A Benchmark to Evaluate Mobile Video Upload to Cloud Infrastructures

Afsin Akdogan, Hien To, Seon Ho Kim, Cyrus Shahabi
Integrated Media Systems Center
University of Southern California

Introduction

How to evaluate the performance of mobile video applications on these cloud infrastructures and select an appropriate set of resources for a given application?

Given:

1. A massive amount of mobile videos' metadata.
2. Cloud providers (e.g., Amazon, Google, Microsoft) allow users to lease computing resources with varying disk, network and CPU capacities.

Benchmark Design

Upload videos with metadata to cloud has three stages:

1) **network** to transfer videos from mobile clients to the cloud servers

2) **database** to insert metadata about the uploaded videos

3) **video transcoding** to change the resolution of uploaded videos to use less storage and bandwidth

We define a **single (cross-resource) metric** to evaluate the uploading workflow of video applications on cloud.

Type	Amazon EC2		Microsoft Azure		Google Compute	
	price ($/hour)		price ($/hour)		price ($/hour)	
	smallest	largest	smallest	largest	smallest	largest
General purpose (*m*)	0.07	0.56	0.02	0.72	0.077	1.232
Compute optimized (*c*)	0.105	1.68	2.45	4.9	0.096	0.768
Memory optimized (*r*)	0.175	2.8	0.33	1.32	0.18	1.44
Disk optimized (*i*)	0.853	6.82	-	-	-	-
Micro (*t*)	0.02	0.044	-	-	0.014	0.0385
GPU	0.65	0.65	-	-	-	-

Tab. 1) Categorization of the server types with the prices (dollars/hour) of the smallest and the largest servers at each group.

Type	Memory	CPU	Disk	Network Bandwidth
m-small	3.75 GB	1 VCPU	4 GB SSD	no info.
c-small	3.75 GB	2 VCPUs	32 GB SSD	no info.
r-small	15.25 GB	2 VCPUs	32 GB SSD	no info.
i-small	30.5 GB	4 vCPUs	800 GB SSD	no info.
m-large	30 GB	8 VCPU	160 GB SSD	no info.
c- large	60 GB	32 VCPUs	640 GB SSD	no info.
r- large	244 GB	32 VCPUs	2 x 320 GB SSD	10 Gigabit Ethernet
i- large	244 GB	32 vCPUs	8x800 GB SSD	10 Gigabit Ethernet

Tab. 2) Hardware specifications of the smallest and largest servers of 4 types on EC2

MediaQ Architecture

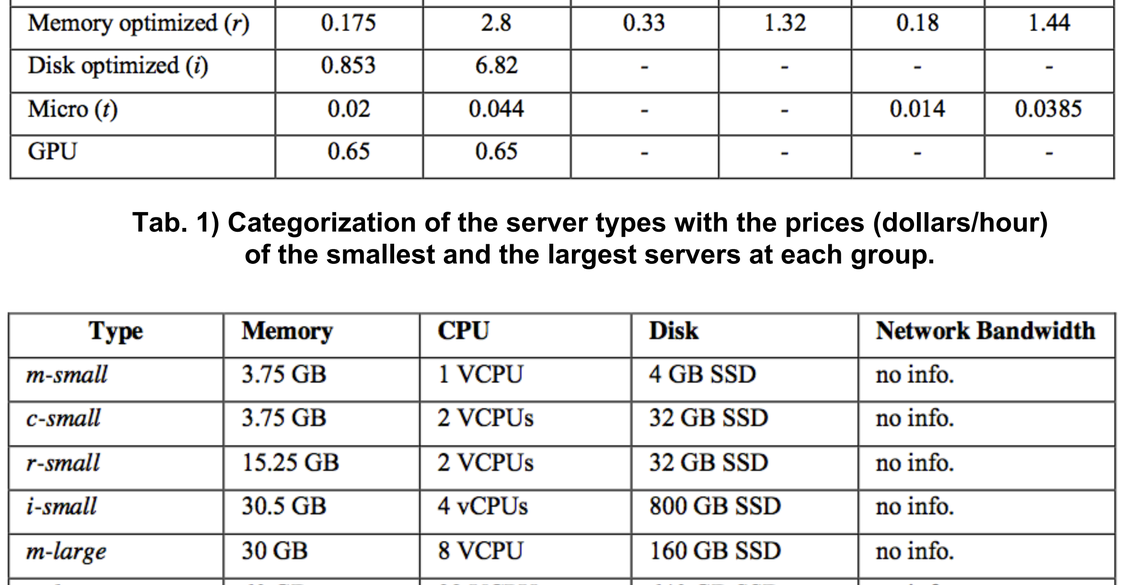


Fig. 1) Overall architecture of MediaQ system

Performance Evaluation

Overall Cost Analysis

Other than general purpose instance, the performance difference between the optimized servers (c, r, i) is not significant even though the prices vary widely.

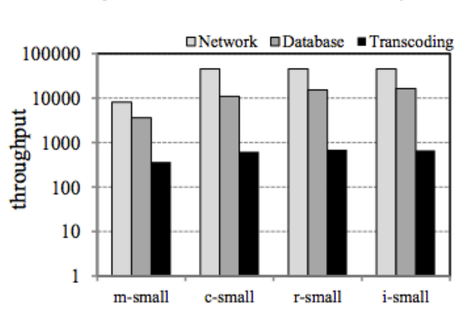


Fig. 2a) Comparison of system components on smallest servers of 4 server types

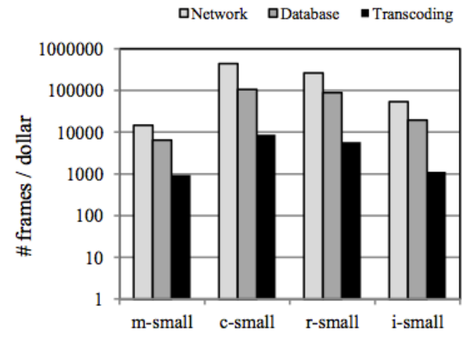
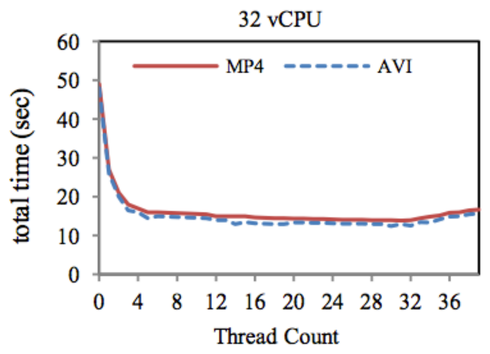


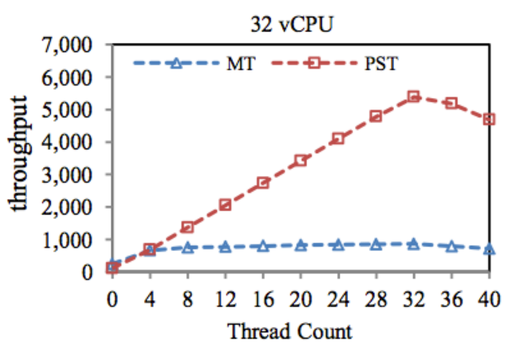
Fig. 2b) Number of frames that can be processed for each dollar spent (log-scale).

Transcoding Analysis

Multi-threading: While multi-thread technique suffers from low parallelism, parallel-single thread can utilize available CPUs better.



a) multi-thread ffmpeg on the same video for different video output types



b) multi-thread (MT) vs. parallel single-thread (PST) ffmpegs

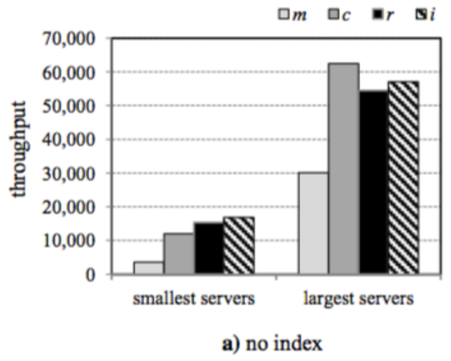
Fig. 3) Scale up performance of transcoding (230MB video)

Reducing video quality: Throughput increases significantly as the resolution decreases. However, the percentage improvement diminishes when the output video resolution becomes too small.

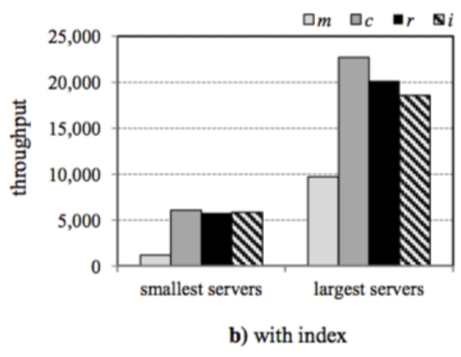
Output resolution	MP4		AVI	
	throughput	% improvement	throughput	% improvement
480x270	623	-	626	-
240x136	842	35%	839	34%
120x68	980	57%	982	57%
60x34	1038	66%	1048	67%

Tab. 2) Transcoding throughput for mp4 and avi types with various output resolutions. The input video is in .m4v format with 960x540 resolution.

Database Analysis



a) no index



b) with index

Fig. 4) Insertion throughput on the smallest and largest servers in 4 instance families

Conclusion and Future Work

- ✓ This study identifies the main cost components of a multimedia system
- ✓ Transcoding is the major bottleneck, which can be more efficient by trading off the quality the video
- ✓ In the future, we aim to partition the data across multiple servers and to provide high scalability

USC Viterbi School of Engineering
Integrated Media Systems Center

With in-kind sponsorship from Google Microsoft

IMSC Retreat 2015