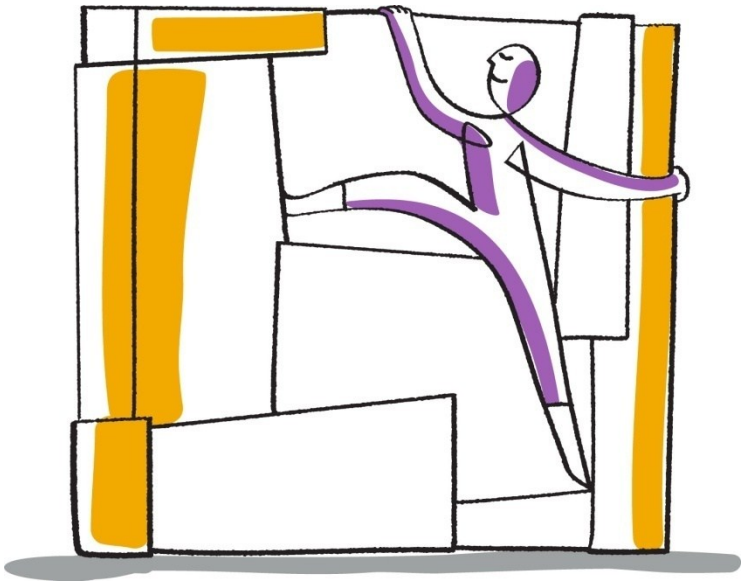
Analyzing Compute vs. Storage Tradeoff for Video-aware Storage Efficiency

Atish Kathpal, Mandar Kulkarni
Ajay Bakre

Advanced Technology Group
NetApp Inc.



© 2012 NetApp, Inc. All rights reserved.

Context and Overview

- Trend: Number of devices to view videos is increasing.

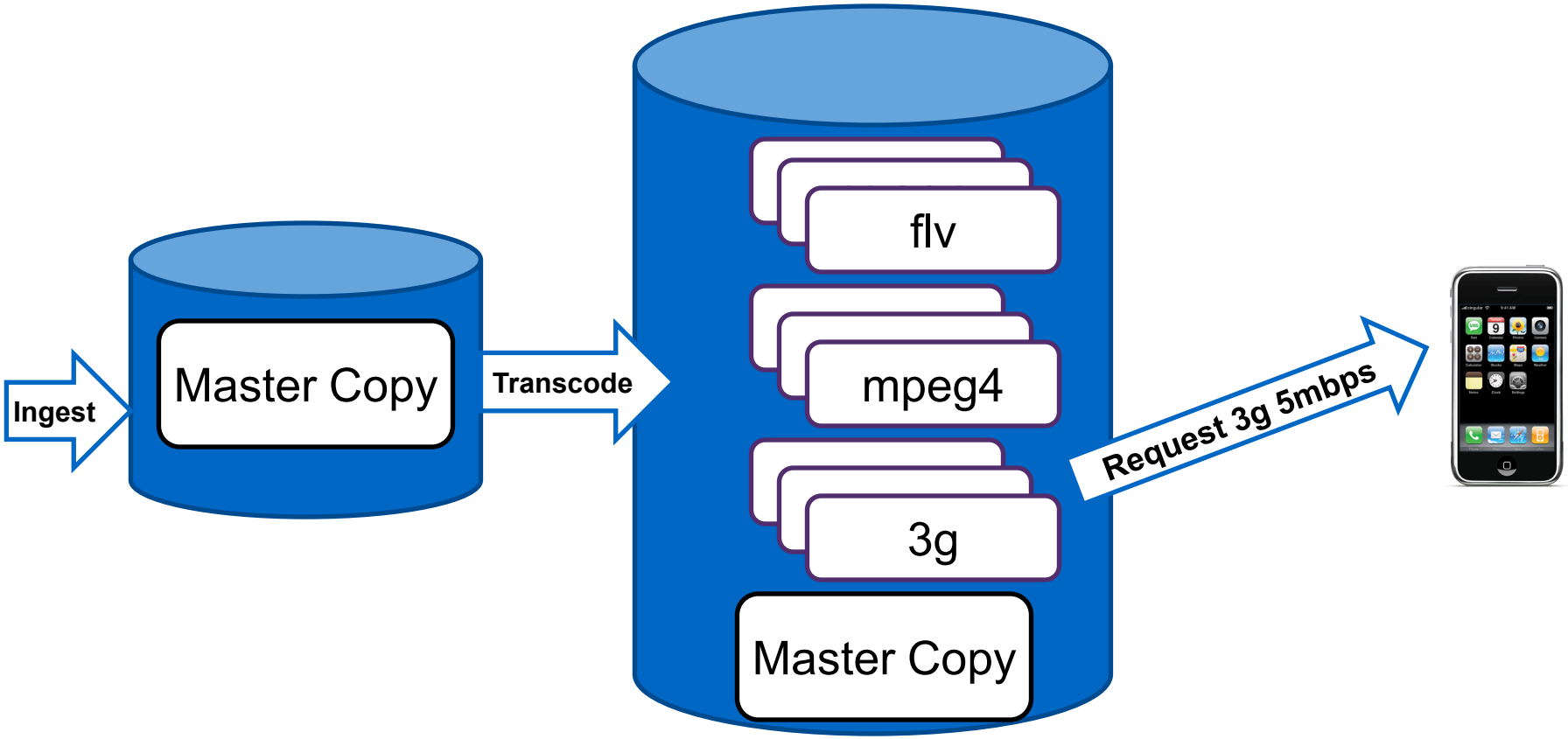







© 2012 NetApp, Inc. All rights reserved.

Typical video distribution workflow



© 2012 NetApp, Inc. All rights reserved.

Issues

Data Explosion: 100-1000s versions of the same video

Poor Dedup: Versions of the same video do not deduplicate

© 2012 NetApp, Inc. All rights reserved.

Intuition

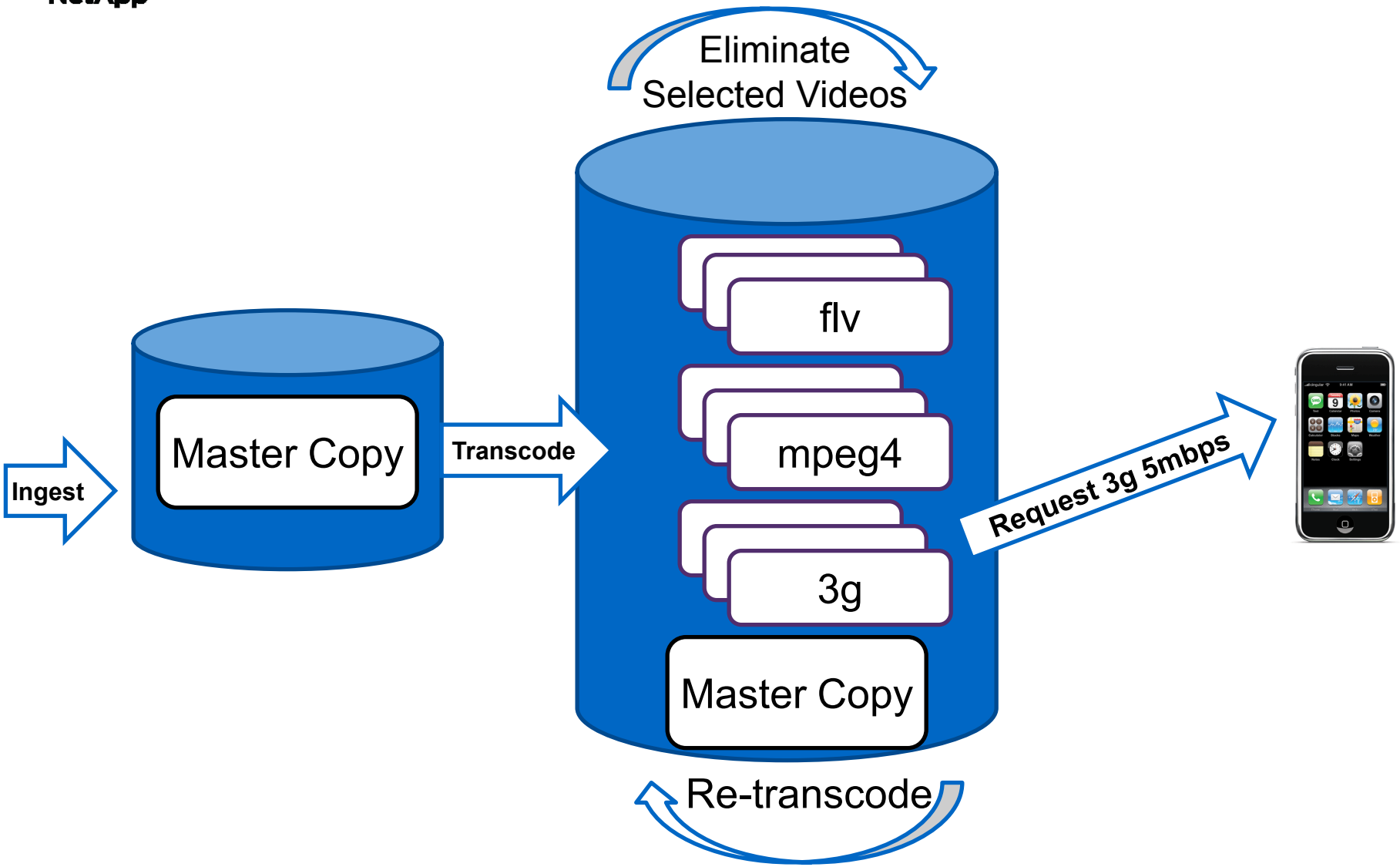


Eliminate unpopular video versions,
transcode on the fly

Optimize on costs

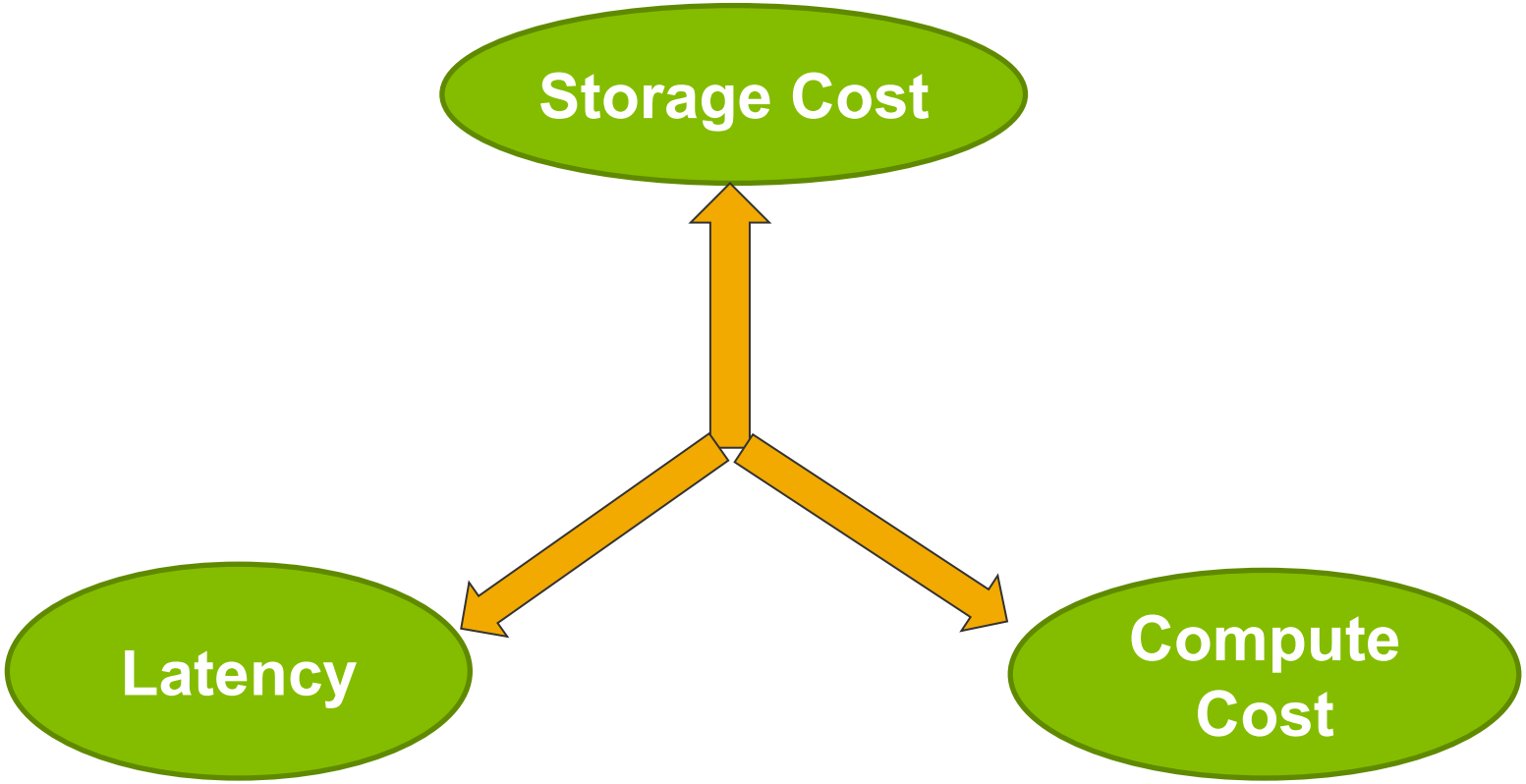
© 2012 NetApp, Inc. All rights reserved.

Proposed video distribution workflow



© 2012 NetApp, Inc. All rights reserved.

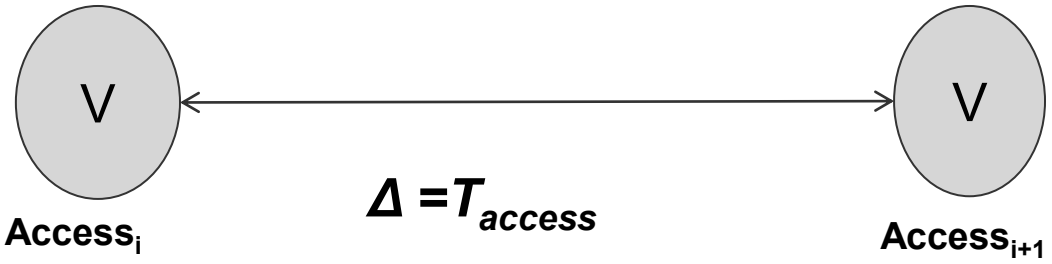
Transcoding-on-fly Trade-offs



© 2012 NetApp, Inc. All rights reserved.

Towards Cost Effective Elimination

- Assume, known T_{access}

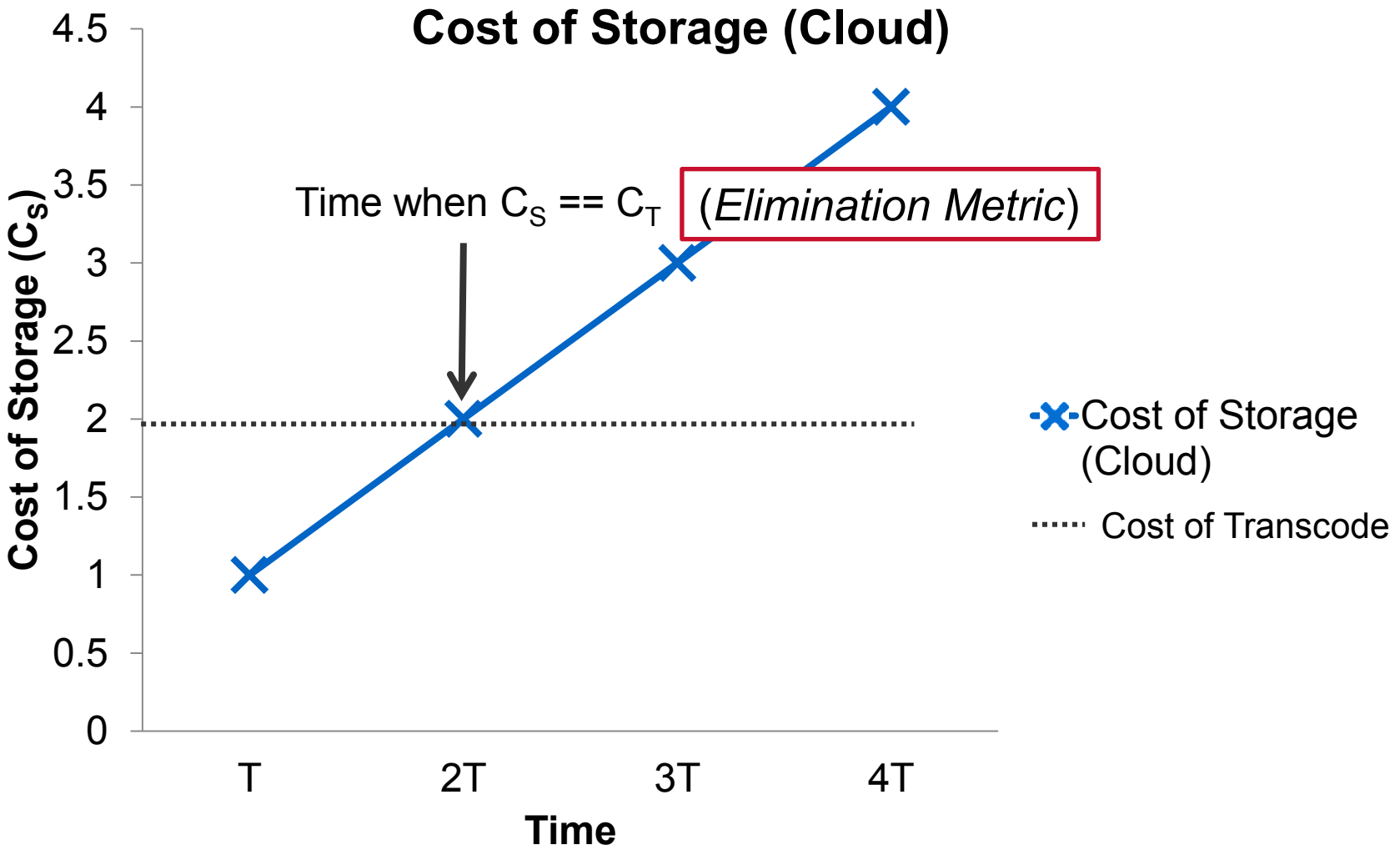


- Cost effective to eliminate if:

Cost of storing for T_{access} time > Cost of transcoding

© 2012 NetApp, Inc. All rights reserved.

Towards Cost Effective Elimination



© 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 (Time) X Storage Cost per unit Time = Transcode Cost

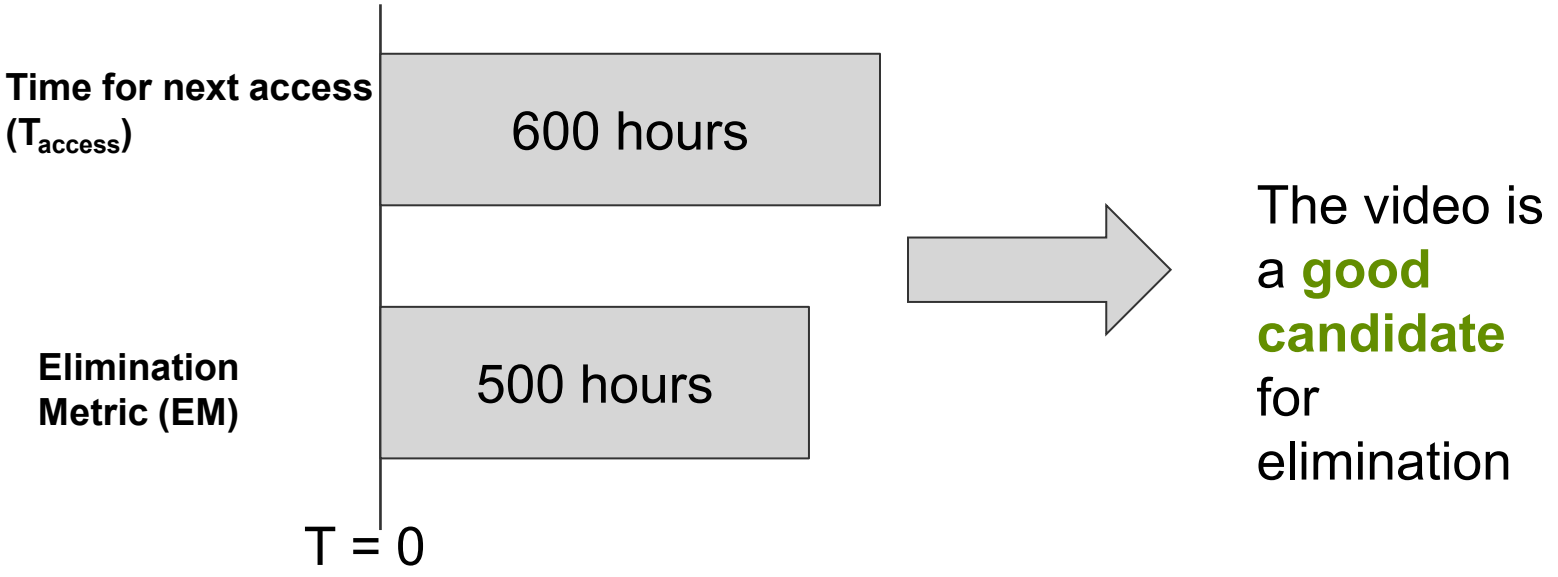
$$\Rightarrow \textbf{EM (Time)} = \frac{\textbf{Transcode Cost}}{\textbf{Storage Cost per unit Time}}$$

*For mathematical derivation, refer paper.

© 2012 NetApp, Inc. All rights reserved.

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



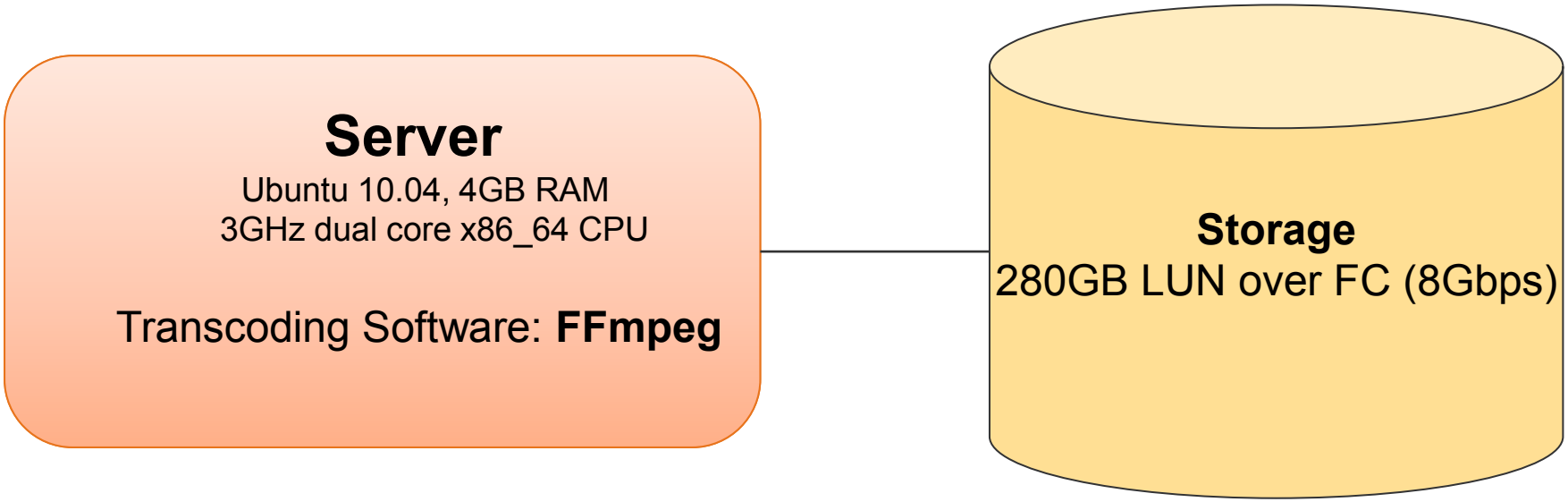
© 2012 NetApp, Inc. All rights reserved.

Analysis



© 2012 NetApp, Inc. All rights reserved.

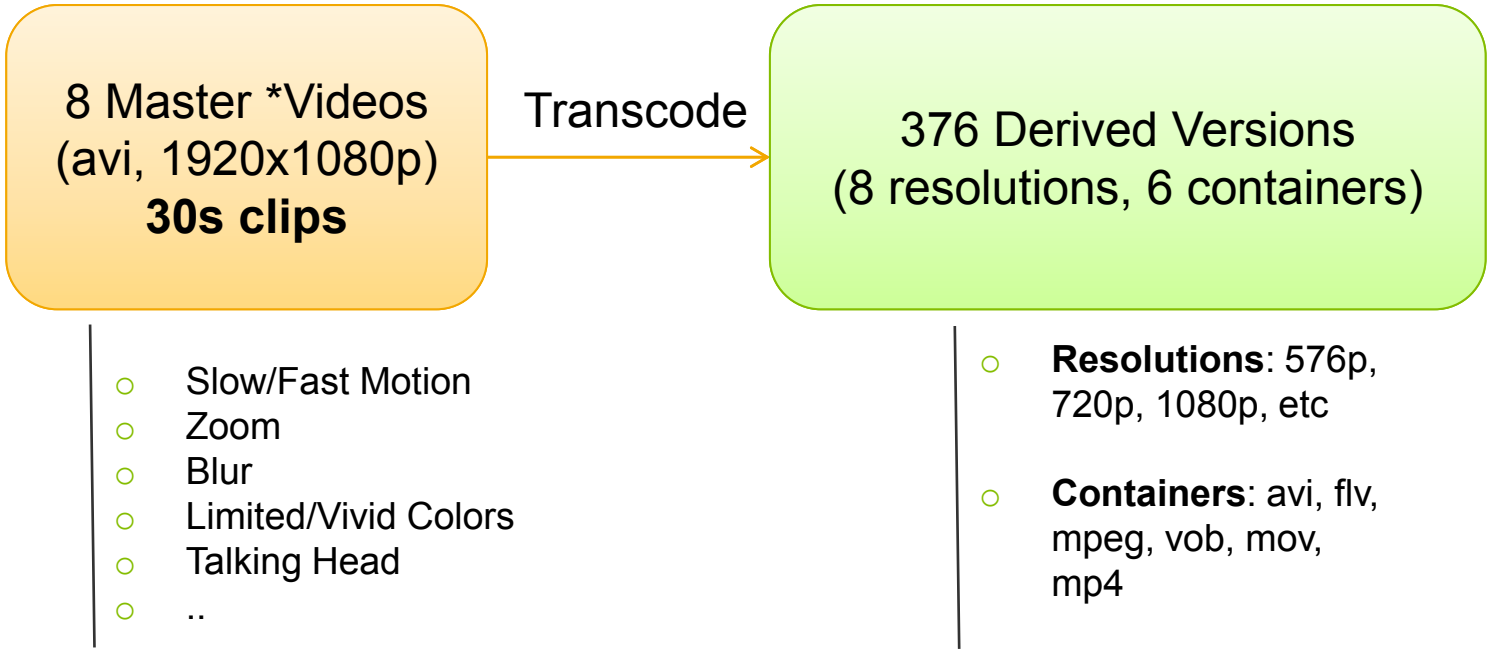
Experimental setup



Cost Model: Amazon EC2 + S3

© 2012 NetApp, Inc. All rights reserved.

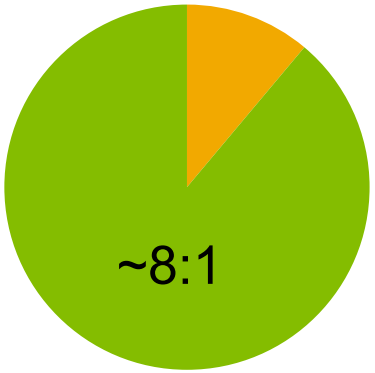
Dataset



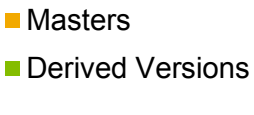
*Videos Source: http://www.testvid.com/

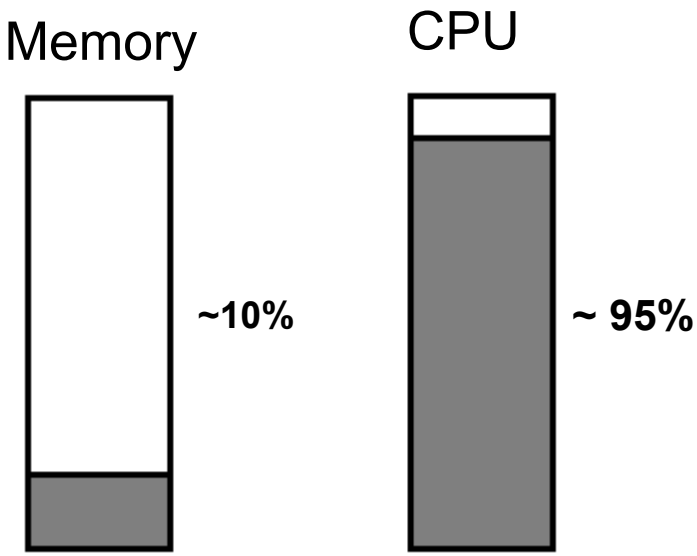
© 2012 NetApp, Inc. All rights reserved.

Observations



Storage footprint

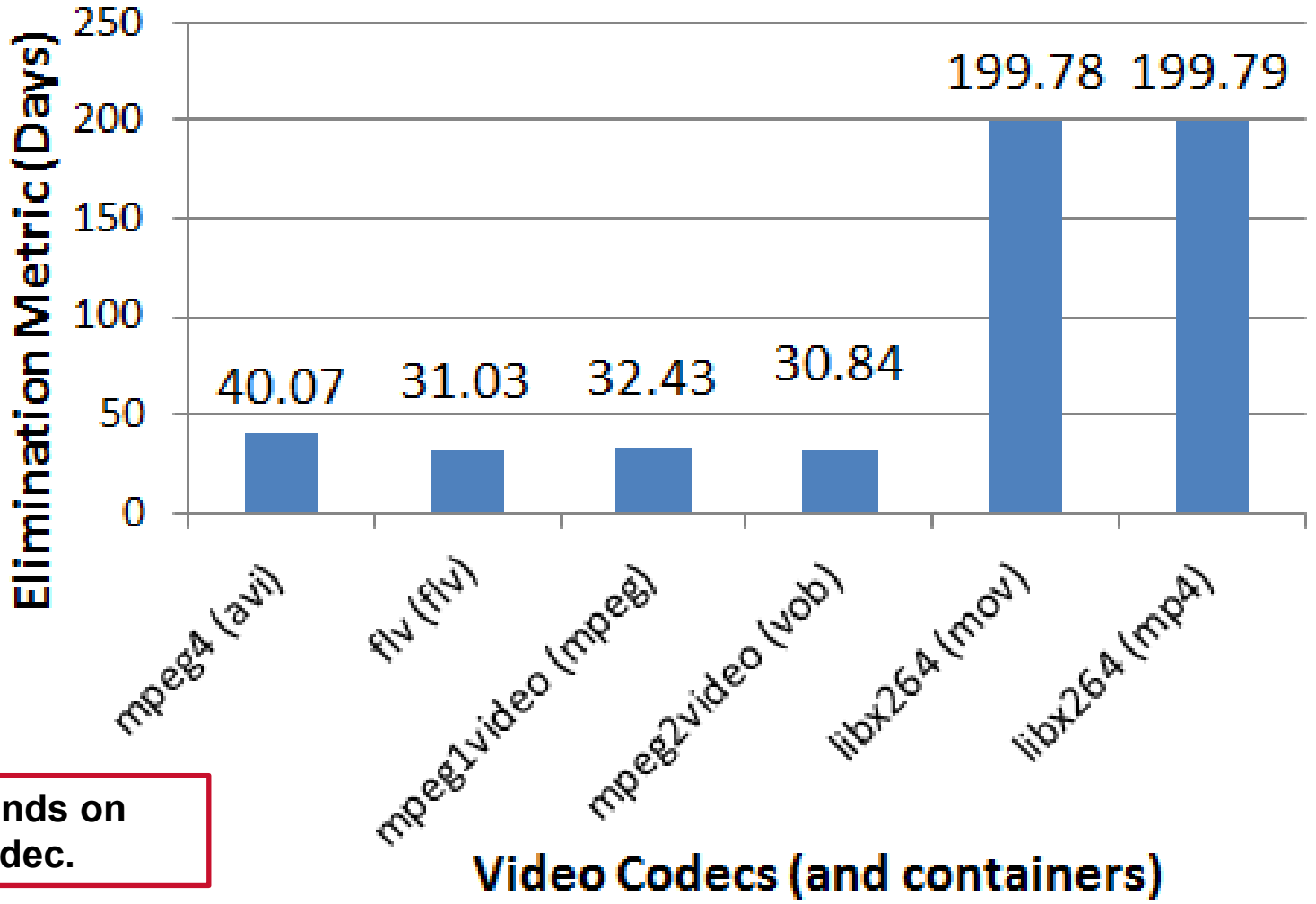




0% block based Deduplication

© 2012 NetApp, Inc. All rights reserved.

EM v/s Video Codecs

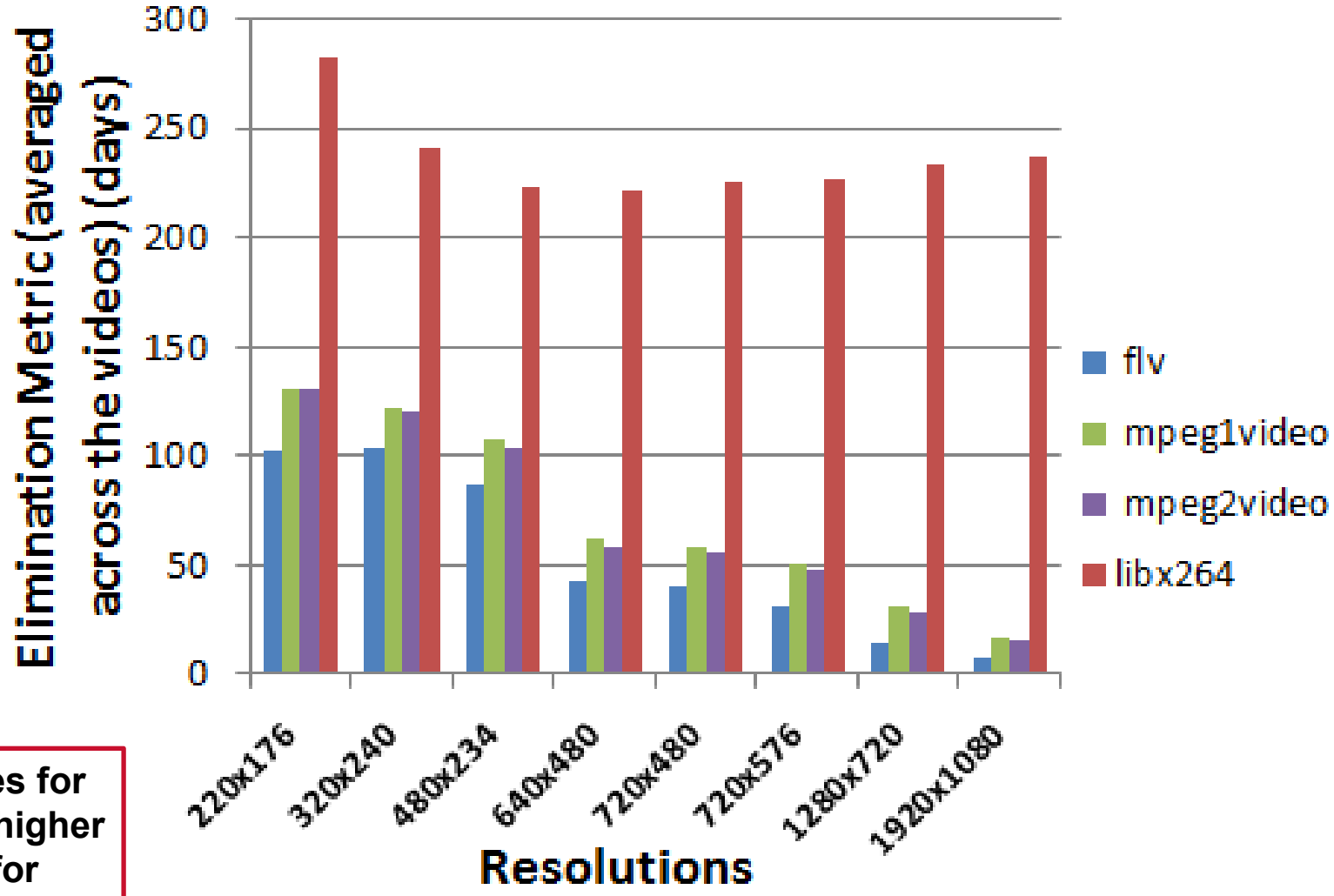


EM depends on video codec.

© 2012 NetApp, Inc. All rights reserved.

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

EM v/s Resolutions

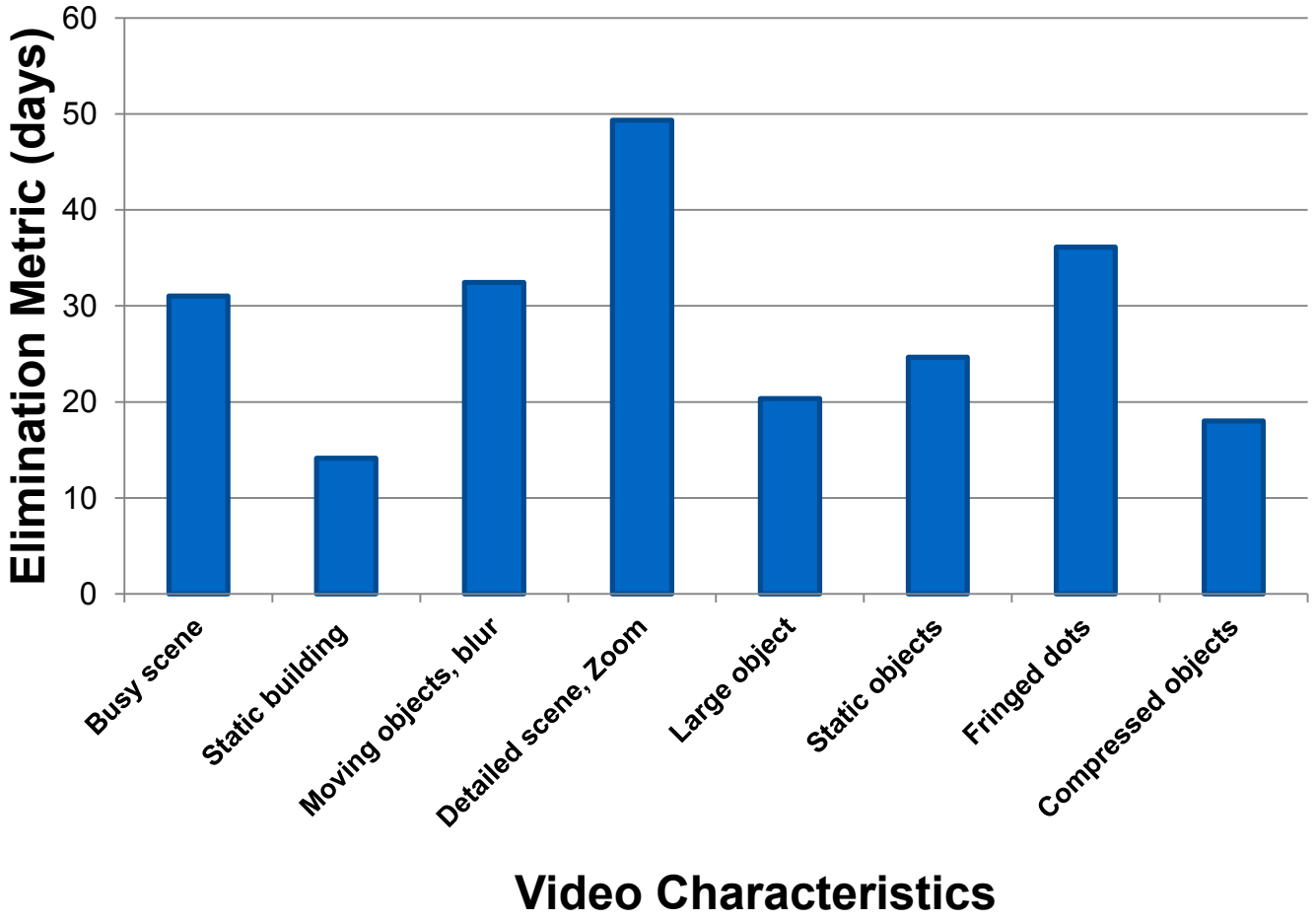


EM decreases for videos with higher resolutions for specific codecs

© 2012 NetApp, Inc. All rights reserved.

* 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

© 2012 NetApp, Inc. All rights reserved.

* 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

© 2012 NetApp, Inc. All rights reserved.

Challenge



Time for video's next access is unknown

© 2012 NetApp, Inc. All rights reserved.

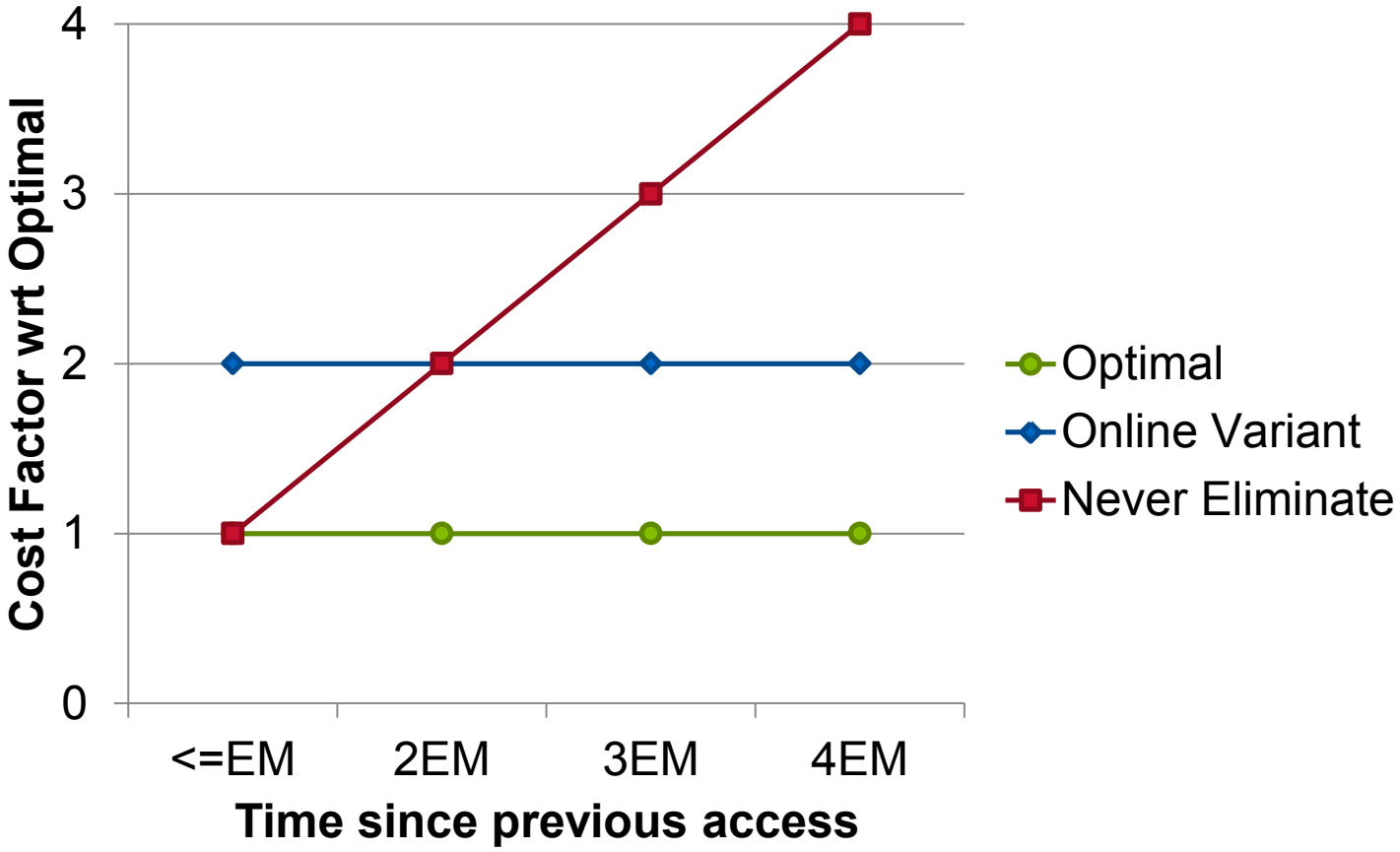
Solution Directions

- 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**

© 2012 NetApp, Inc. All rights reserved.

Elimination Strategies



© 2012 NetApp, Inc. All rights reserved.

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

© 2012 NetApp, Inc. All rights reserved.

References

- Katiyar et al. *ViDeDup: An Application Framework for Video De-duplication*. Usenix HotStorage'11
- Adams et al. "*Maximizing efficiency by trading storage for computation*," HotCloud'09
- Zhou et al. *A Video Replacement Policy based on Reve-nue to Cost Ratio in a Multicast TV-Anytime System*. Parallel and Distributed Processing Symposium 2011
- 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
- S. Albers. *Online algorithms. In Interactive Computa-tion:* The New Paradigm edited by D.Q. Goldin, S.A. Smolka and P. Wegner, 143-164, 2006
- Content-Aware Storage Optimization http://www.ocarinanetworks.com/resources/147

© 2012 NetApp, Inc. All rights reserved.

NetApp

Q&A



© 2012 NetApp, Inc. All rights reserved.