Cloud Transcoding for Mobile Video Content Delivery

Adriana GARCIA, and Hari KALVA, Dr.
Department of Computer & Electrical Engineering and Computer Science, Florida Atlantic University

***Abstract*—This study analyzes a use case where computer clouds are used in the transcoding of media content and points to their advantages in a low cost media distribution solution such as HTTP Live.**

I. INTRODUCTION

Mobile device communication demands adaption to varying bandwidths, different communication links and a multitude of screen sizes [3]. Transcoding, the process used for the conversion of one data representation to another type of representation including transcoding within media types (conversion from one standard into another), between media types (such as speech to text for example) and further data compression [3], is one of the methods that allows such adaption by adjusting video resolution to match a mobile's reduced display, for example.

Apple's HTTP Live Streaming Protocol supports the streaming of audio or video to mobiles such as iPhones and iPads, among others, and the transfer of video on demand with encryption and authentication. The HTTP Live streaming architecture consists of an encoding part, a distribution part and a client. The encoder takes the input stream from a video source and transforms it into an MPEG2 transport stream, dividing it into 10 second segments during/after the encoding stage. The content is delivered through this transport stream segments together with a playlist file (M3U8). These files are stored on a web server and accessed by the client using a compatible player software or web browser [5]. The earliest content availability time for access is tightly bound to the speed of the preceding data encoding stage.

The exploration of cloud computing to harness the large processing power of less expensive or on-demand hardware for this purpose could increase the speed at which content is available to the end-user after the encoding process. In this study, Hadoop streaming jobs are used as a way to exploit the power of the computer cloud to speed-up server content availability in the case when a video playlist is made available to/requested by users who have a transcoding requirement, like mobile users who seek resolution reduction.

II. RELATED CONCEPTS AND WORK

Although there is no single definition of what cloud computing stands for, several characteristics are common to all definitions, including the presence of multiple distributed computing resources that are made available to the end consumer which are greater and less costly than those provided by a single machine. Specifically in [2] "Computing in the "cloud" eludes to ubiquitous and inexhaustible on-demand IT resources accessible through the Internet." An example of such services is Amazon's Web Services cloud offering.

In this section, the concepts related to cloud/distributed computing and its reported use as a transcoding tool are presented.

A. Hadoop Map/Reduce

Hadoop Map/Reduce is a software framework for writing applications which process vast amounts of data (multi-terabyte data-sets) in-parallel on large clusters (thousands of nodes) of commodity hardware in a reliable way. It implements the MapReduce model suggested by Google [1] in free distributions such as Apache Hadoop, Yahoo Hadoop and Cloudera.

MapReduce's value lies in the possibility of automatic parallelization of activities and their distributed execution.

B. Implementations

Solutions prior to Hadoop such as x264farm approached encoding in a distributed environment. x264farm utilizes x264, a free h264/avc encoder [4] to accomplish encoding into the H.264 format.

Commercially, today video transcoding on cloud environments is available through HDCloud.com and Encoding.com, which provide flexible but proprietary cloud based video transcoding services integrated with Amazon Web Services's EC2, S3, and CloudFront CDN services.

III. EXPERIMENTS

In the experimental set up shown in Figure 1 a Hadoop streaming job is launched on demand by the server when a video request mandates it. The computer cloud relieves the server from encoding and optionally the segmentation of the input video. After encoding/transcoding the resulting video/video segments are stored on a location accessible to both parties and the normal HTTP Live streaming distribution is continued normally.

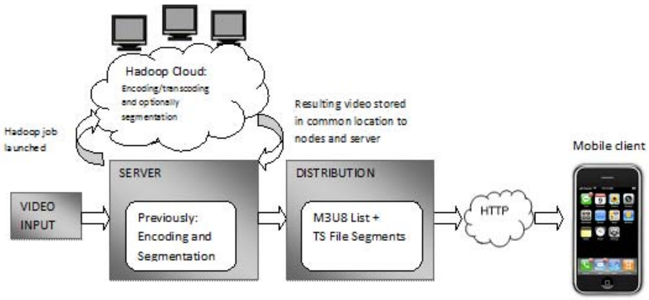

Fig. 1. System setup

For the purpose of exploring the server's content availability speed-up consequence of implementing cloud computing concepts in the encoding stage of the video content delivery process the following use case was explored:

A text file representing a video playlist derived from a DVD's content is processed under three different experiments:

- When the DVD content is split by context in this case, the DVD's chapters
- When the DVD is split into 10s segments consistent with HTTP Live streaming recommendation [5].
- When the content is split into a number of pieces equal to the number of processing nodes available in order to explore speed-up thanks to distributed processing.

The experiments are run in a three node testbed with 2047MiB System memory and two 3.8GHz Intel Xeon processors each; all connected to an external SCSI storage where the input and output data is stored and is accessible to all nodes concerned.

Transcoding was achieved through the use of Hadoop streaming jobs that utilize the ffmpeg tool.

IV. Results

The results for the three tests are shown in Figure 2 to Figure 4.

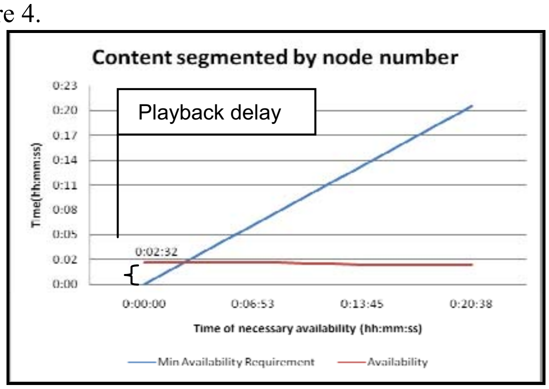

Fig. 2. Availability time for a DVD chapter (20m 18s long) split in 3 equal parts. First segment delay: 2m 32s.

Figure 2 shows that for a content partition equal to the number of nodes, availability is delayed by a significant amount (more than 2 minutes). Counterbalancing this fact, all the content is available almost at once, which points to the possibility of supporting trick modes such as rewinding and fast forwarding.

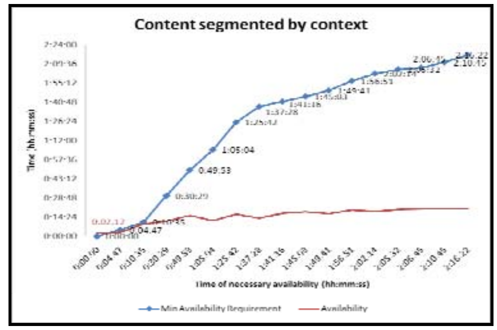

Fig. 3. Availability time for all the segments in the DVD (2h 18m in long). First segment delay: 2m 12s. Note: segments by context have different durations.

Figure 3 shows quite a random behavior as context specific segments are not equal in size and cannot be easily distributed among the nodes in a balanced manner.

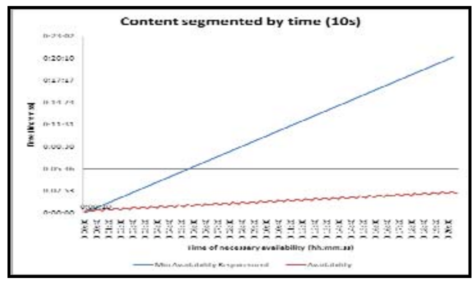

Fig. 4. Availability time for a DVD chapter (20m 18s long) split in 124 10s segments. First segment delay: 10s.

Figure 4 shows an availability curve for the 10s segments well below the minimum playback requirement. Its serrated behavior can be explained by a faster availability of some segments due to the load distribution among the participating nodes.

V. Conclusions

The availability results presented above reveal the potential advantage of the use of cloud computing in a server transcoding phase that can reduce storage size (data is not saved in raw form but actually compressed) as well as enable the access to a wide variety of clients with different requirements. Its use beyond an exclusive HTTP Live scenario is worth exploring in the future.

References

[1] J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," in Proceedings of OSDI '04: 6th Symposium on Operating System Design and Implementation, San Francisco, CA, Dec. 2004.

[2] D. Hilley, "Cloud Computing: A Taxonomy of Platform and Infrastructure-level Offerings", Georgia Institute of Technology, April (2009)

[3] M. Fouquet, H. Niedermayer, G. CarleCloud, "Computing for the Masses" Technische Universität München, Proceedings of the 1st ACM workshop on User-provided networking: challenges and opportunities, Rome, Italy (2009), p.31-36.

[4] L. Merrit, R. Vanam, "x264: A high performance H. 264/AVC encoder" Penn State University

[5] iOS Reference Library, "HTTP Live Streaming-Overview" Online at: http://developer.apple.com/iphone/library/documentation/networkinginternet/conceptual/streamingmediaguide/introduction/introduction.html Last accessed: 08/04/2010