on one aspect of video streaming processing and how that aspect can be performed efficiently on the cloud. Further studies are required to position these works in a bigger picture and provide a higher level of vision on the efficient use of cloud services for the entire workflow of video streaming.

1.2 THE MYSTERY OF VIDEO STREAMING OPERATION

1.2.1 Structure of a Video Stream

As shown in Figure 1.1, a video stream consists of a sequence of multiple smaller segments. Each segment contains several *Group Of Pictures* (GOP) with a segment header at the beginning of each GOP. The segment header includes information such as the number of GOPs in that segment and the type of the GOPs. A GOP is composed of a sequence of frames. The first frame is an I (intra) frame, followed by several P (predicted) and B (bi-directional predicted) frames. A frame in a GOP is divided into multiple *slices* and each slice consists of several *macroblocks* (MB). The MBs are considered as the unit for video encoding and decoding operations.

Two types of GOP can exist, namely closed-GOP and open-GOP. In the former, GOPs are independent, *i.e.*, there is no relation between GOPs. Therefore, closed-GOPs can be processed independently. Alternatively, in open-GOP, a GOP is dependent on another GOP, hence, cannot be processed independently.

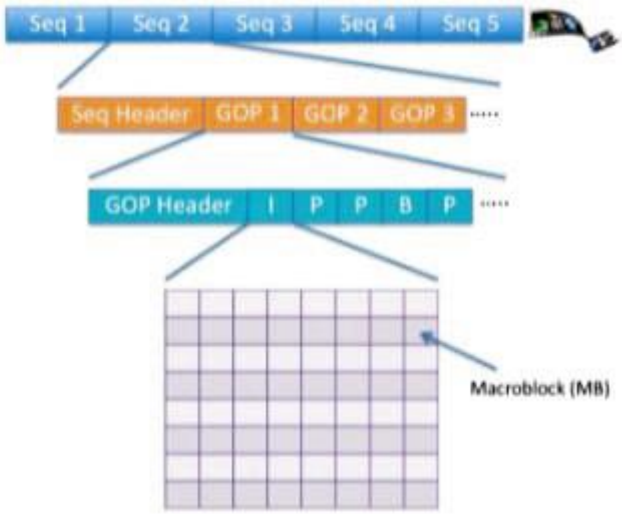

FIGURE 1.1 The structural overview of a video stream. A video stream contains multiple segments and each segment contains multiple GOPs. A frame in a GOP includes a number of macroblocks (MB).

To process video streams, they can be split at different levels, namely segment level, GOP level, frame level, slice level, and macroblock level. Sequence-level contains several GOPs that can be processed independently. However, due to the large size of a sequence, its transmission and processing time become a bottleneck [75]. Processing at the frame, slice, and macroblock levels implies dealing with the spatio-temporal dependencies that makes the processing complicated and slow [75]. In practice, video stream providers generally perform processing at the segment or GOP levels. That is, they define a segment or a GOP as a unit of processing (*i.e.*, a *task*) that can be processed independently [65].

1.2.2 Video Streaming Operation Workflow

In both Video On Demand (VOD) (*e.g.*, Hulu, YouTube, Netflix) and Live streaming (*e.g.*, Livestream[9]), the video contents generated by cameras have to go through a complex workflow of processes before being played on viewers' devices. In this section, we describe these processes.

Figure 1.2 provides a bird's-eye view of the main processes performed for streaming a video—from video production to playing the video on viewers' devices. These processes collectively enable raw and bulky videos, generated by cameras, be played on a wide variety of viewers' devices in a real-time manner and with the minimum delay. It is noteworthy that, in addition to these processes, there are generally other processes to enable features such as video content protection and cost-efficiency of video streaming. In the rest of this section, we elaborate on the main processes required for video streaming. Additional processes (*e.g.*, for video content protection and analysis of video access rates) are discussed in the later parts of the paper (Sections 1.3.6, 1.3.7, and 1.3.8, respectively).

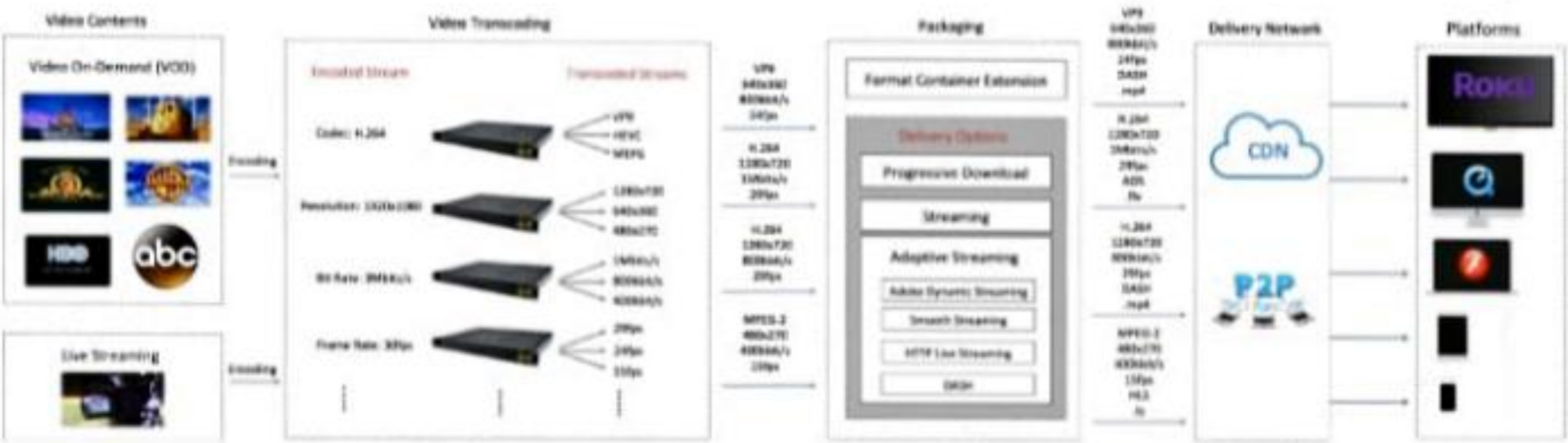

FIGURE 1.2 A bird's-eye view to the workflow of processes performed on a video stream from the source to the viewer's device. The workflow includes compression, transcoding, packaging, and delivery processes.

1.3 VIDEO STREAMING: CHALLENGES AND SOLUTIONS

1.3.1 Overview

In the previous section, we described the high-level workflow of processes for video streaming. However, there are various approaches and challenges to accomplish those processes. In this section, we elaborate on all issues, challenges and current solutions that are related to video streaming.

Figure 1.3 presents a taxonomy of all the issues that a video streaming system needs to deal with. The taxonomy shows different types of video streaming (*i.e.*, VOD and live streaming). It discusses variations of video transcoding and packaging processes along with important streaming protocols. The taxonomy shows possible ways of distributing video content, such as CDN [146] and P2P [92]. Video content protection is further detailed on video privacy for video streaming [144] and copyright issues, known as Digital Right s Management (DRM) [91]. The taxonomy covers analytical studies that have been conducted to understand the viewers' behaviors and discovering their access patterns to videos streams. Last, but not the least, the taxonomy covers the storage issue for video streaming and possible strategies for video streaming repository management, such as in-house storage versus outsourcing solutions via cloud storage services.

FIGURE 1.3 Taxonomy of all aspects we need to deal with in video streaming.

1.3.2 Video Streaming Types

The taxonomy shown in Figure 1.4 expresses possible ways a video streaming service can be provided to viewers. More Specifically, video streaming service can be offered in three main fashions: *Video On Demand* (VOD) streaming, *live* streaming, and *live-to-VOD* streaming.

FIGURE 1.4 Taxonomy of different types of video streaming.

1.3.2.1 VOD Streaming

In VOD streaming, which is also known as Over-The-Top (OTT) streaming, the video contents (*i.e.*, video files) are already available in a video streaming repository and are streamed to viewers upon their requests. By far, VOD is the most common type of video streaming and is offered by major video streaming providers, such as YouTube, Netflix, and Amazon Prime Video.

Some VOD providers (*e.g.*, Netflix and Hulu) offer professionally made videos and movies that are subscription-based and viewers are generally required to pay a monthly fee to access their service. Alternatively, other VOD services (*e.g.*, YouTube) operate based on user-provided videos. Such services are generally advertisement-based and free of charge. VODs have also applications in e-learning systems [120], Internet television [107], and in-flight entertainment systems [40].

1.3.2.2 Live Streaming

In live video streaming, the video contents are streamed to the viewer(s), as they are captured by a camera. Live video streaming has numerous applications, such as event coverage and video calls. The live video streaming used in different applications have minor differences that are mainly attributed to the buffer size on the sender and receiver ends [76]. In general, a larger buffer size causes a more stable streaming experience but imposes more delay. This delay can be tolerated in live broadcasting applications, however, delay-sensitive applications (*e.g.*, video telephony) cannot bear the delay, thus need a shorter buffer size. Accordingly, live streaming can have four variations as follows:

(A) **One-to-one** (unicast) streaming is when a user streams video contents to another user. This type is primarily used in video chat and video call applications that require two live streams, one from each participant. This streaming type requires short delays to enable smooth conversation between participants. As such, these applications generally operate with a short buffer size and low picture quality to make the delay as