

Building Cloud-ready Video Transcoding System for Content Delivery Networks(CDNs)

Zhenyun Zhuang and Chun Guo

Speaker: 饒展榕

Outline

- INTRODUCTION
- BACKGROUND
- MOTIVATION
- DESIGN
- CONCLUSION

INTRODUCTION

- Video streaming traffic of both VoD (Video on Demand) and Live is exploding.
- Given the vast number of video stream formats and transmission protocols for supporting varying types of playback terminals, video content providers often need to transcode videos to multiple formats in order to stream to different types of users.

- Being time-sensitive and requiring high bandwidth, video streaming exerts high pressure on underlying delivery networks.
- Content Delivery Network (CDN) providers can help their customers quickly and reliably distribute stream contents to end users.

- In addition to distributing video streams, CDN providers typically allow their customers to perform video transcoding on CDN platforms.

BACKGROUND

- A typical CDN infrastructure consists of Ingesting Servers (for accepting customer contents), Origin Servers (for serving edge servers) and Edge Servers (for serving end users directly), forming a layered structure.
- Today's video playback devices, particularly mobile ones, are highly diversified.

- Each of these devices has particular screen size , prefer different video format , and support only certain transmission protocols.
- All these properties require the transcoding service to convert videos to multiple formats.
- Cloud computing provides the elastic compute capability.
- By leveraging the power of cloud computing, a service provider can better adapt to changing needs for better performance and reduced cost.

MOTIVATION

- First, as the first step towards delivering, videos/streams need to be uploaded/ingested to CDNs before being transcoded.
- With multiple available ingesting servers typically provided by CDNs, there is a need to select the optimal ingesting server.
- Second, CDNs may involve large bulk of transcoding jobs.

- Given such characteristics, the transcoding system needs to be scalable.
- CDNs typically allow customers to do region-based provisioning.
- Because of this, the physical locations of chosen transcoding servers can cause different bandwidth and time cost.

A. Video/stream Ingesting

- Video/stream ingesting consists of both VoD video uploading and Live stream ingesting. Customers have to ingest their videos/streams to CDN ingesting servers before the videos can be transcoded and streamed.

B. Diversified Transcoding Jobs

- After videos/streams are ingested into CDN, the transcoding system takes control for converting to various video/streaming formats.
- It is important to note that a simple balancing algorithm that randomly assigns jobs to transcoding nodes won't work well if the incoming transcoding jobs vary in terms of the compute resource is required.

- A desired balancing algorithm needs to be able to meter the workload associated with each transcoding job and has to consider the following factors at the minimum: workload and node capabilities.
- For VoD streaming, after transcoding, the converted videos need to be moved to provisioned streaming servers as determined by the customer profiles.

- For live streaming, the transcoded streams need to be continuously transmitted to the provisioned live streaming servers.
- Because of this, choosing different transcoding nodes have the impact on two further things: bandwidth cost and moving latency.

C. Bursty Transcoding Load

- During some periods only a few videos come while during other periods many more videos come.
- With such a bursty behavior, a transcoding system needs to be able to dynamically allocate/de-allocate transcoding nodes both for the concerns of cost and job processing time.

IV.DESIGN

- Author now present the design of C3.
- Author first give a high-level design overview, followed by the software architecture and the two major components of C3.

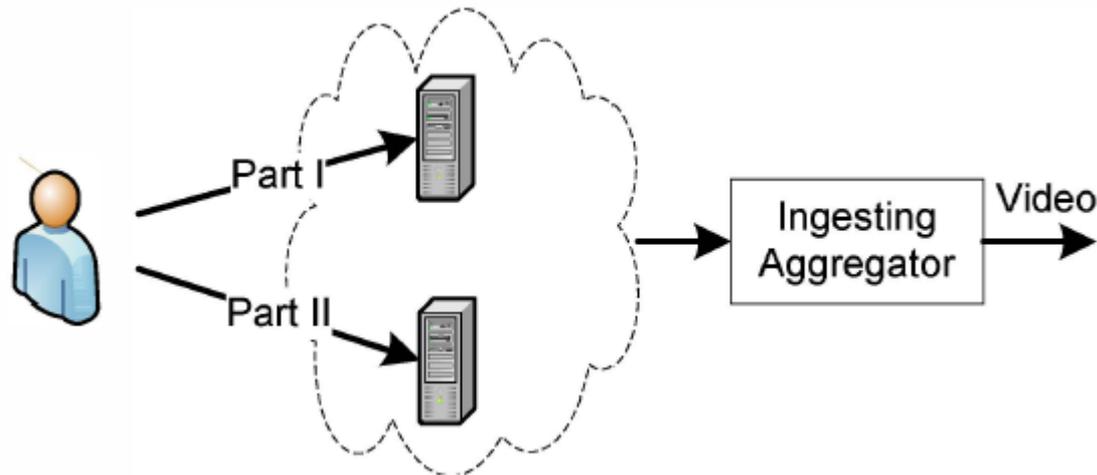
A. Design Overview

- First, the video is moved into CDN infrastructure by being uploaded to CDN ingesting servers.
- For live streams, the ingesting process is similar; but instead of being uploaded as video files, live streams are continuously ingested into the ingesting servers.
- Second, after a video is uploaded into CDN servers, C3 will invoke the second component of Transcoding Cloud to perform the transcoding based on the customer's profile.

- A customer's profile contains information including the video output formats, output screen size, number of multi-bitrate levels, provisioned target streaming servers, and so on.

B. Software Architecture

- Customer videos are uploaded to a optimal ingesting server which is determined by a specialized DNS component.



- GeoDNS answers a DNS request by providing a server that is geographically closer to the requesting client.
- To deal with the bursty behavior of incoming transcoding tasks, C3 can dynamically spin up/down transcoding servers to handle the transcoding more smoothly.

C. C3 Components

- 1) Ingesting Cloud: The ingesting cloud component of C3 relies on GeoDNS service, but it is designed to use multiple ingesting servers at the same time.
- 2) Transcoding Cloud: One of the key features of C3 is that the number of transcoding servers that are actively transcoding is dynamically adjusted to the current transcoding load.

V. EVALUATION

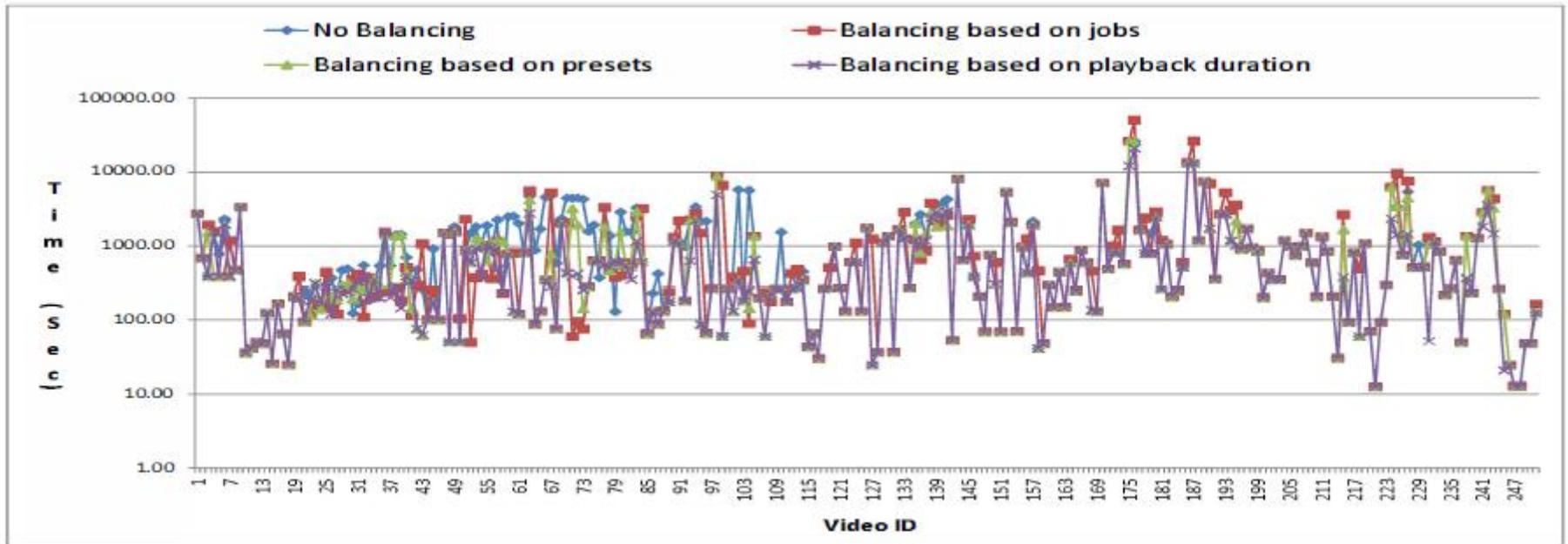
- Author built a prototype with Microsoft Expression Encoder 4 (EE4) [4].
- EE4 can encode videos for various types of devices and the web Silverlight videos with customizable templates .

A. Ingesting Cloud

- We first evaluate the performance of the Ingesting Cloud.
- Starting from a single ingesting server, we gradually add more ingesting servers into the cloud.

B. Transcoding Cloud

- Depending on the transcoding workload, C3 may spin up more transcoding servers to handle the jobs.



- Apparently, the last algorithm is working at the finest level, and we name it playback-based.

V I. CONCLUSION

- In this work, we propose a cloud-based video transcoding system for CDNs.
- The design considers the characteristics of CDNs and transcoding requirements and is able to dynamically adjust the transcoding tasks to reduce operation cost.
- We also build a prototype for evaluating the proposed the design.