Analyzing Compute vs. Storage Tradeoff for Video-aware Storage Efficiency

Atish Kathpal, Mandar Kulkarni
Ajay Bakre

Advanced Technology Group
NetApp Inc.
Context and Overview

- Trend: Number of devices to view videos is increasing.
Typical video distribution workflow
Data Explosion: 100-1000s versions of the same video

Poor Dedup: Versions of the same video do not deduplicate
Intuition

Eliminate unpopular video versions, transcode on the fly

Optimize on costs
Proposed video distribution workflow

- Ingest
- Transcode
- Master Copy
- Eliminate Selected Videos
- flv
- mpeg4
- 3g
- Master Copy
- Re-transcode
- Request 3g 5mbps

© 2012 NetApp, Inc. All rights reserved.
Transcoding-on-fly Trade-offs

- Storage Cost
- Compute Cost
- Latency
Towards Cost Effective Elimination

- Assume, known $T_{\text{access}}$

- Cost effective to eliminate if:

\[ \Delta = T_{\text{access}} \]

\[
\text{Cost of storing for } T_{\text{access}} \text{ time} > \text{Cost of transcoding}
\]
Towards Cost Effective Elimination

Cost of Storage (Cloud)

Time when $C_S = C_T$ (Elimination Metric)

- Cost of Storage (Cloud)
- Cost of Transcode

© 2012 NetApp, Inc. All rights reserved.
Elimination Metric (EM)

- Time (since video creation) when cost of storage is equal to cost of transcoding.

- Assuming cloud based cost model

\[ EM \text{ (Time)} \times \text{Storage Cost per unit Time} = \text{Transcode Cost} \]

\[ \Rightarrow EM \text{ (Time)} = \frac{\text{Transcode Cost}}{\text{Storage Cost per unit Time}} \]

*For mathematical derivation, refer paper.*
Example: Elimination Metric (EM)

- Transcoding time = 1hr, Storage required = 2GB
- Compute cost = 1000$ per hr
- Storage cost = 1$ per GB per hr
- EM = Transcode cost/Storage Cost per unit time
  = (1hr x 1000$ per hr)/(2GB x 1$ per GB per hr)
  = 500 hours

The video is a good candidate for elimination

Time for next access
\( T_{\text{access}} \)  

Elimination Metric (EM)

\( T = 0 \)

600 hours

500 hours
Analysis
Experimental setup

Server
Ubuntu 10.04, 4GB RAM
3GHz dual core x86_64 CPU
Transcoding Software: FFmpeg

Storage
280GB LUN over FC (8Gbps)

Cost Model: Amazon EC2 + S3
Dataset

8 Master *Videos (avi, 1920x1080p) 30s clips

Transcode

376 Derived Versions (8 resolutions, 6 containers)

- Slow/Fast Motion
- Zoom
- Blur
- Limited/Vivid Colors
- Talking Head
- ..

- **Resolutions:** 576p, 720p, 1080p, etc
- **Containers:** avi, flv, mpeg, vob, mov, mp4

Observations

Storage footprint:
- ~8:1
  - Masters
  - Derived Versions

Memory:
- ~10%

CPU:
- ~95%

0% block based Deduplication
EM v/s Video Codecs

EM depends on video codec.

* Using cost of Amazon High-CPU Medium Instance & S3 cost for the first 1TB
EM decreases for videos with higher resolutions for specific codecs

* Using cost of Amazon High-CPU Medium Instance & S3 cost for the first 1TB
EM v/s Different Videos

EM depends on the content of the video

Video Characteristics

* Using cost of Amazon High-CPU Medium Instance & S3 cost for the first 1TB
Strategies for Elimination of Video versions

- **Never Eliminate:** Store all versions of all videos
  - Pay for storing all videos
  - Good if all videos are popular

- **Always Eliminate:** Eliminate all derived versions
  - Pay for transcoding the videos
  - Good if all videos are unpopular

- **Eliminate Cost Effectively:** Leverage EM!
  - Eliminate unpopular videos
    - (Videos with $T_a > EM$)
  - Store popular videos
Challenge

Time for video’s next access is unknown
Solution Directions

- Predict patterns based on history
  - Statistics
  - Eliminate the video based on predicted access time (if $T_{Predicted} > EM$)

- Leverage Online Algorithms
  - Draw an upper bound on costs without prediction of access patterns
  - Eliminate video when time since last access exceeds EM
Elimination Strategies

Cost Factor wrt Optimal

Time since previous access

Optimal
- Online Variant
- Never Eliminate
Summary & Future Work

- Transcoding on the fly
- Cost Effective Elimination
- Elimination Metric
- Elimination Strategies

Next Steps
- Prediction of EM without transcoding
- Evaluation with video access patterns
- Handle latencies due to transcode on the fly
- EM for a datacenter
References

- Adams et al. “Maximizing efficiency by trading storage for computation,” HotCloud’09
- Gürsun et al. “Describing and forecasting video access patterns”. Infocom 2011
- Veeraraghavan et al. “quFiles: The Right File at the Right Time”. ACM Transactions on Storage
- Gürsun et al. “Describing and forecasting video access patterns”. Infocom 2011
- Content-Aware Storage Optimization
  http://www.ocarinanetworks.com/resources/147
Q&A