

Storage Efficiency Opportunities and Analysis for Video Repositories

HotStorage ‘15

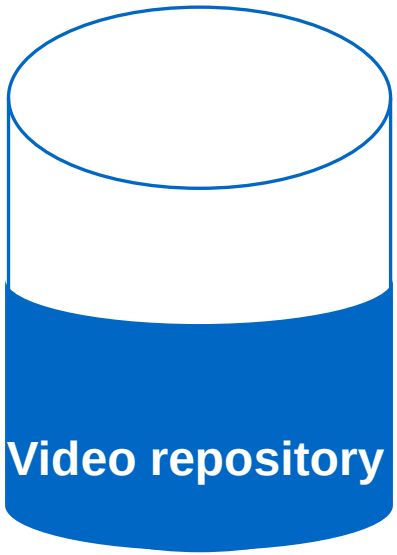
Suganthi Dewakar, Sethuraman Subbiah, Gokul Soundararajan, Mike Wilson, Mark W. Storer, Kishore Udayashankar, Kaladhar Voruganti, Minglong Shao

NetApp Inc.

07/07/2015

© 2015 NetApp, Inc. All rights reserved.

Video files have deduplication



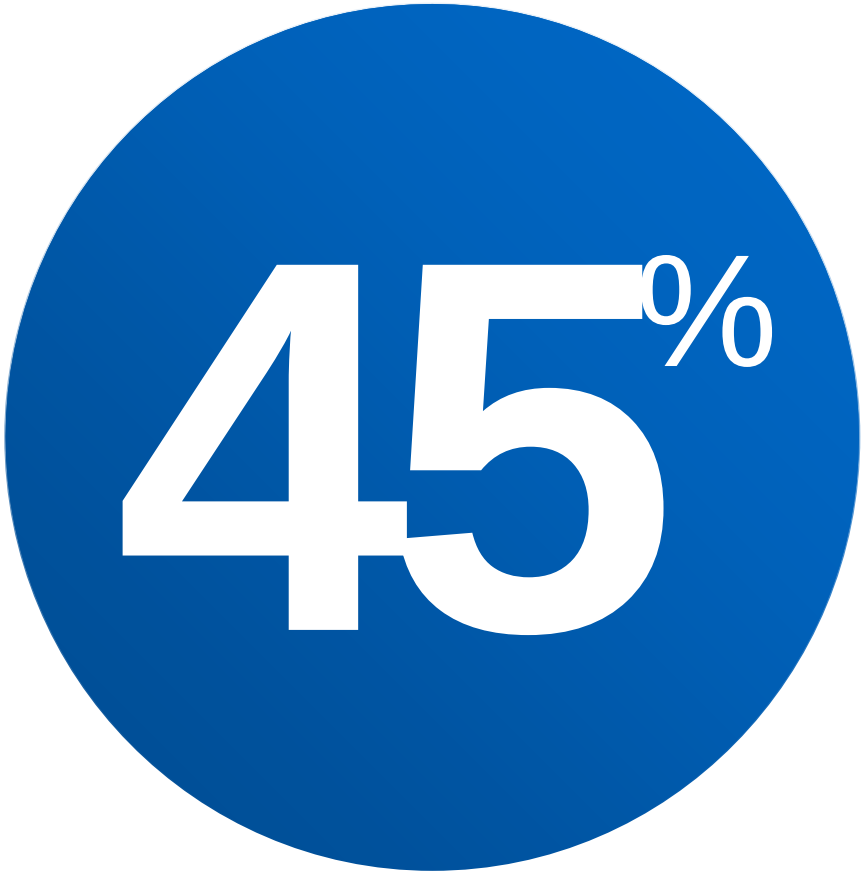


No deduplication!

© 2015 NetApp, Inc. All rights reserved.

Video files have deduplication





© 2015 NetApp, Inc. All rights reserved.

But where..?



© 2015 NetApp, Inc. All rights reserved.

But where..?



Video Production

© 2015 NetApp, Inc. All rights reserved.

But where..?



Video Production



Video Distribution

© 2015 NetApp, Inc. All rights reserved.

But where..?



Video Production



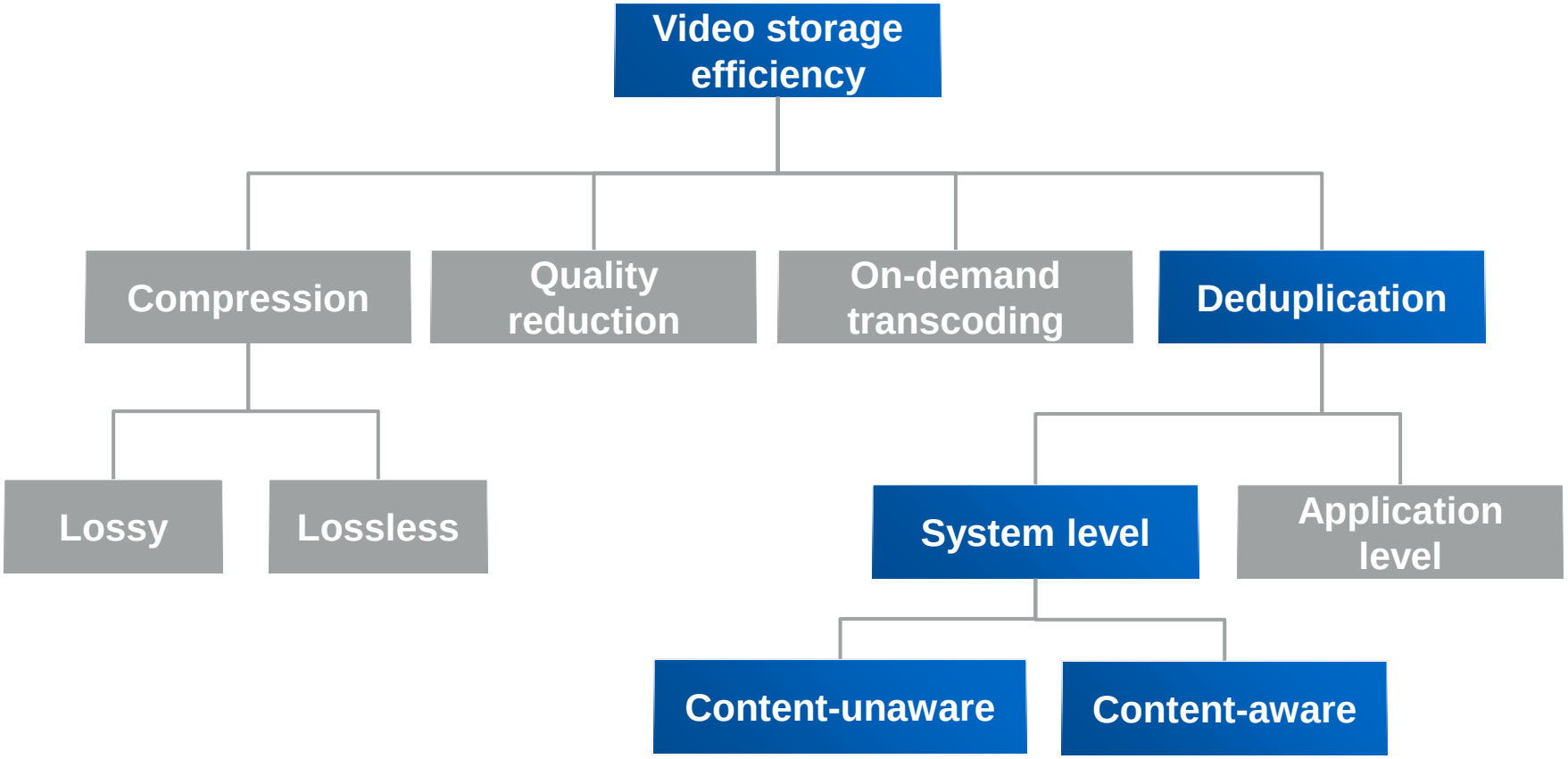
Video Distribution



DVR

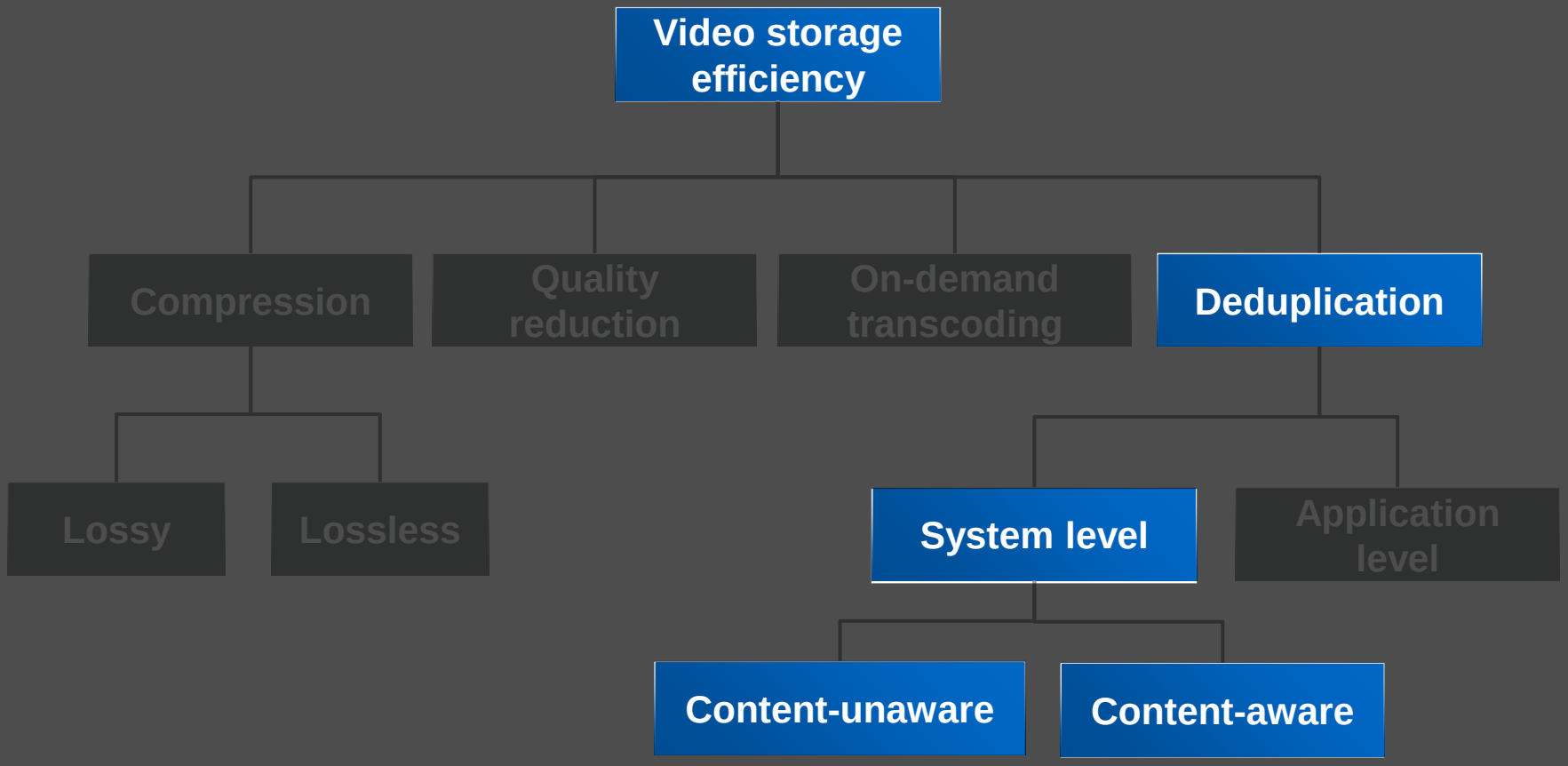
© 2015 NetApp, Inc. All rights reserved.

Storage efficiency options



© 2015 NetApp, Inc. All rights reserved.

Storage efficiency options



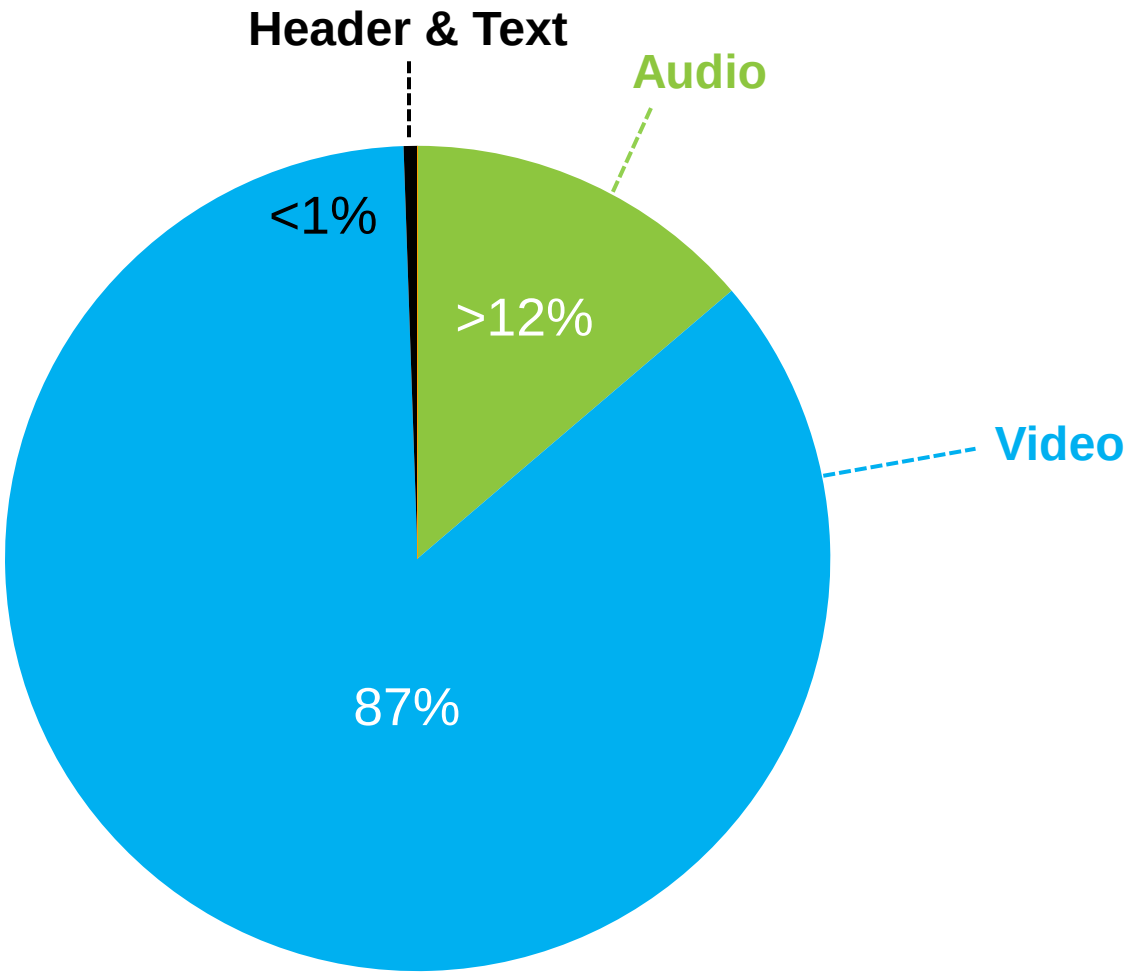
© 2015 NetApp, Inc. All rights reserved.

Outline

1) Introduction
2) Classification of storage efficiency techniques
3) Content-aware deduplication techniques
4) Evaluation
5) Conclusions and future work

© 2015 NetApp, Inc. All rights reserved.

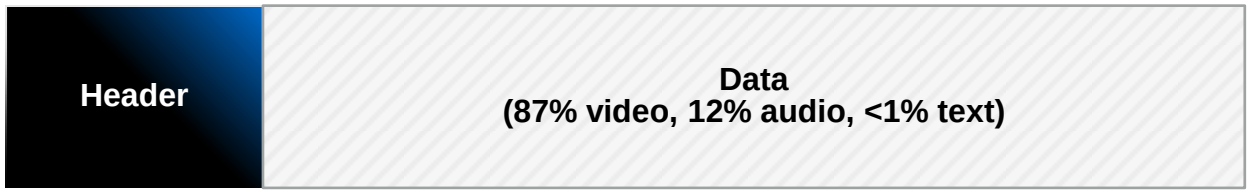
Video file content



© 2015 NetApp, Inc. All rights reserved.
NetApp

Video file layout

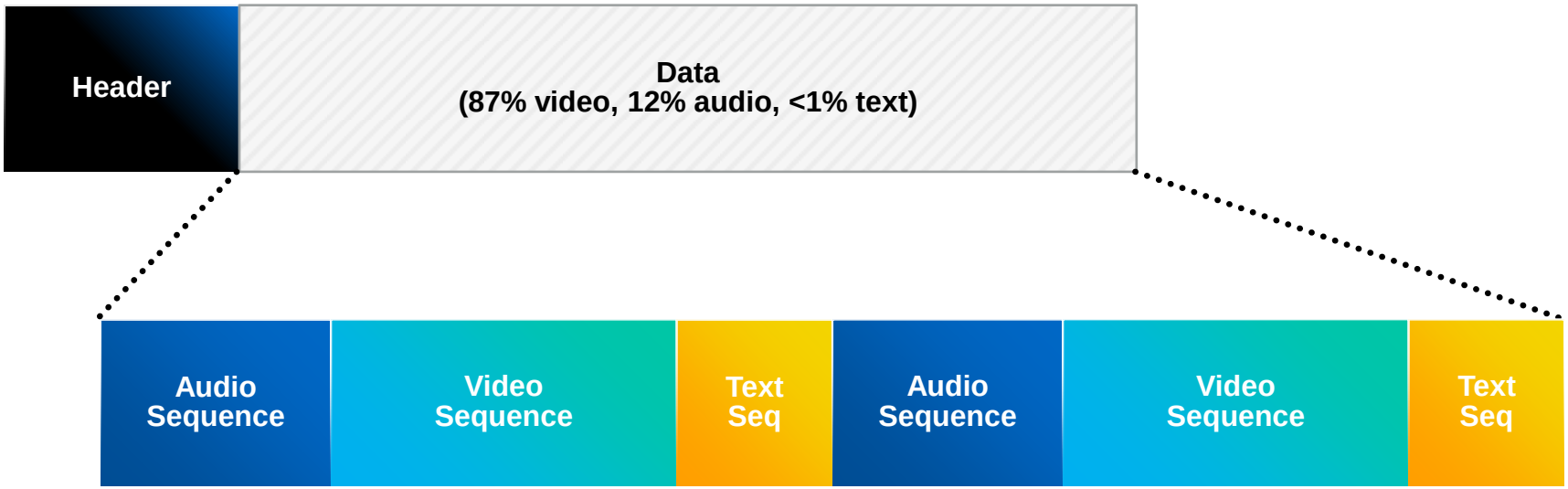
ISO Base Media Format



© 2015 NetApp, Inc. All rights reserved.

Video file layout

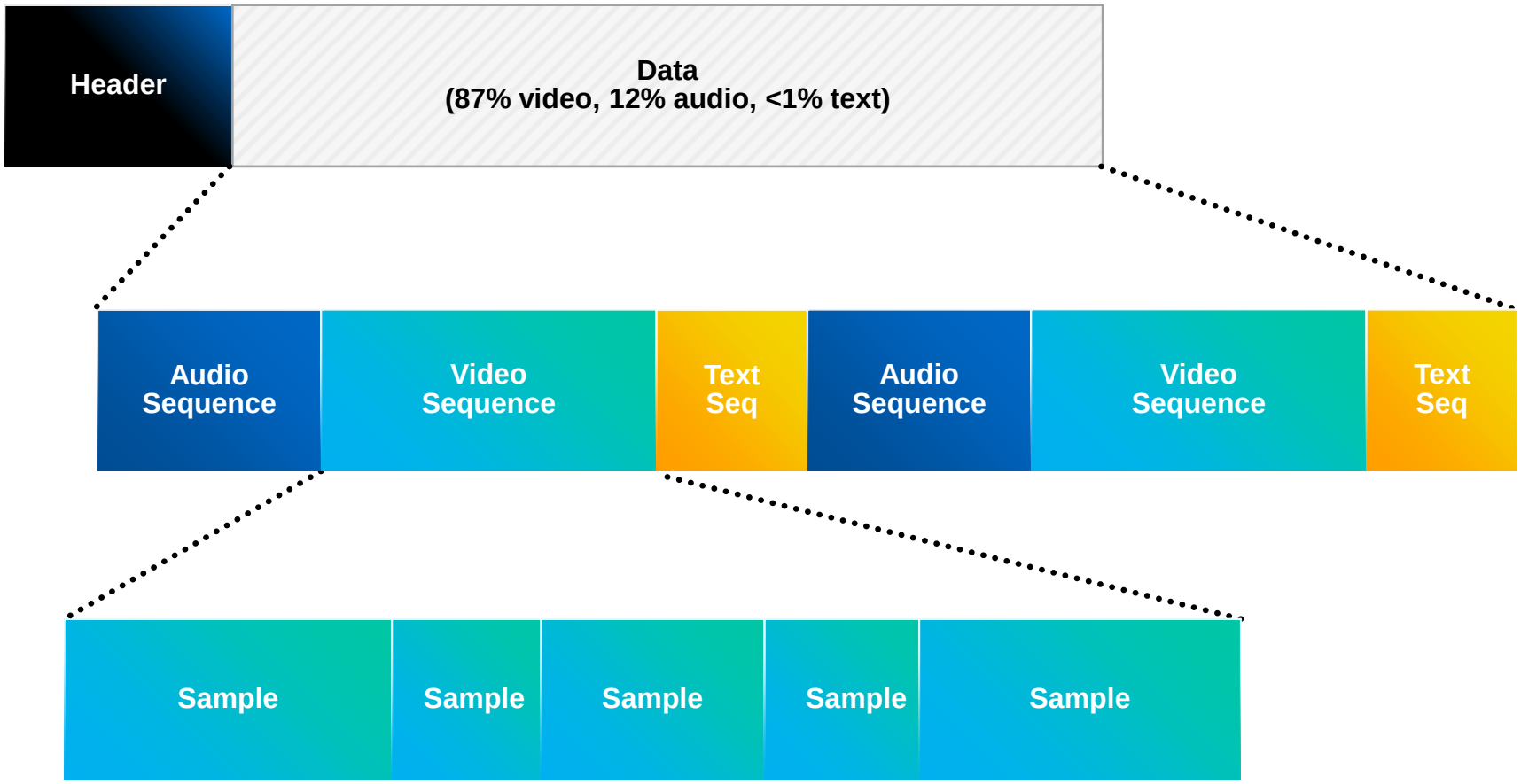
ISO Base Media Format



© 2015 NetApp, Inc. All rights reserved.

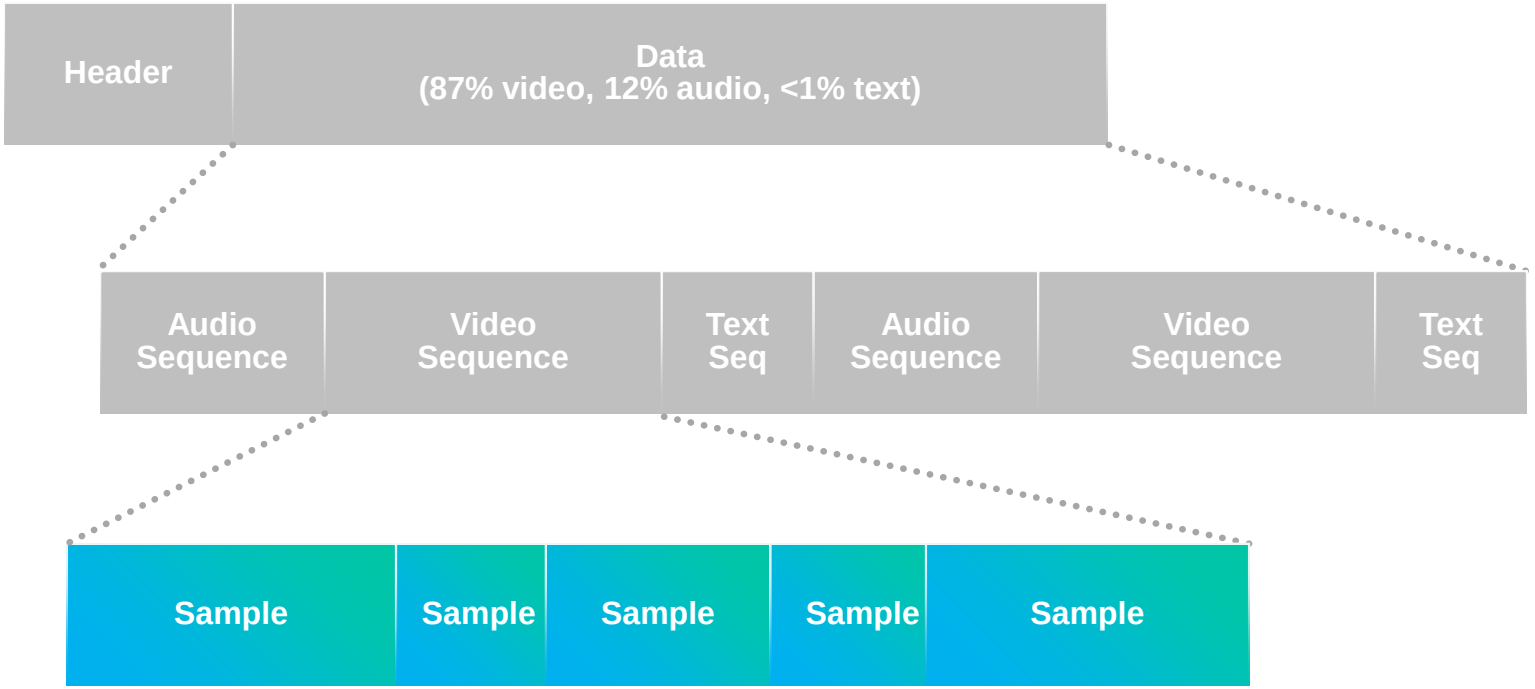
Video file layout

ISO Base Media Format



© 2015 NetApp, Inc. All rights reserved.
NetApp

Sample-based deduplication



© 2015 NetApp, Inc. All rights reserved.
NetApp

Sample-based deduplication technique

Signature table

Finger print	Link to data	Length
V1	40024	894
V2	54204	984
V3	62683	683

© 2015 NetApp, Inc. All rights reserved.

Sample-based deduplication technique

Signature table

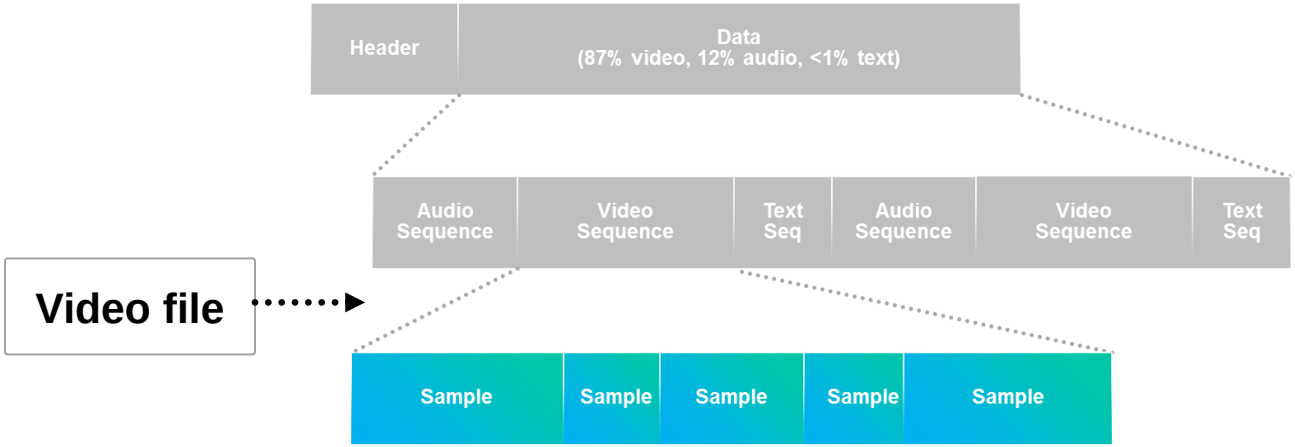
Finger print	Link to data	Length
V1	40024	894
V2	54204	984
V3	62683	683

Video file

Video File

© 2015 NetApp, Inc. All rights reserved.

Sample-based deduplication technique



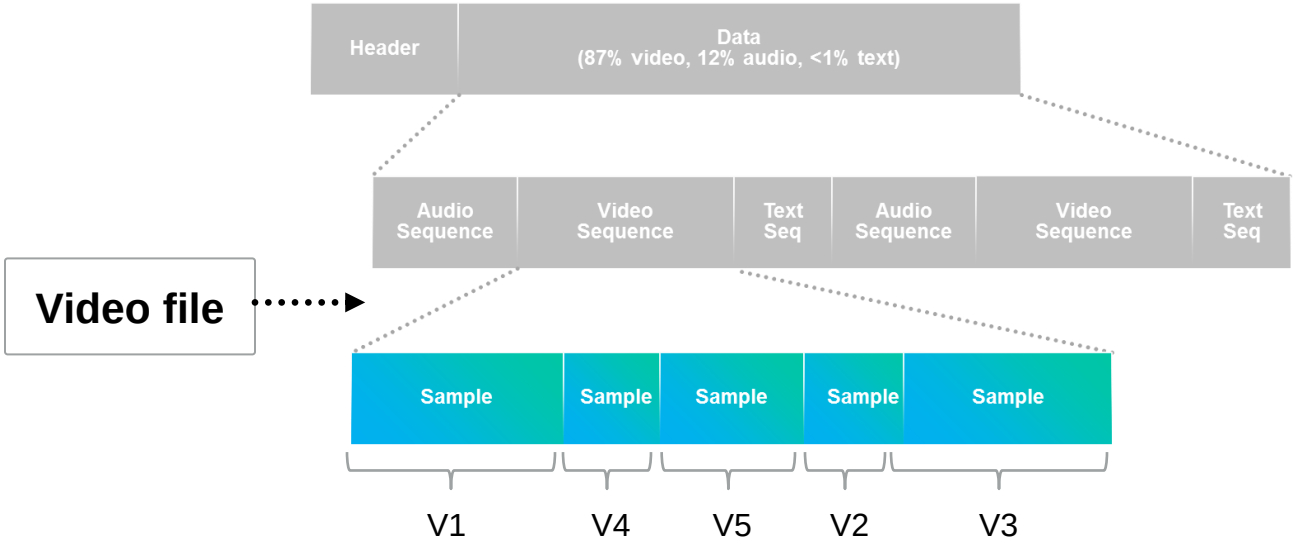
Signature table

Finger print	Link to data	Length
V1	40024	894
V2	54204	984
V3	62683	683



© 2015 NetApp, Inc. All rights reserved.

Sample-based deduplication technique



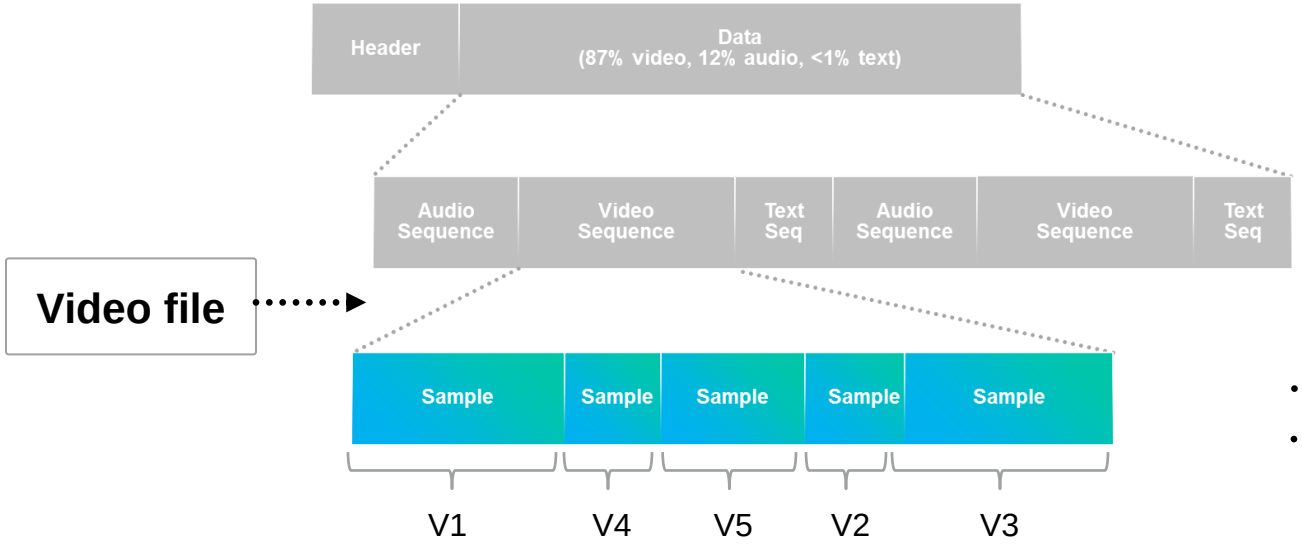
Signature table

Finger print	Link to data	Length
V1	40024	894
V2	54204	984
V3	62683	683



© 2015 NetApp, Inc. All rights reserved.
NetApp

Sample-based deduplication technique



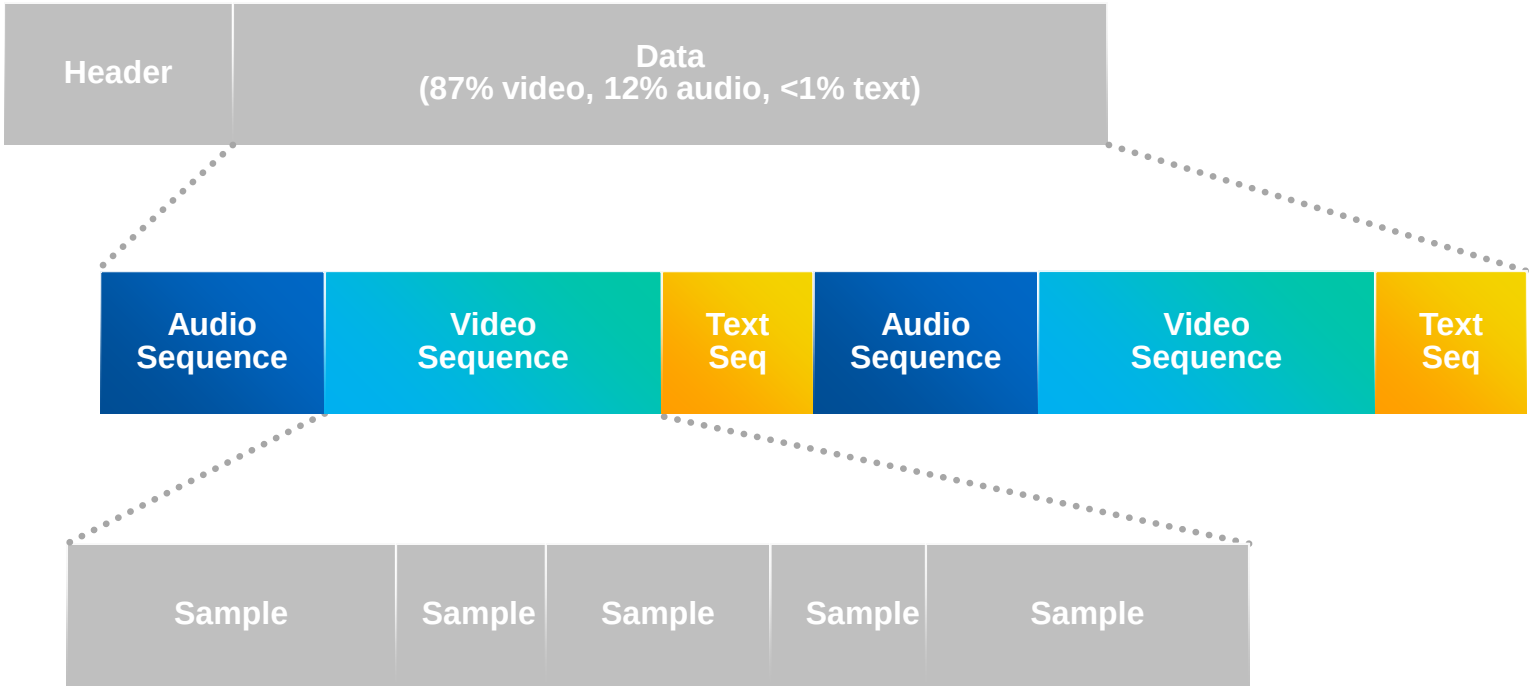
Signature table

Finger print	Link to data	Length
V1	40024	894
V2	54204	984
V3	62683	683
V4	74000	937
V5	80137	873



© 2015 NetApp, Inc. All rights reserved.
NetApp

Sequence-based deduplication



© 2015 NetApp, Inc. All rights reserved.
NetApp

Sequence-based deduplication technique

Audio signature table

Signature	Link to data	Length
A1	40024	894
A2	62683	683

Video signature table

Signature	Link to data	Length
V1	30995	312
V2	12685	646
V3	19865	857

© 2015 NetApp, Inc. All rights reserved.

Sequence-based deduplication technique

Audio signature table

Signature	Link to data	Length
A1	40024	894
A2	62683	683

Video file

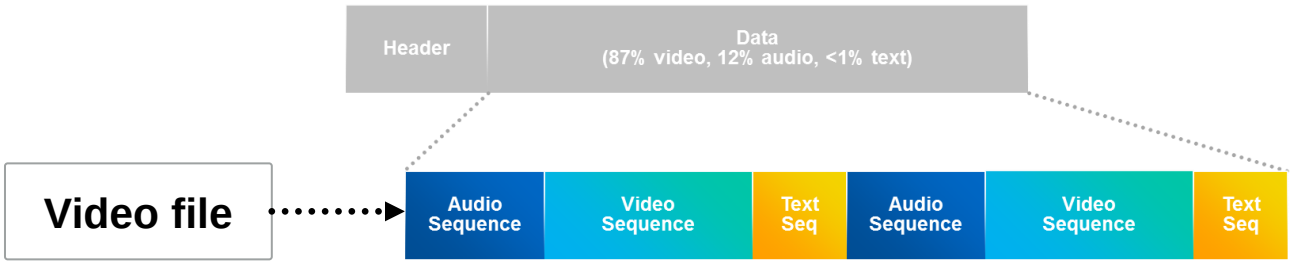
Video signature table

Signature	Link to data	Length
V1	30995	312
V2	12685	646
V3	19865	857

Video File

© 2015 NetApp, Inc. All rights reserved.

Sequence-based deduplication technique



Audio signature table

Signature	Link to data	Length
A1	40024	894
A2	62683	683

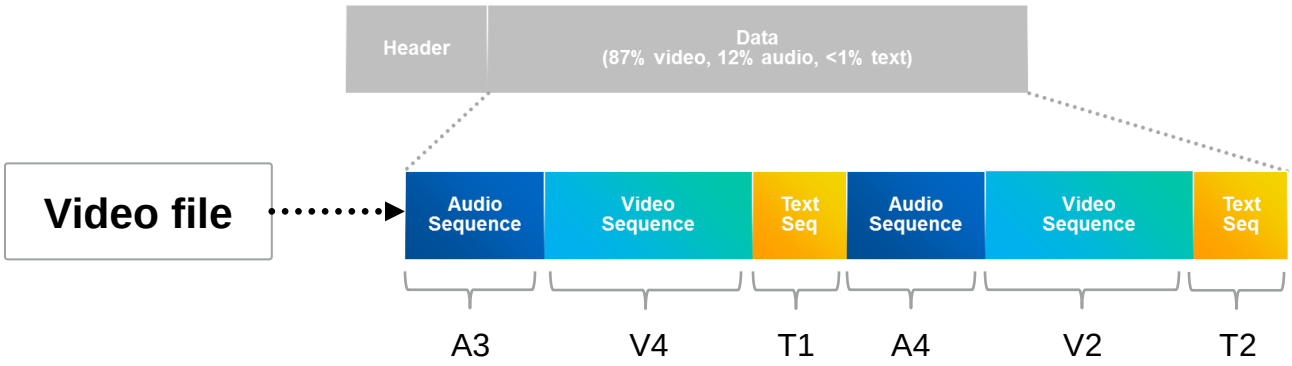
Video signature table

Signature	Link to data	Length
V1	30995	312
V2	12685	646
V3	19865	857

Video File → **Extract sequences**

© 2015 NetApp, Inc. All rights reserved.
NetApp

Sequence-based deduplication technique



Audio signature table

Signature	Link to data	Length
A1	40024	894
A2	62683	683

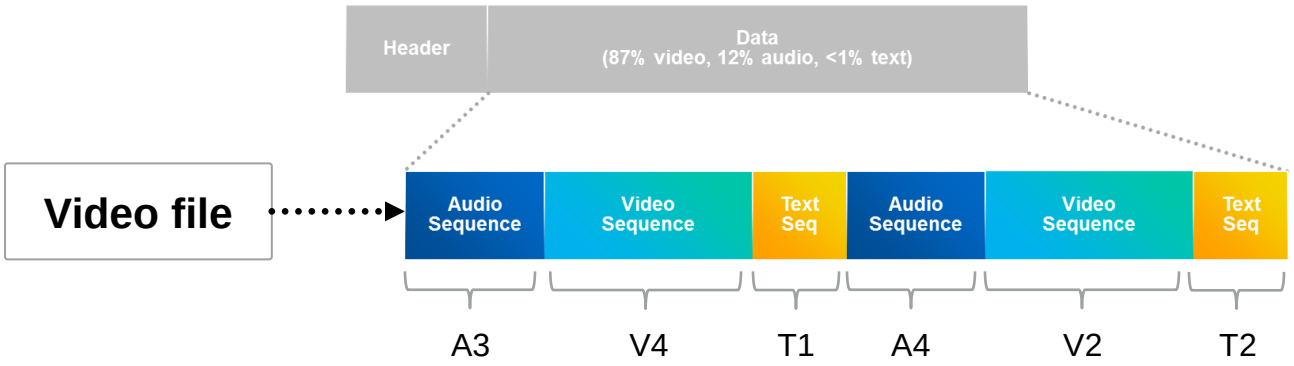
Video signature table

Signature	Link to data	Length
V1	30995	312
V2	12685	646
V3	19865	857



© 2015 NetApp, Inc. All rights reserved.
NetApp

Sequence-based deduplication technique



Audio signature table

Signature	Link to data	Length
A1	40024	894
A2	62683	683
A3	74000	937
A4	80137	873

Video signature table

Signature	Link to data	Length
V1	30995	312
V2	12685	646
V3	19865	857
V4	50687	757



© 2015 NetApp, Inc. All rights reserved.
NetApp

Outline

1) Introduction
2) Classification of storage efficiency techniques
3) Content-aware deduplication techniques
4) Evaluation
   - Experiments
   - Input dataset
   - Metrics
   - Results
5) Conclusions and future work

© 2015 NetApp, Inc. All rights reserved.

Existing deduplication techniques

- Content-unaware deduplication techniques
  - Fixed-sized deduplication
  - Variable-sized deduplication
- Fixed-sized
  - Block length – 4096 Bytes
- Variable-sized
  - Rabin fingerprint
  - Average block size: 4096 Bytes
  - Polynomial degree: 48

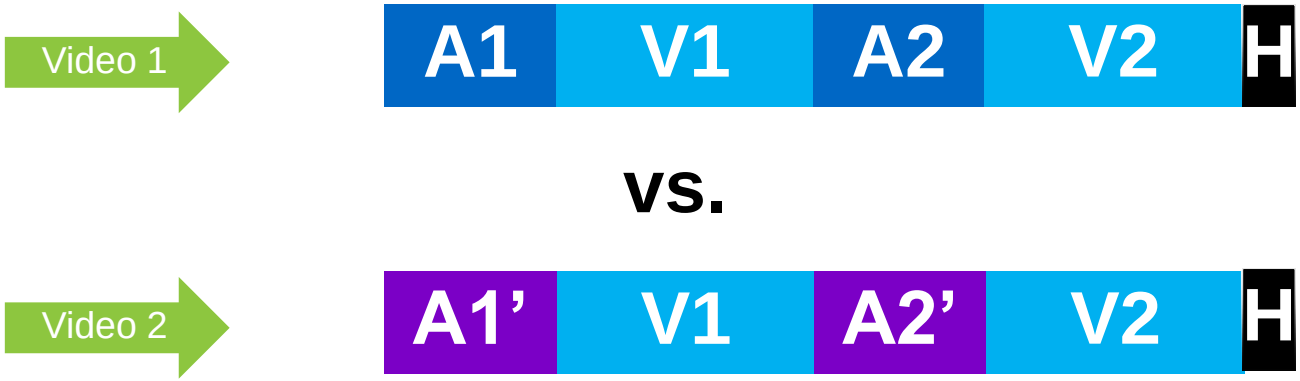
© 2015 NetApp, Inc. All rights reserved.

Experiments



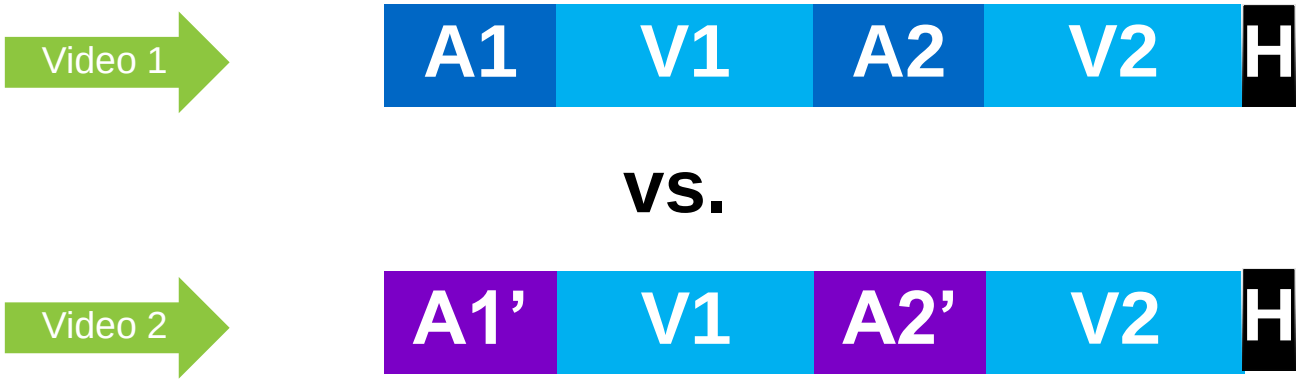
© 2015 NetApp, Inc. All rights reserved.

Experiments



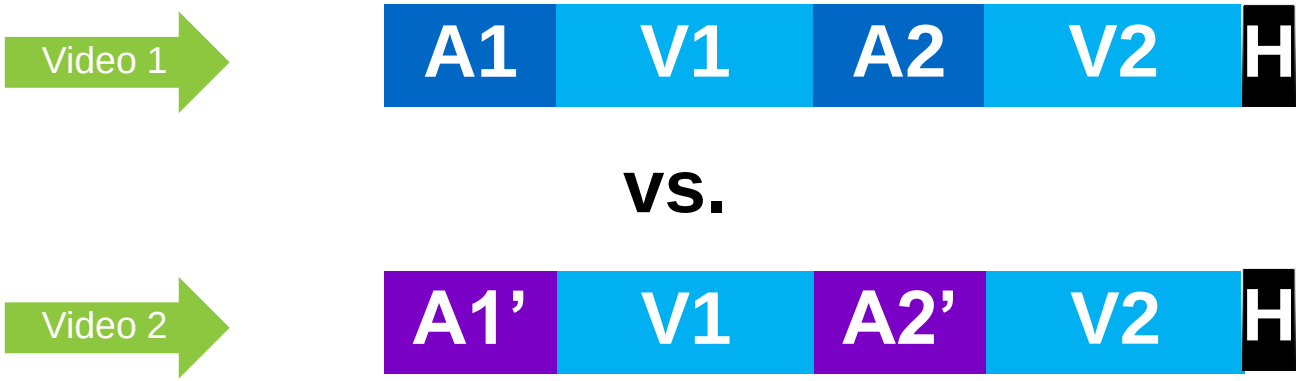
© 2015 NetApp, Inc. All rights reserved.

Experiments



© 2015 NetApp, Inc. All rights reserved.

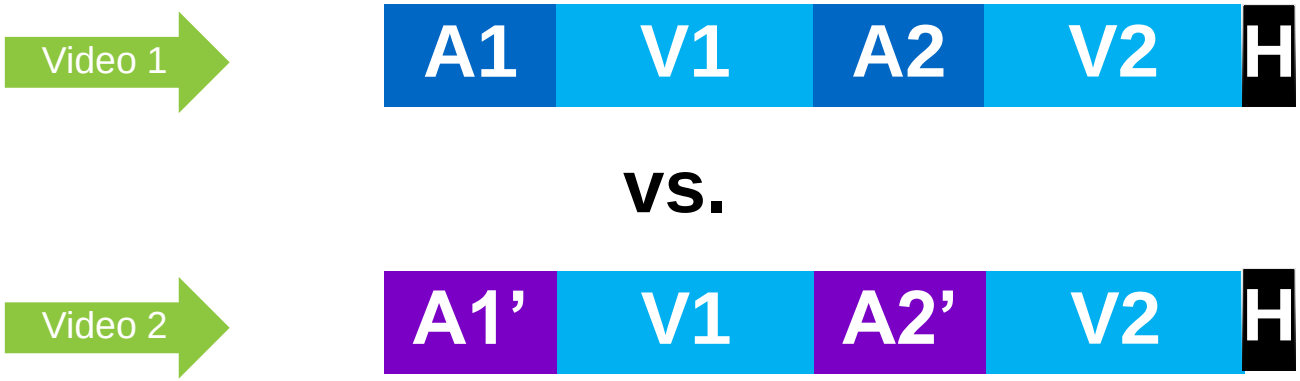
Experiments



After deduplication:

© 2015 NetApp, Inc. All rights reserved.

Experiments

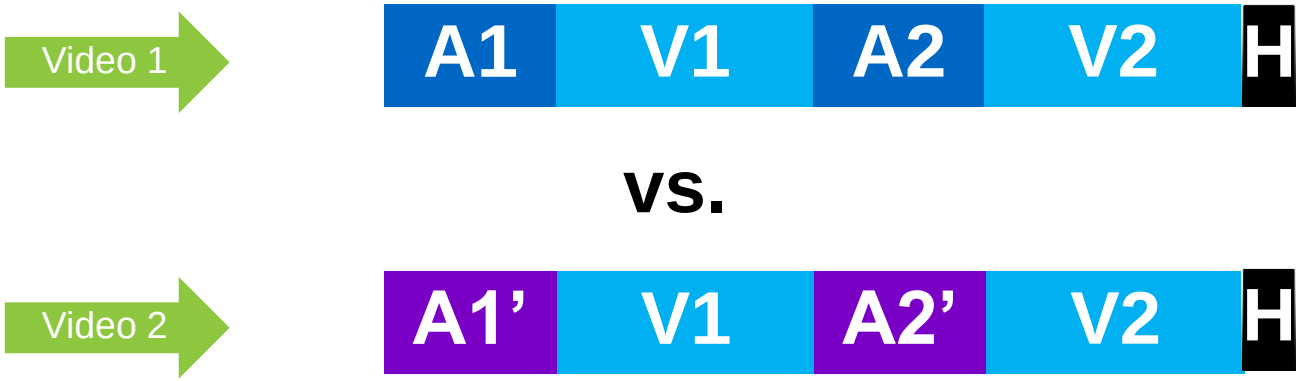


After deduplication:

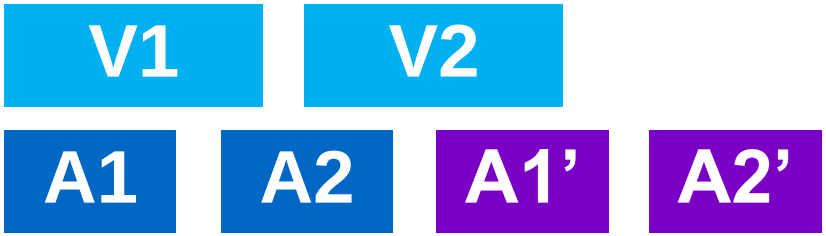


© 2015 NetApp, Inc. All rights reserved.

Experiments



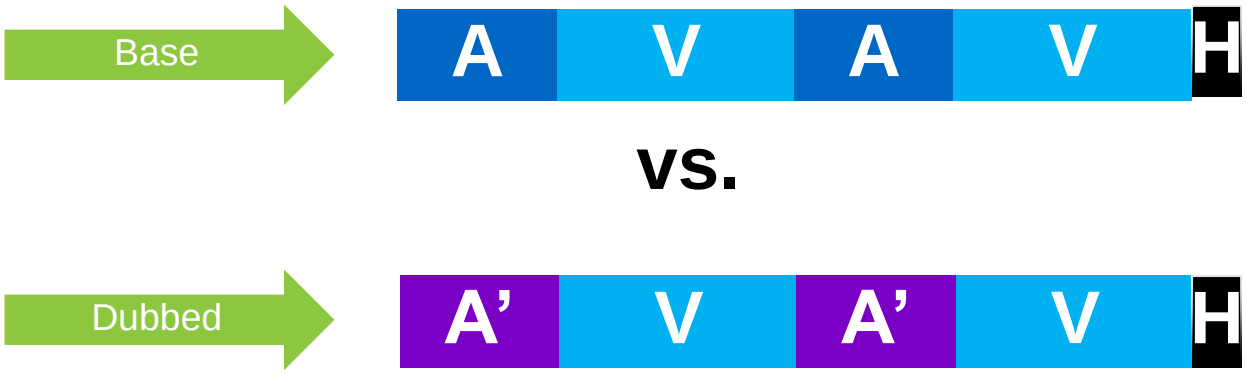
After deduplication:



© 2015 NetApp, Inc. All rights reserved.

Experiments

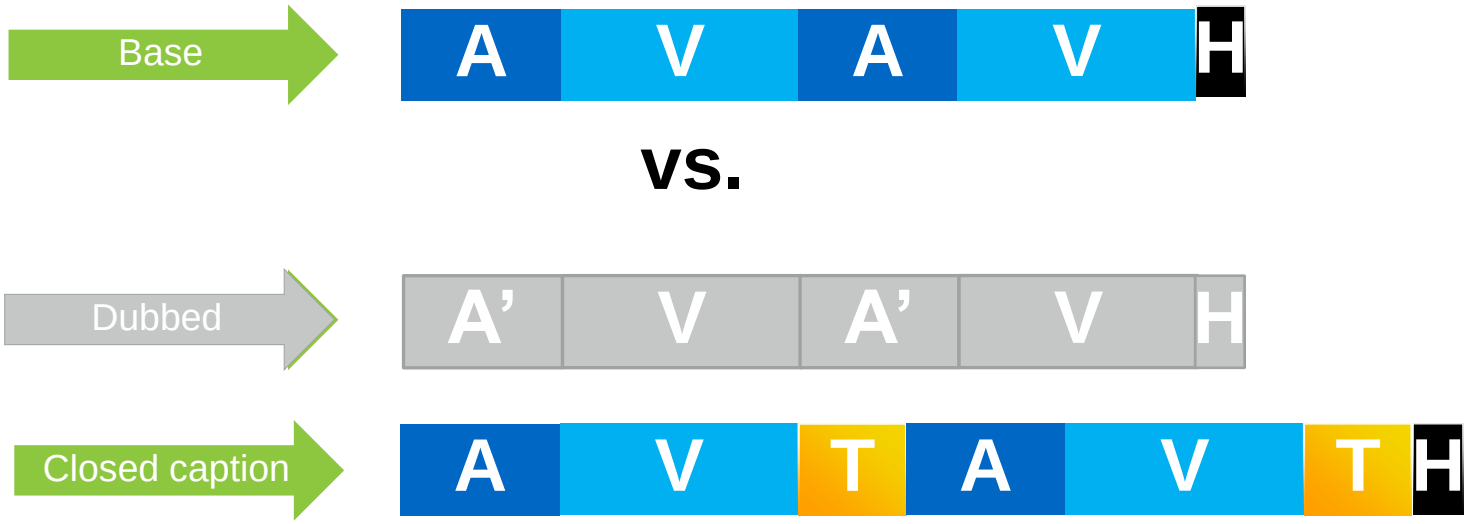
Content variations



© 2015 NetApp, Inc. All rights reserved.

Experiments

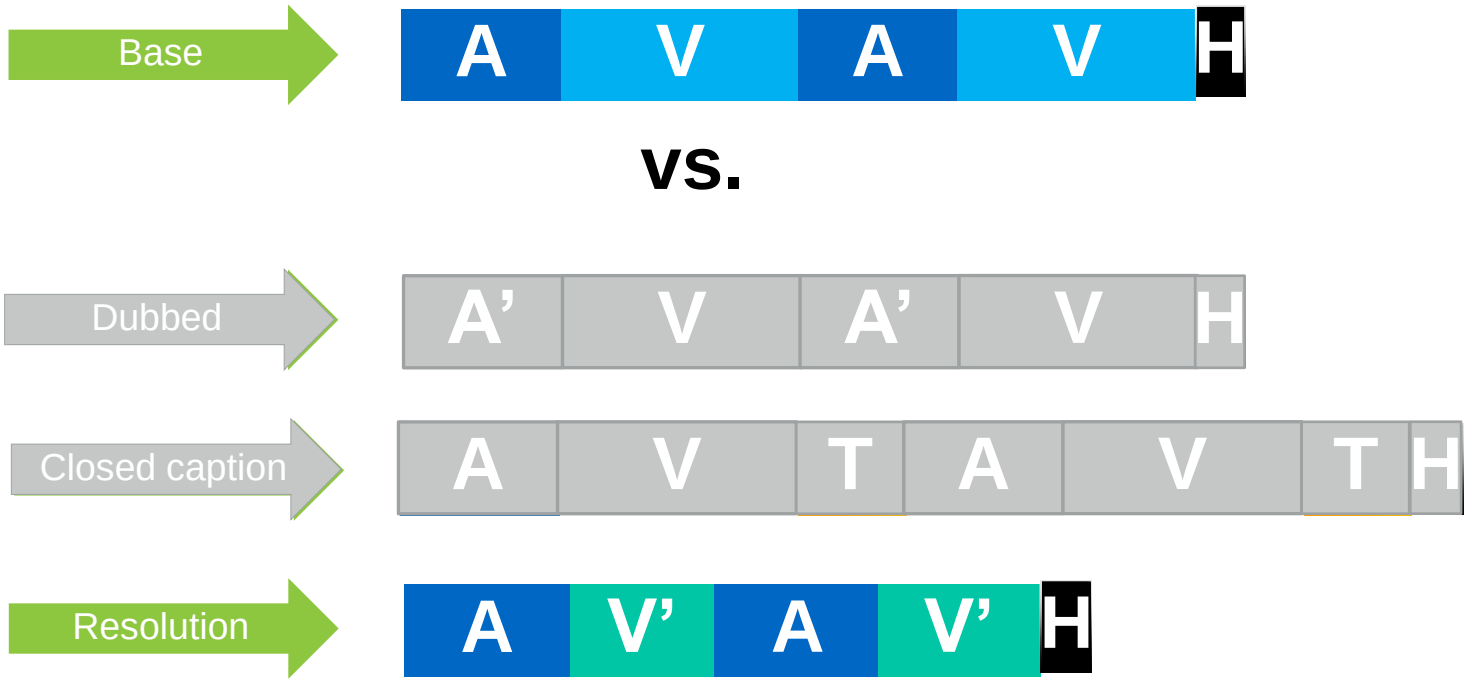
Content variations



© 2015 NetApp, Inc. All rights reserved.
NetApp

Experiments

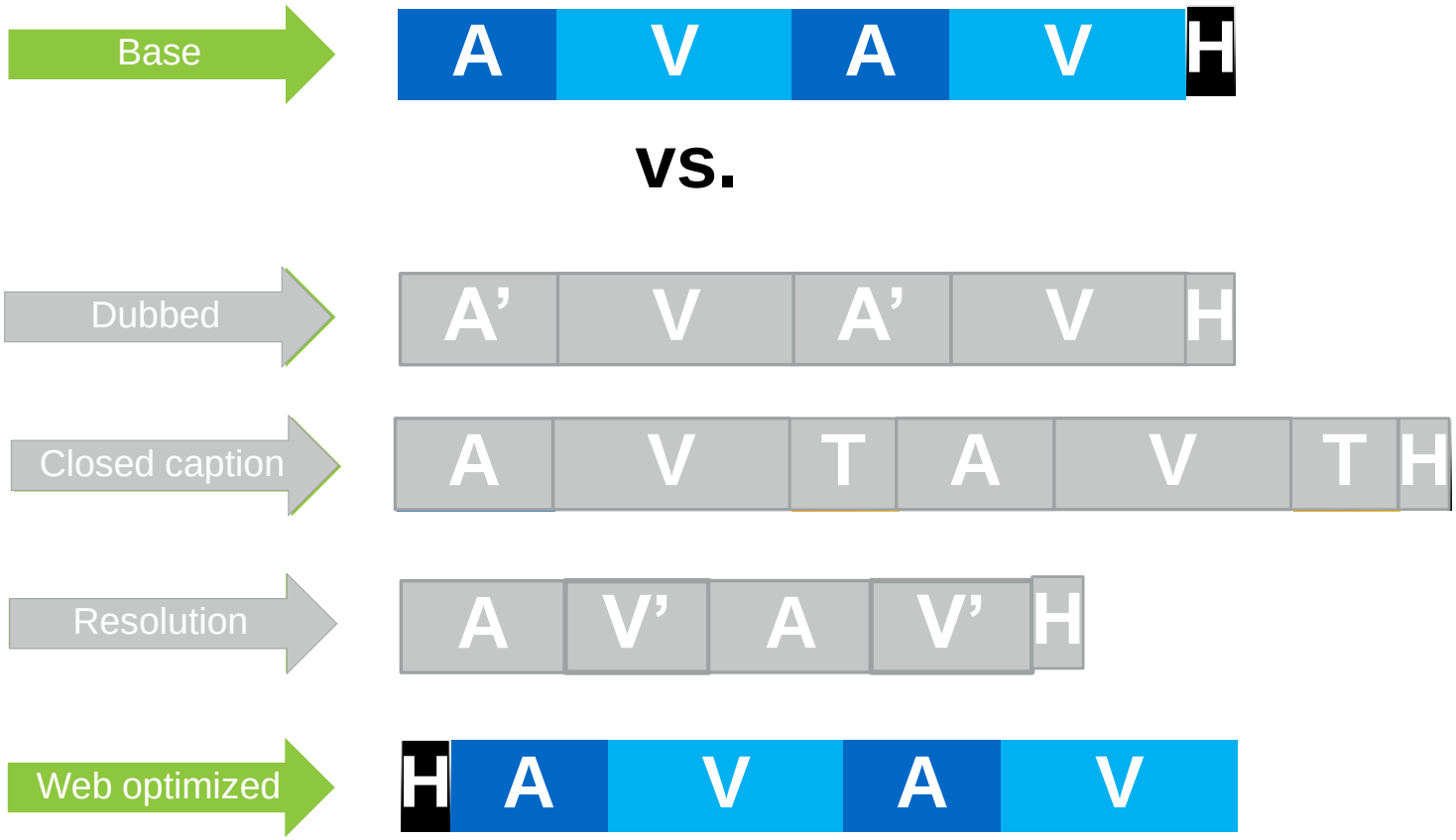
Content variations



© 2015 NetApp, Inc. All rights reserved.
NetApp

Experiments

Content variations



© 2015 NetApp, Inc. All rights reserved.
NetApp

Experiments

Video Segments

Base



vs.

Source: Madagascar

© 2015 NetApp, Inc. All rights reserved.

Experiments

Video Segments

Base



vs.

Different start time



Source: Madagascar

© 2015 NetApp, Inc. All rights reserved.

Experiments

Video Segments

Base



vs.

Different start time



Different end time



Source: Madagascar

© 2015 NetApp, Inc. All rights reserved.

Experiments

Video Segments

Base



vs.

Different start time



Different end time



Subset of base



Source: Madagascar

© 2015 NetApp, Inc. All rights reserved.

Input dataset

Video DVDs

- Movies and TV series
- Base, Dubbed, Closed Caption, Resolution
- Metadata in the beginning (Web Optimized)
- Different start time, different end time
- Subset of a video



© 2015 NetApp, Inc. All rights reserved.
NetApp

Evaluation metrics

- **Storage saved**

$$\text{Deduplication ratio \%} = 100 - \left(\frac{\text{storage footprint after dedupe}}{\sum \text{size of files}} * 100 \right)$$

E.g. If a file is stored twice, storage saved% is 50

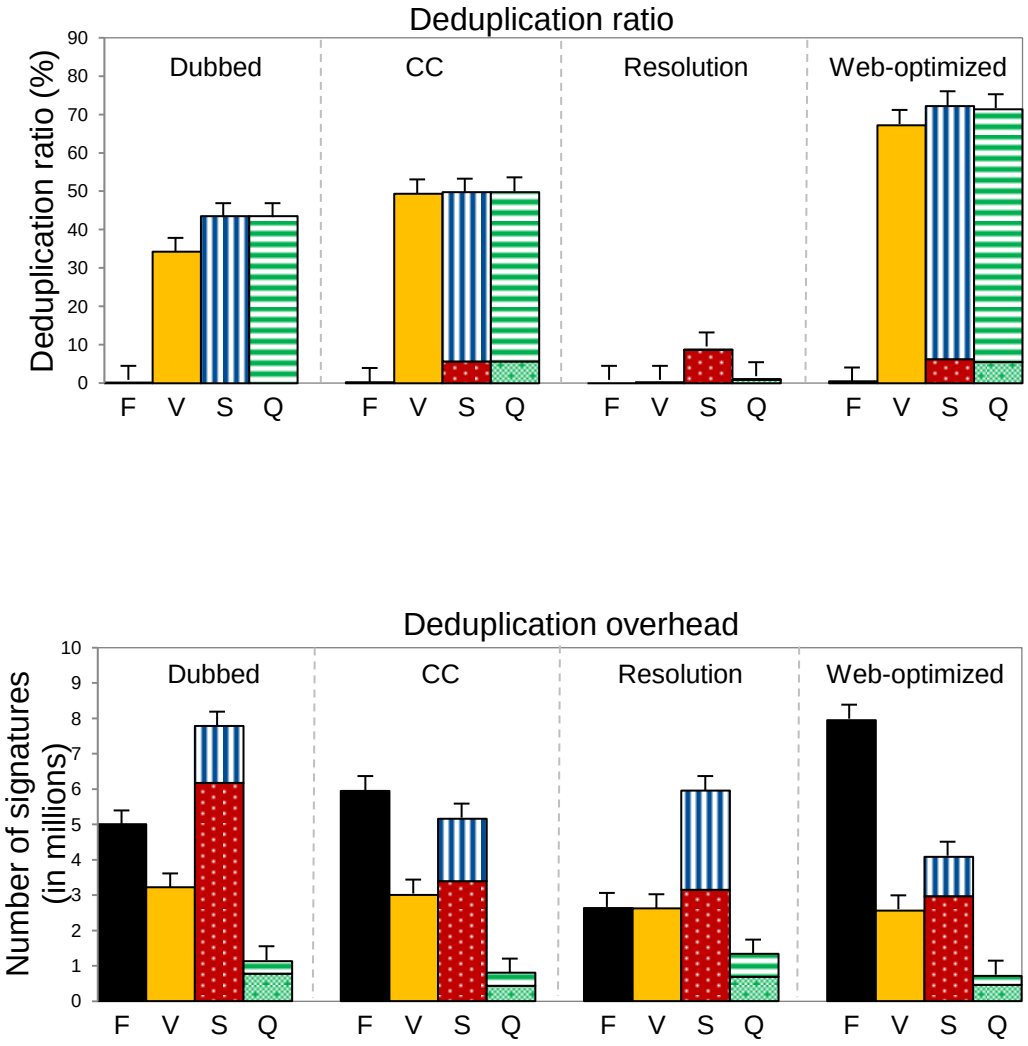
- **Deduplication overhead**

Number of **signatures** maintained in the fingerprint table

© 2015 NetApp, Inc. All rights reserved.

Results

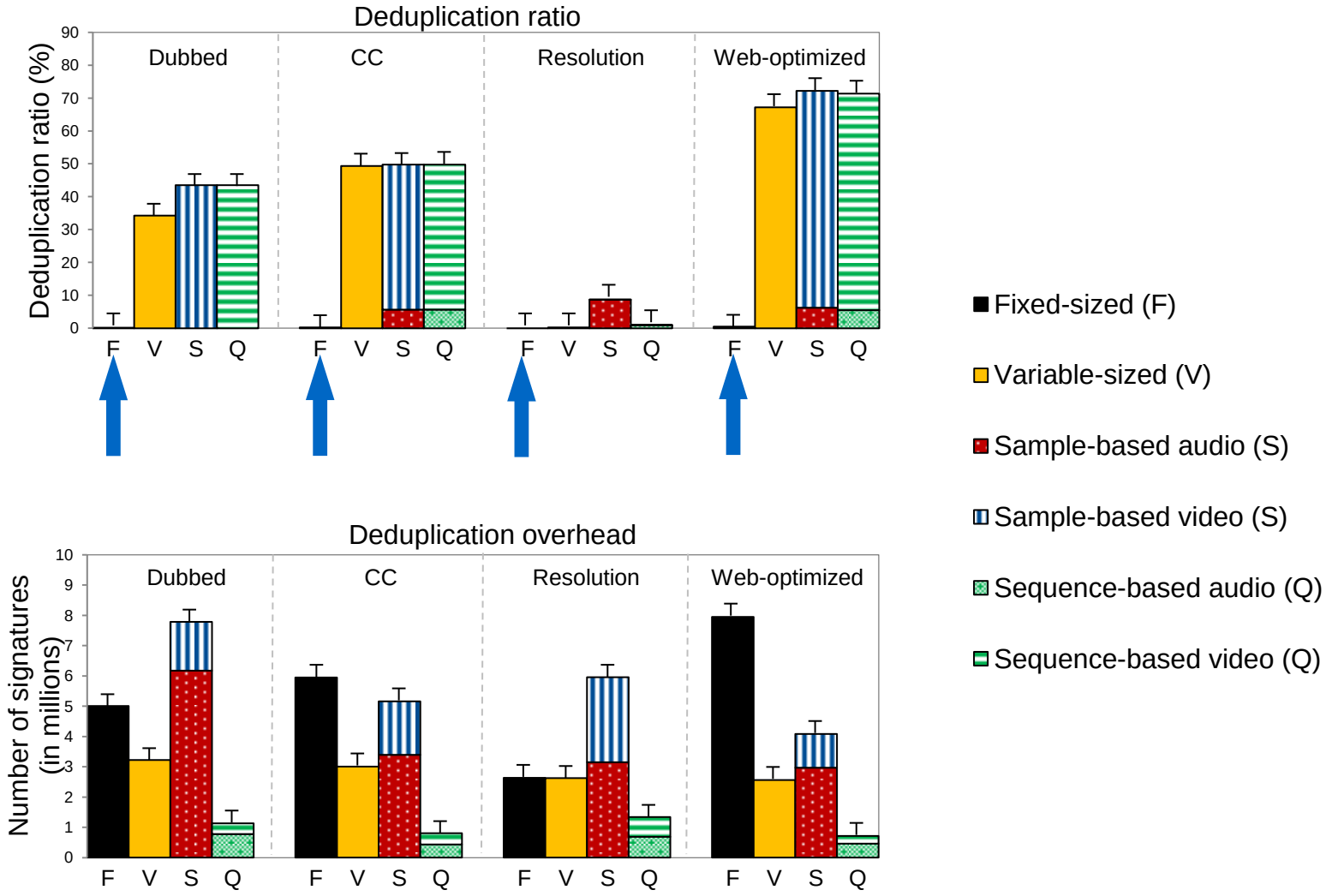
Content variation



© 2015 NetApp, Inc. All rights reserved.

Results

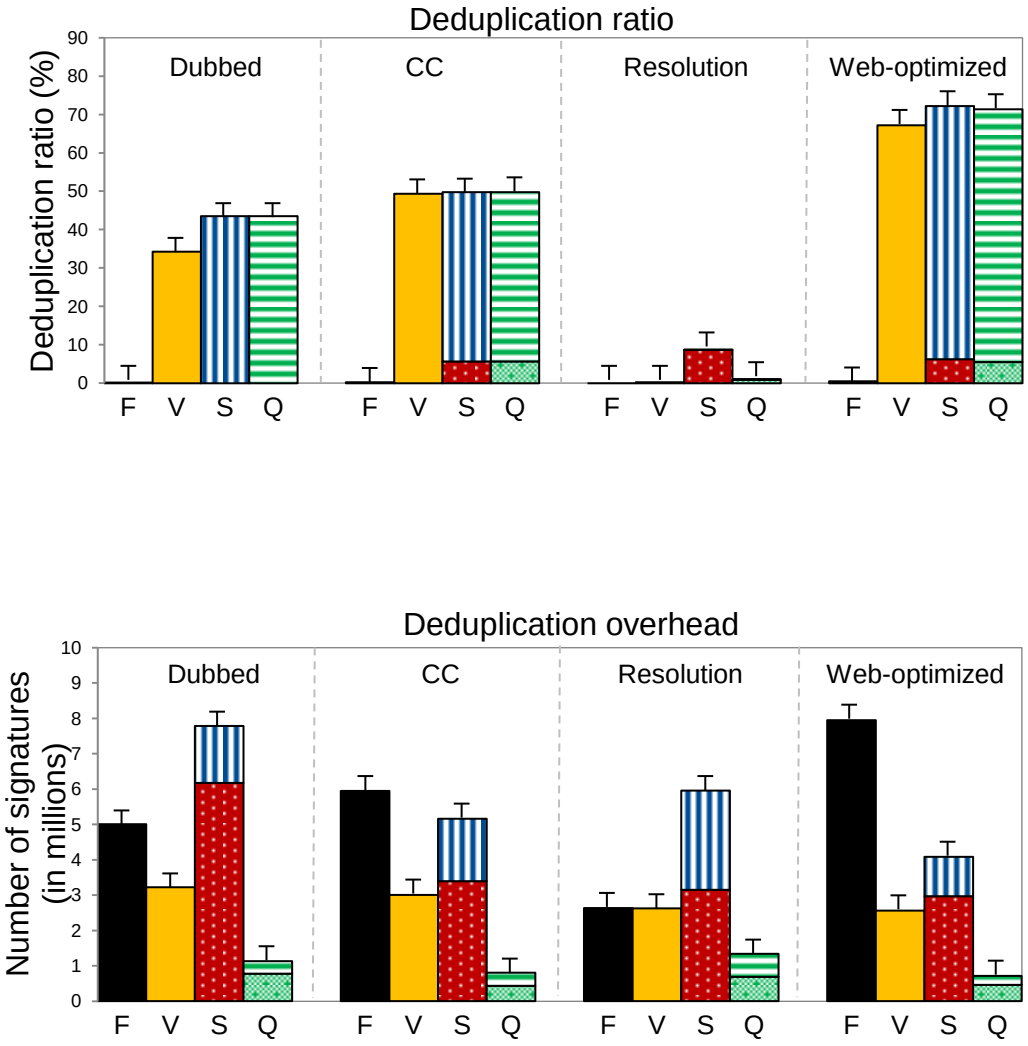
Content variation



© 2015 NetApp, Inc. All rights reserved.

Results

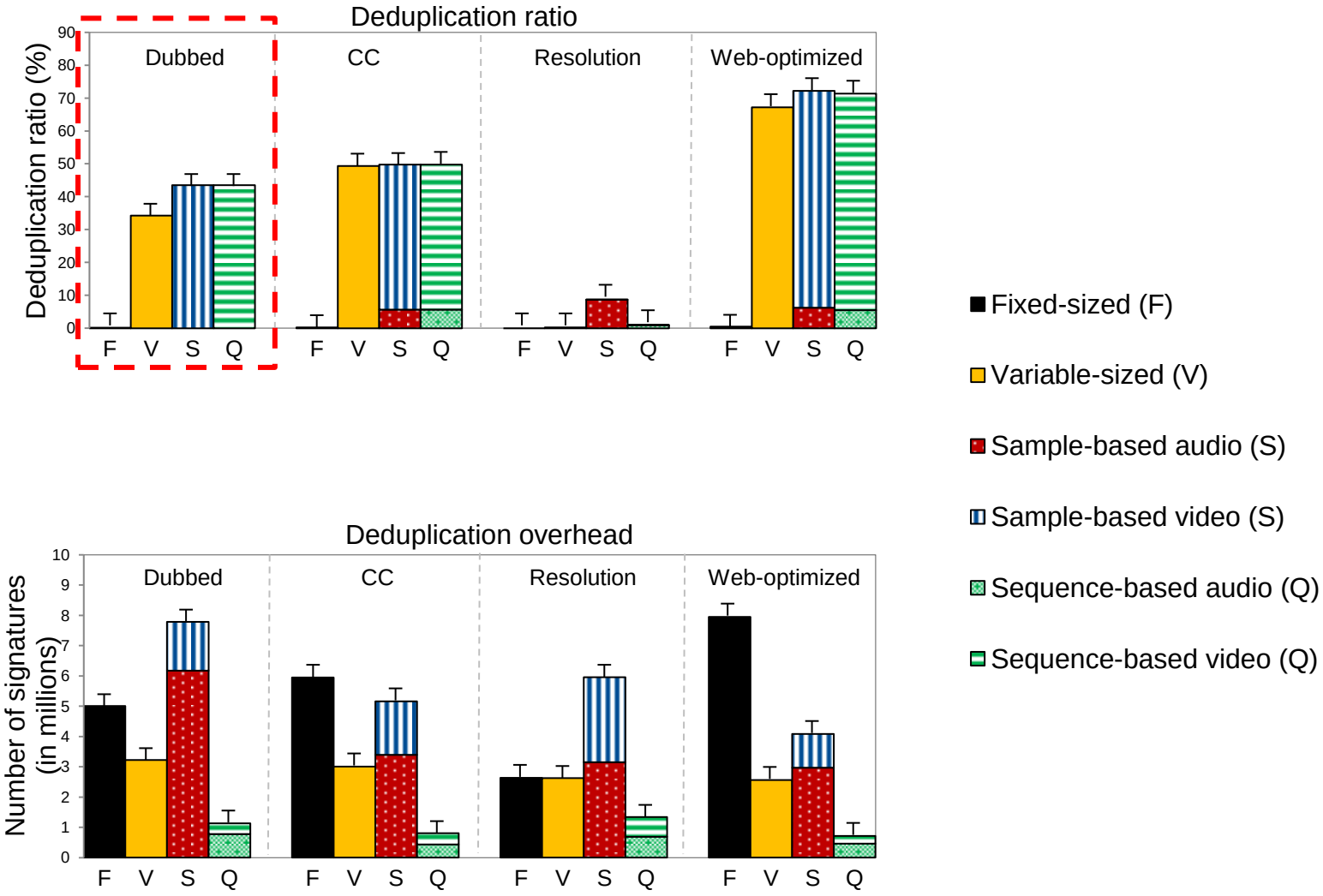
Content variation



© 2015 NetApp, Inc. All rights reserved.
NetApp

Results

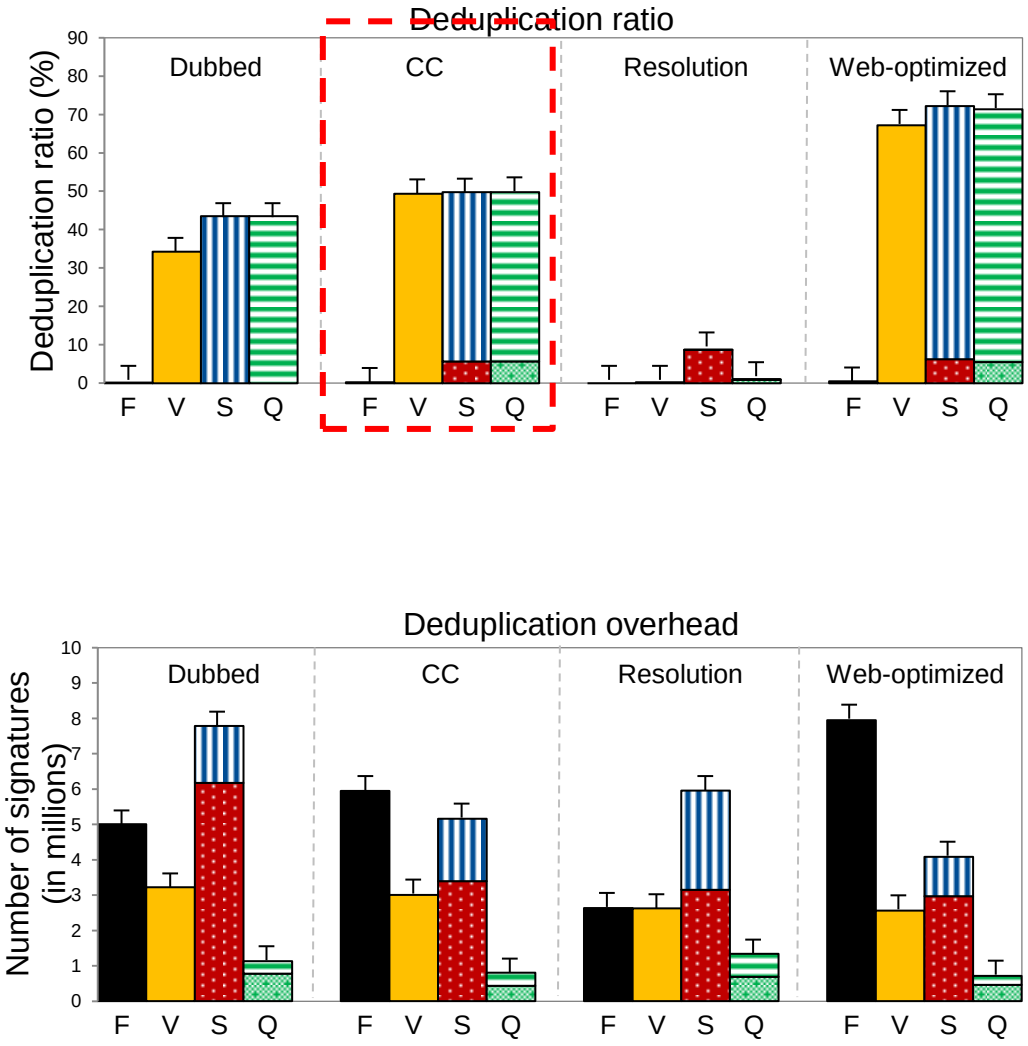
Content variation



© 2015 NetApp, Inc. All rights reserved.

Results

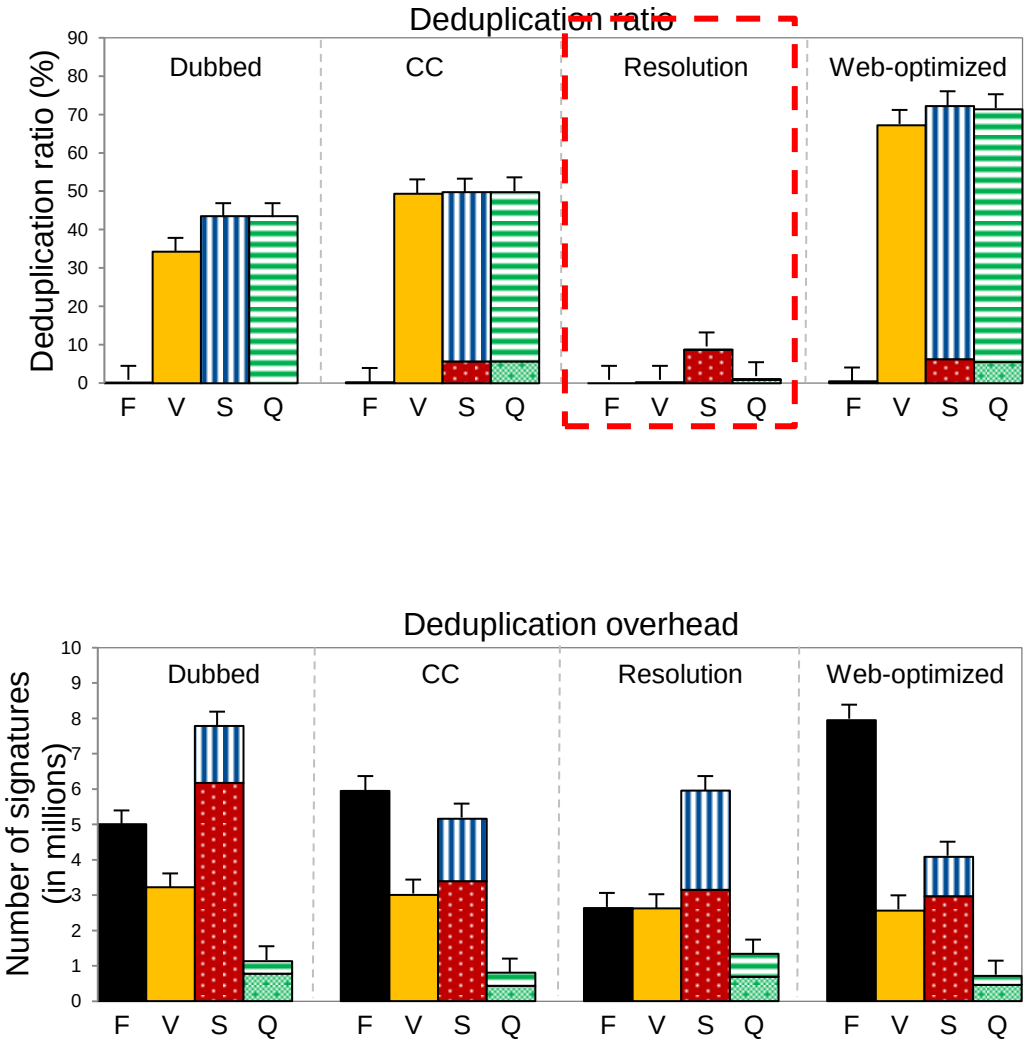
Content variation



© 2015 NetApp, Inc. All rights reserved.

Results

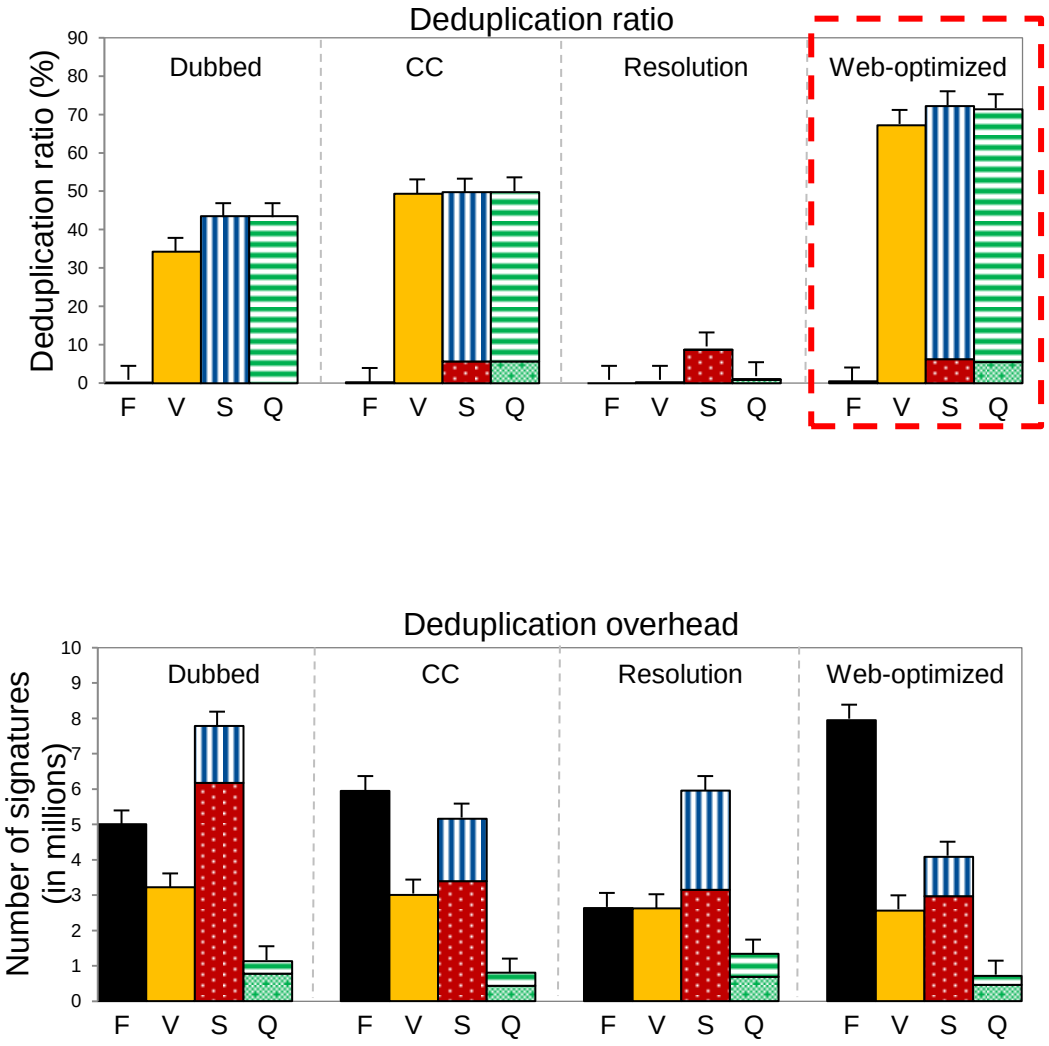
Content variation



© 2015 NetApp, Inc. All rights reserved.

Results

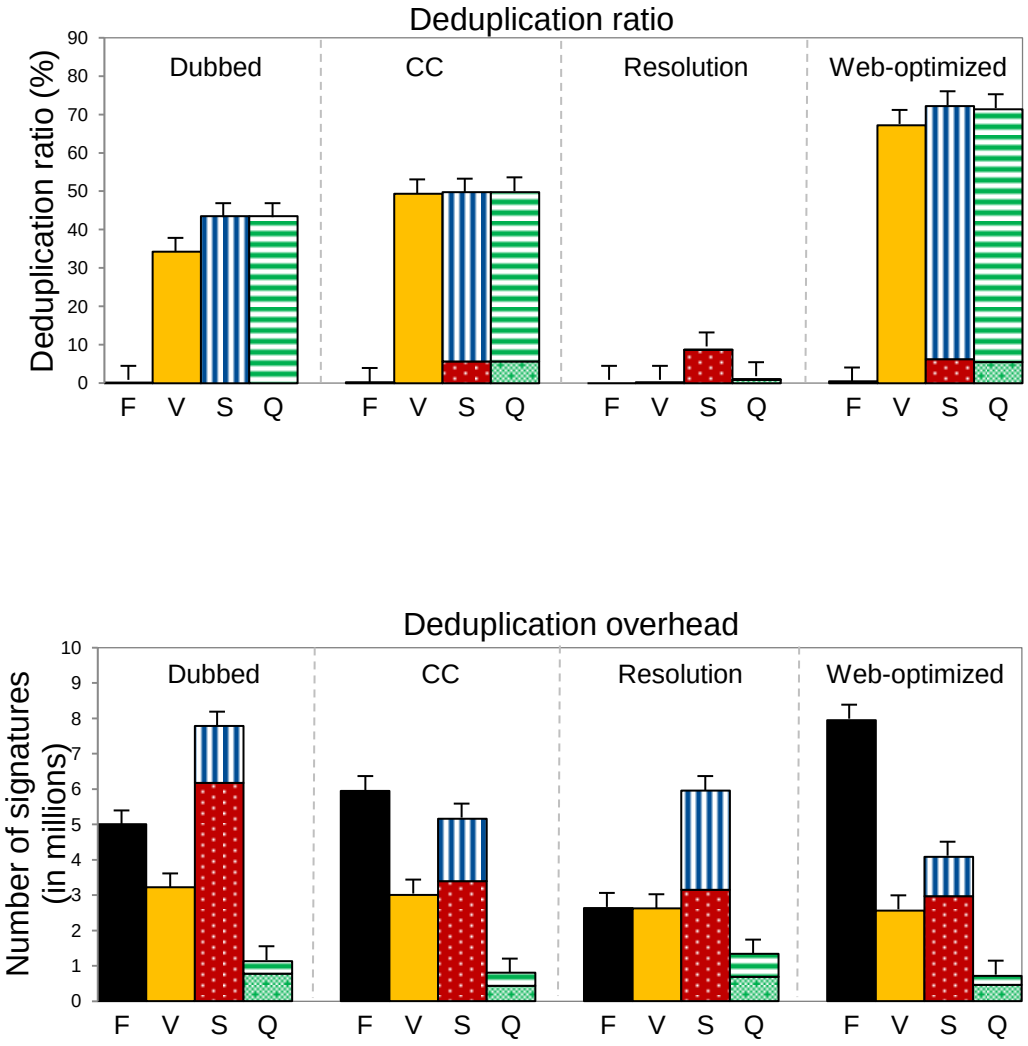
Content variation



© 2015 NetApp, Inc. All rights reserved.

Results

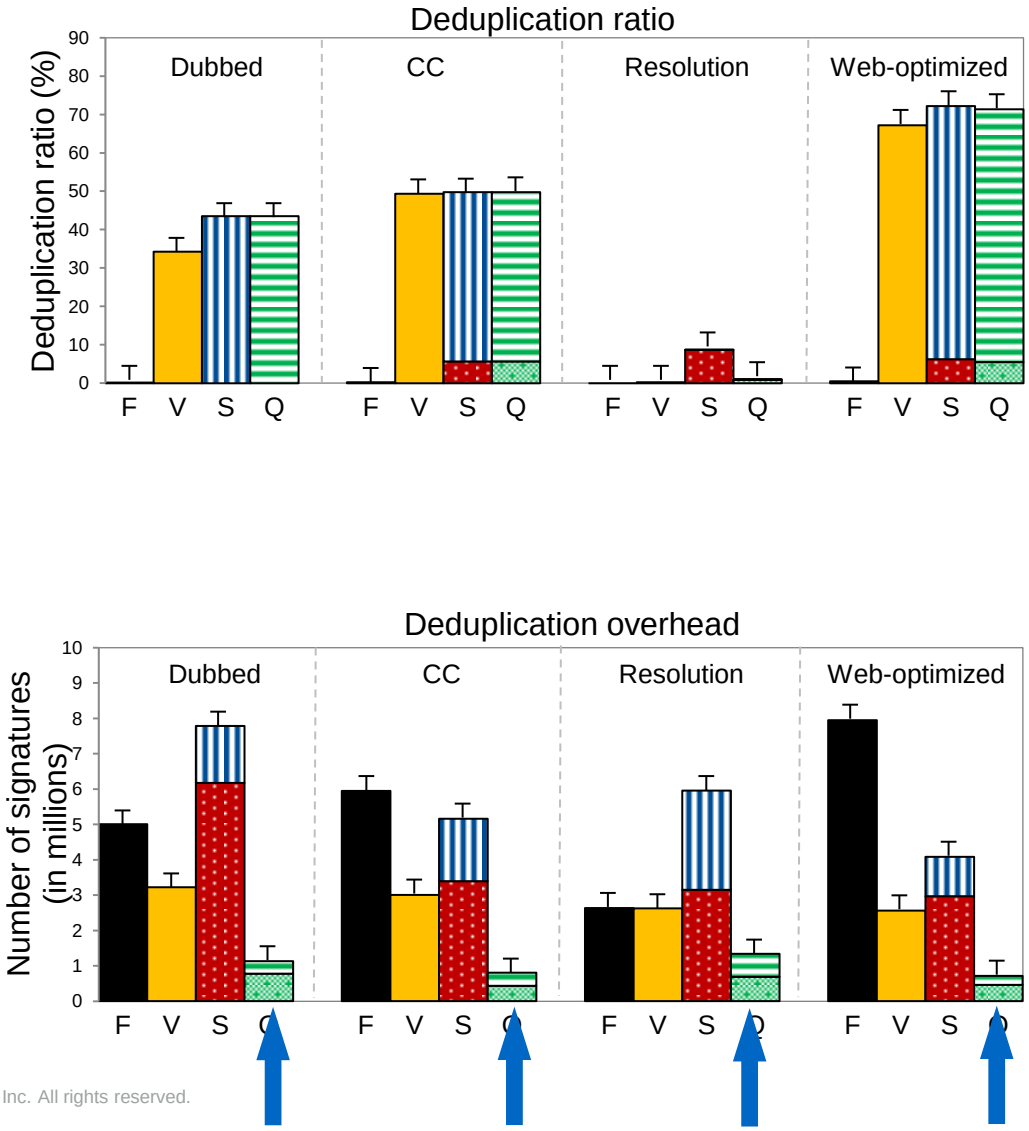
Content variation



© 2015 NetApp, Inc. All rights reserved.
NetApp

Results

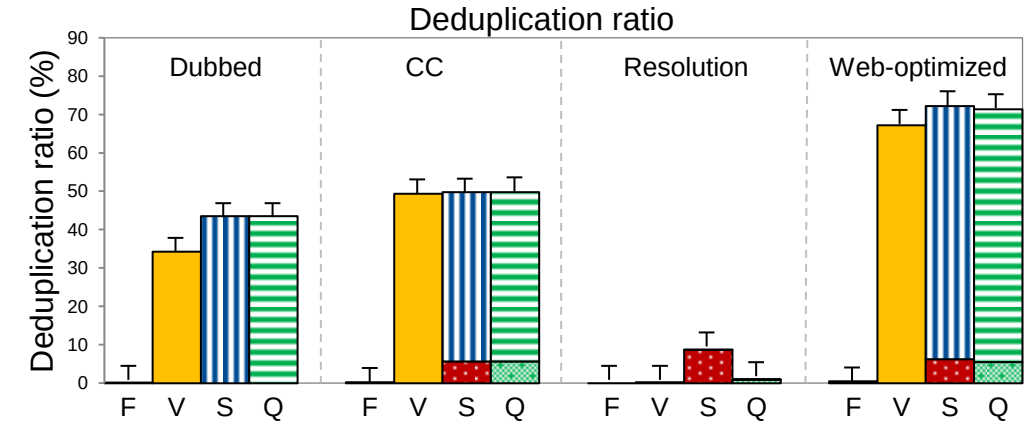
Content variation

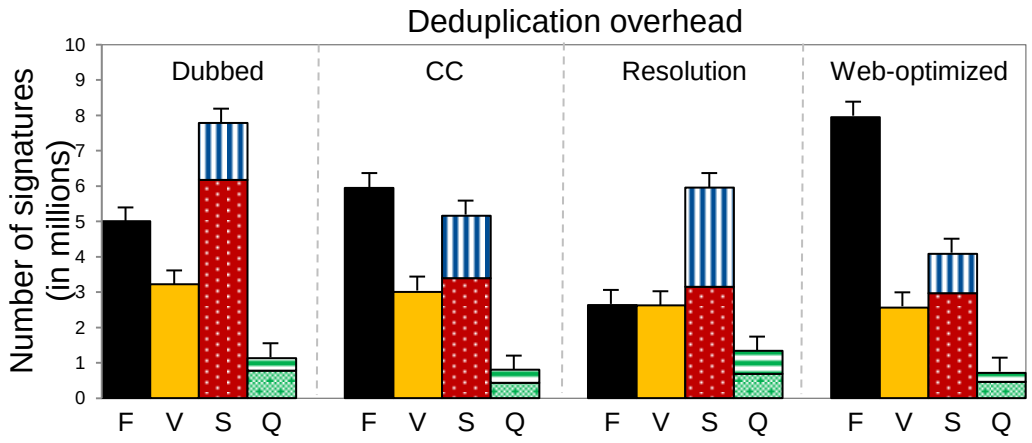


© 2015 NetApp, Inc. All rights reserved.

Results

Content variation



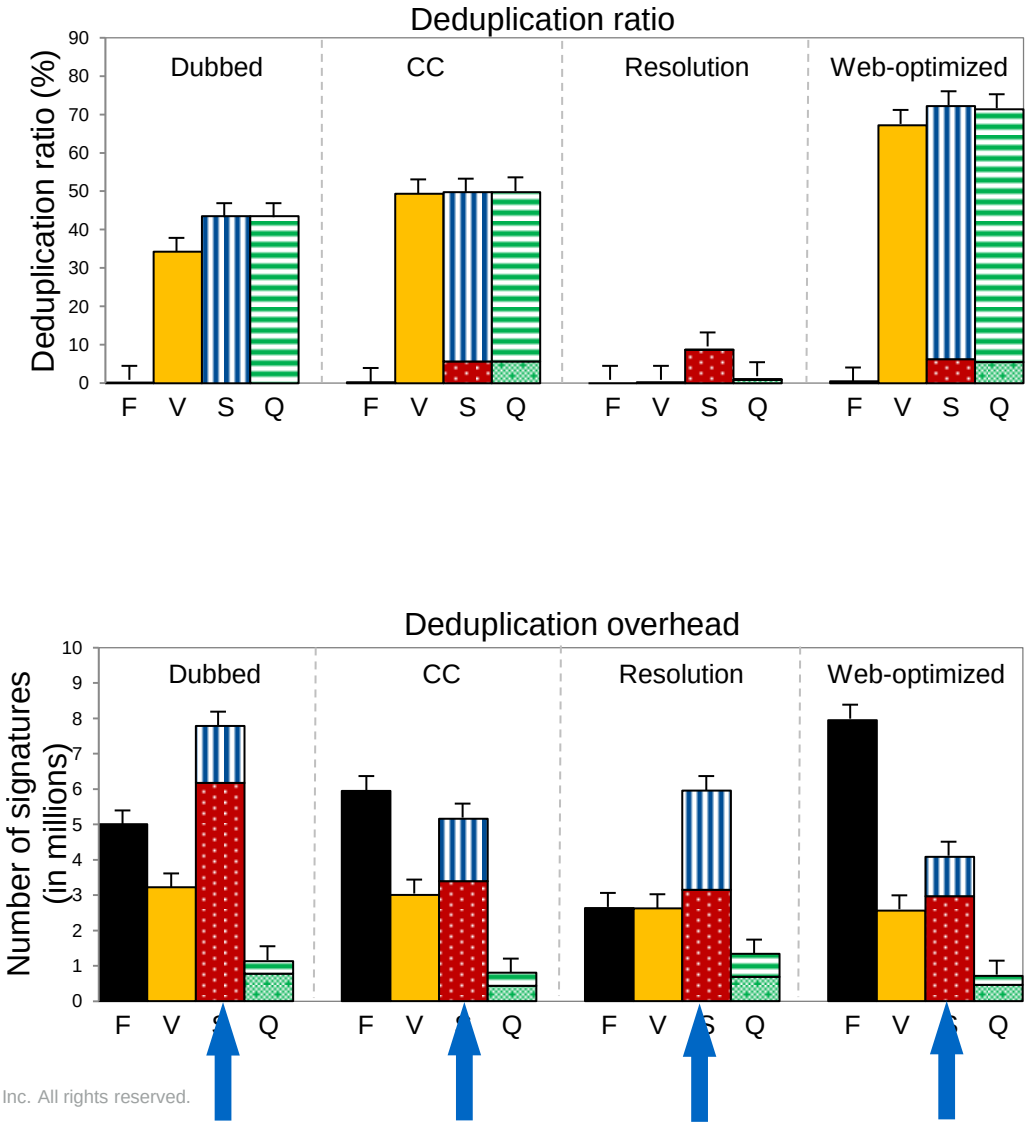


- Fixed-sized (F)
- Variable-sized (V)
- Sample-based audio (S)
- Sample-based video (S)
- Sequence-based audio (Q)
- Sequence-based video (Q)

© 2015 NetApp, Inc. All rights reserved.

Results

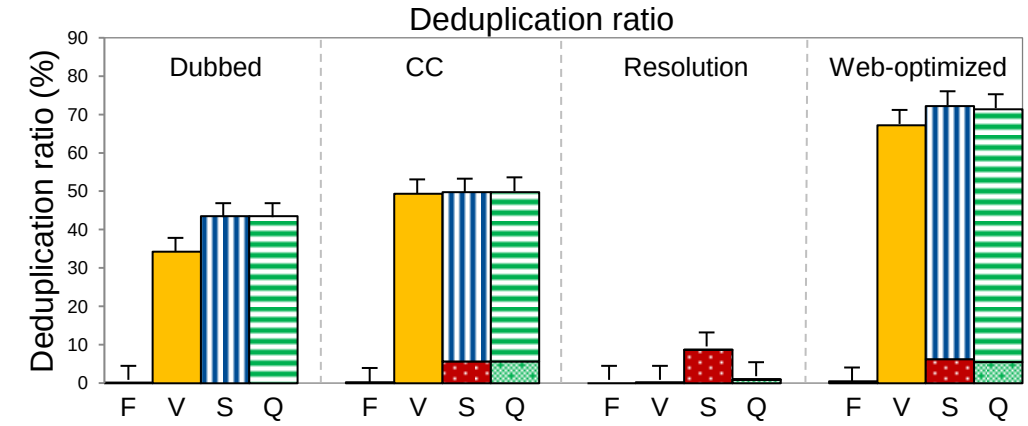
Content variation

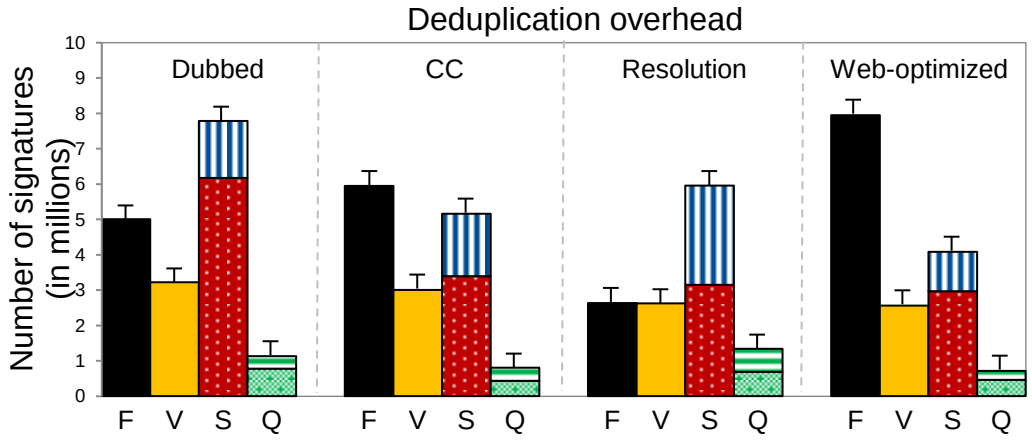


© 2015 NetApp, Inc. All rights reserved.

Results

Content variation



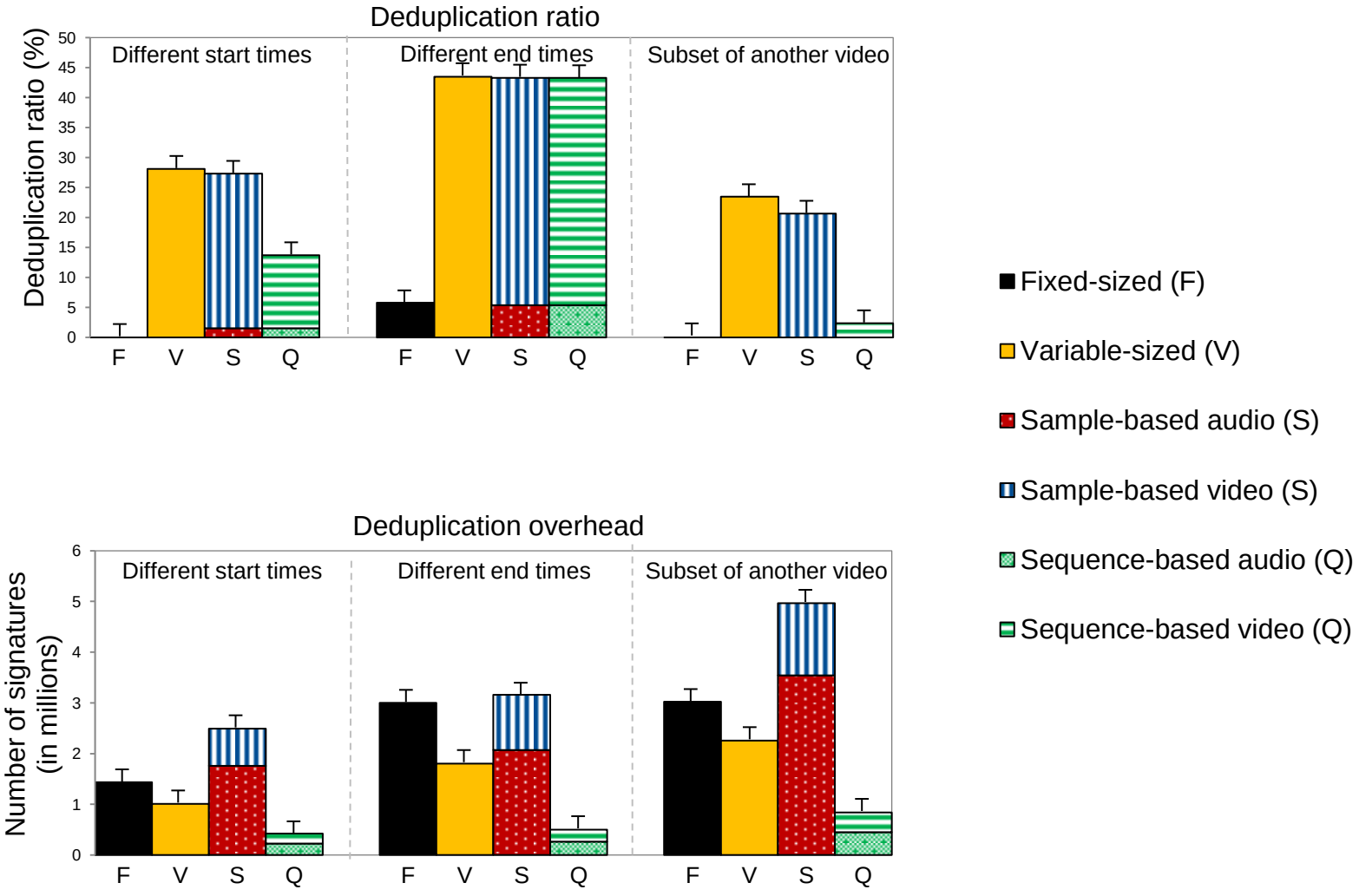


- Fixed-sized (F)
- Variable-sized (V)
- Sample-based audio (S)
- Sample-based video (S)
- Sequence-based audio (Q)
- Sequence-based video (Q)

© 2015 NetApp, Inc. All rights reserved.

Results

Video segments



© 2015 NetApp, Inc. All rights reserved.

What did we learn?

- Different techniques work well for different environments
- Content-aware techniques work well for controlled environments
- Variable-sized works well for repositories with multiple unknown formats
- Sequence-based has 81% lower signature table size
- Sample-based provides maximum deduplication to other techniques
- Exclude audio when using sample-based technique
- Variable-sized fails when sample ordering within sequences are different

© 2015 NetApp, Inc. All rights reserved.

Conclusion

- Deduplication works for video files
- Different ways storage efficiency is achieved in video repositories
- Compared the different system-level deduplication techniques
- Obtained inferences from the results
- In future, study different video file formats like FLV, AVI, etc.

© 2015 NetApp, Inc. All rights reserved.



Thank you

© 2015 NetApp, Inc. All rights reserved.