CHARACTERISTICS OF BACKUP WORKLOADS IN PRODUCTION SYSTEMS

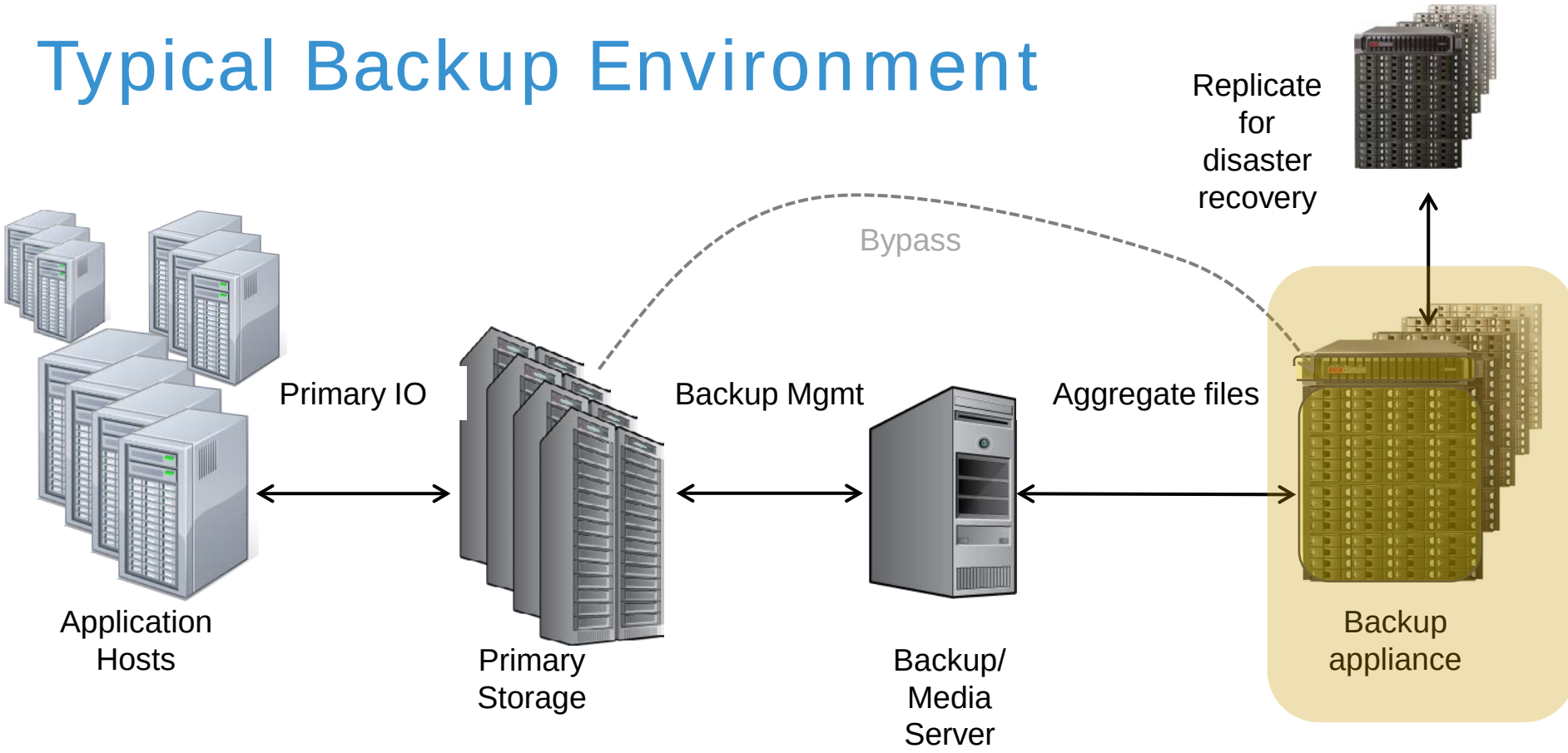
Grant Wallace, **Fred Douglis**, Hangwei Qian*, Philip Shilane, Stephen Smaldone, Mark Chamness, Windsor Hsu

Backup Recovery Systems Division
EMC Corporation

*Case Western Reserve Univ.

EMC²

Typical Backup Environment

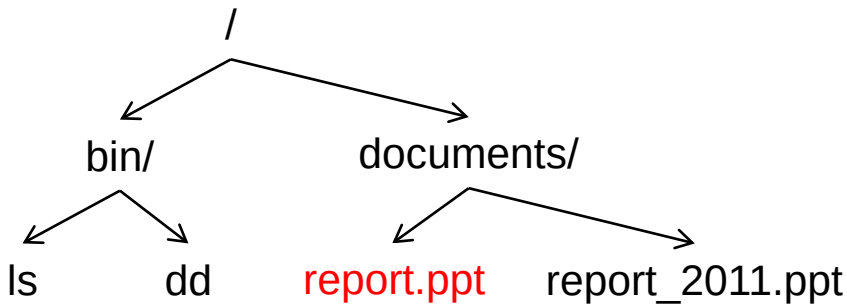


More limited DRAM cache

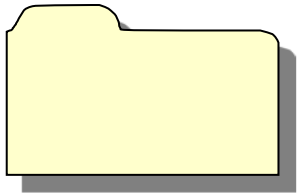
View from the backup appliance

- Aggregated, large backup files

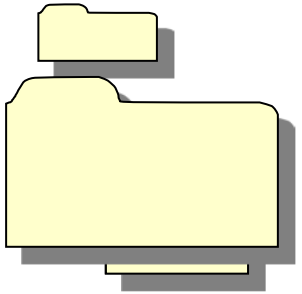
Backup Data Patterns



- Weekend: full backup
  100 GB tar-type file



- Monday - Friday: incremental
  1 GB tar-type file
- Weekend: full backup
  100 GB tar-type file
- Retained for months



- Full backups have the majority of bytes transferred to the backup appliance, though a small fraction of the bytes have changed

EMC Data Domain Appliance

- Purpose-built backup appliance
  - Designed to identify duplicate regions of files and replace with references
- Deduplication
  - Content-defined chunks, fingerprinted with secure hash
  - Check each hash against previously stored data
  - New data written together are stored together [Zhu08]
- Generally claim 5-20X deduplication ratio and 2X compression ratio → 10-40X total data reduction
  - Depends on data change rate, backup pattern, and retention policy
- Deduplication ratio is

$$\frac{\text{Pre-deduplication } \textit{logical} \text{ data}}{\text{Post-deduplication } \textit{physical} \text{ data}}$$

Motivation

- Lots of analyses of “primary” storage systems but little characterization of backup
  - Estimates of 8EB of data stored on disk-based purpose-built protection appliances by 2015
- Backup statistics tend to be single-dimension
  - “Our systems average 10x deduplication or better”
- Performance optimizations supported by limited datasets
  - “We compared our system against that other system using backups from this environment over that interval”
- Validate past design decisions using more extensive data, and provide data for future analyses

EMC²

Two-Pronged Analysis

- Broad study
  - Snapshot of autosupport data to characterize production systems in statistical terms
  - Compare these metrics against primary storage systems
    - Meyer & Bolosky, FAST'11, Microsoft workstation data
- Deep content-based analysis
  - Statistics insufficient for some types of study
  - Collect anonymized metadata from customer and internal Data Domain backup systems
    - By specific agreement
    - Generate time-ordered representation of content ("trace")
    - Analyze impact of chunk size, caching policies

Broad Study: Autosupports

- Over 10,000 systems periodically send statistical information to a centralized repository
  - Deduplication and compression rates
  - Storage usage
  - File counts, ages, etc
  - Many others
- Took ASUPs from one week in July 2011
  - Exclude any systems younger than 3 months or with less than 2.5% of capacity in use

Content-based Analysis: Metadata Collection

- Why metadata?
  - Collecting entire content infeasible (size, privacy)
  - Examples of metadata:
    - Per-chunk (fingerprints, size, physical location)
    - Per-file (comprising fingerprints)
    - Sub-chunk fingerprints
  - Physical chunks, logical files
- Ensuring privacy
  - Anonymize filenames, paths, fingerprints, and any other content that can be matched to actual data

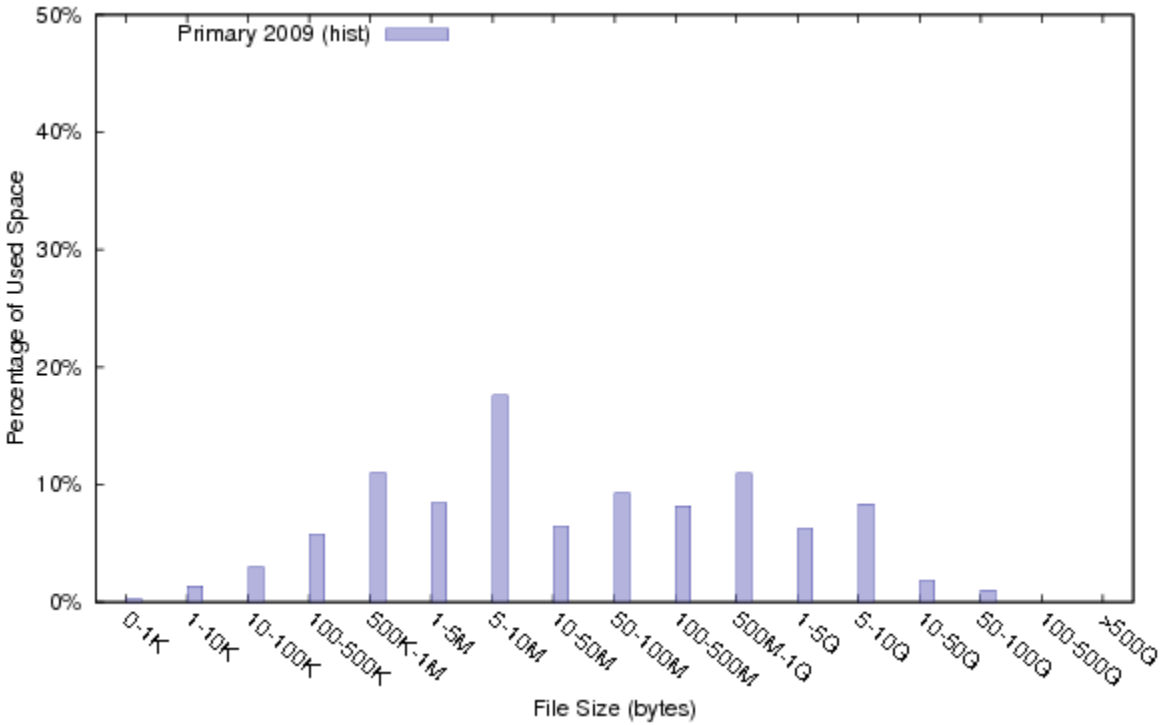
EMC²

Characteristics of Backup Filesystems

Autosupport Analysis

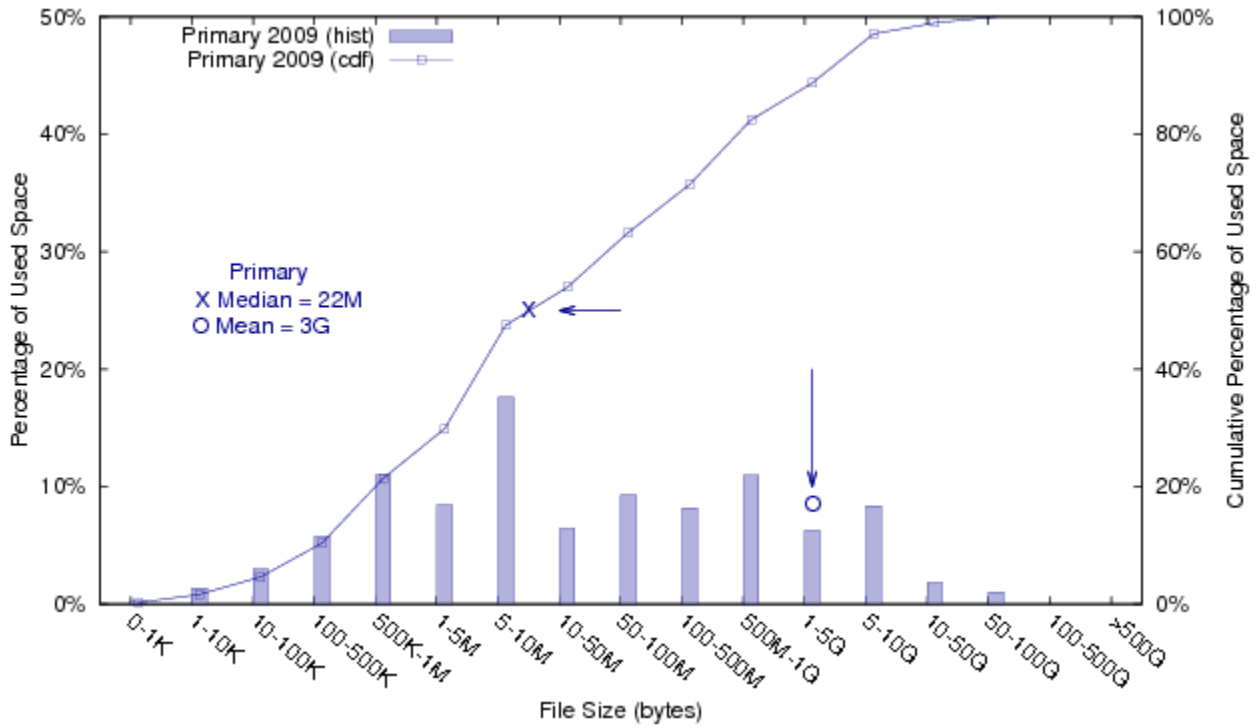
File Size by Space

- How do backup files compare to primary storage files?



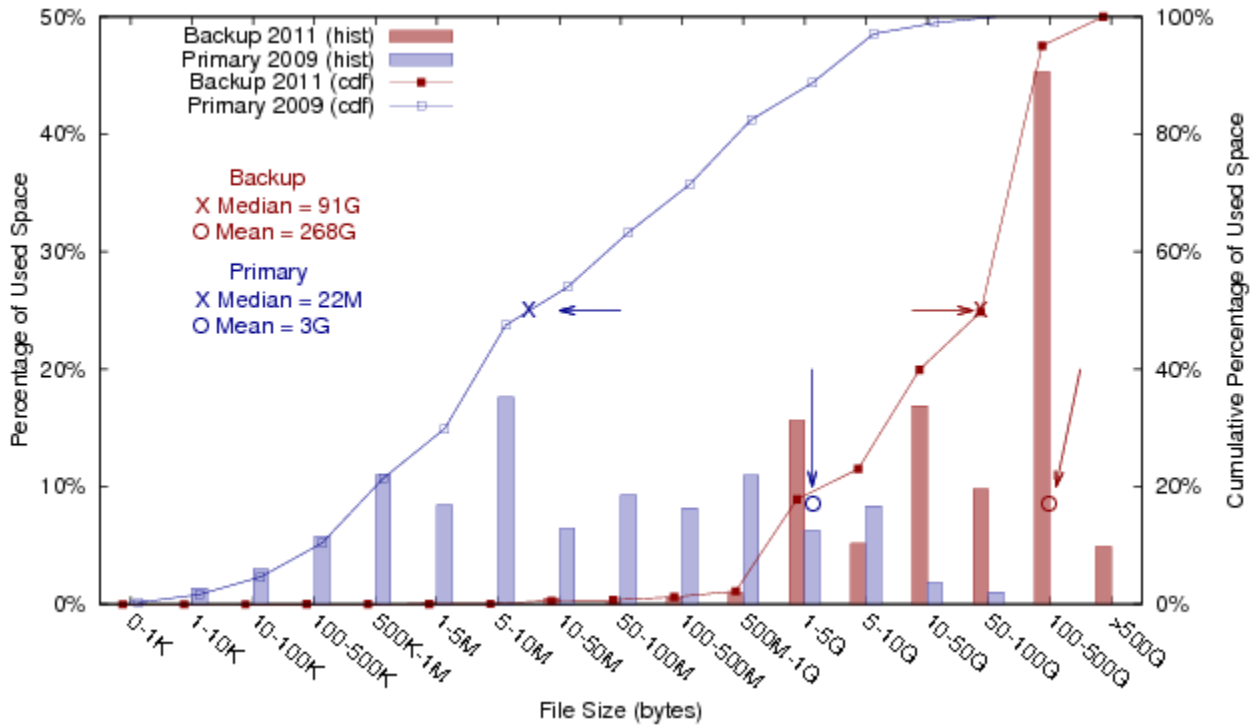
File Size by Space

- How do backup files compare to primary storage files?



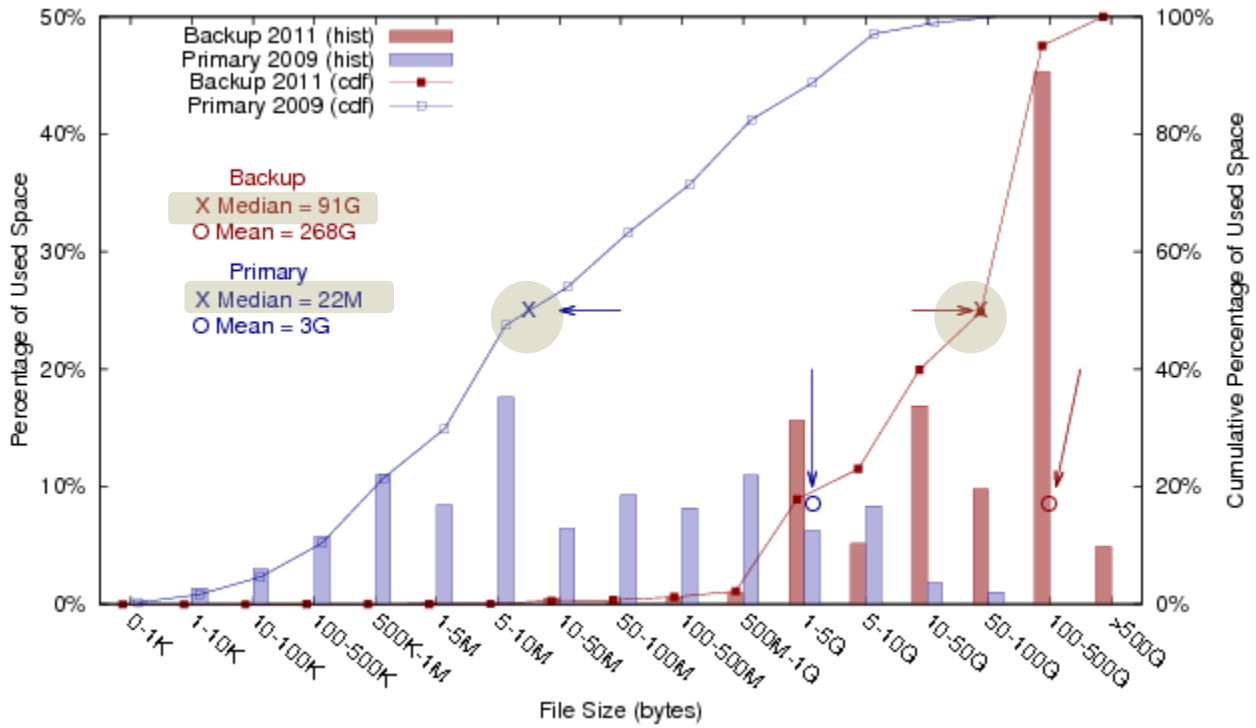
File Size by Space

- How do backup files compare to primary storage files?



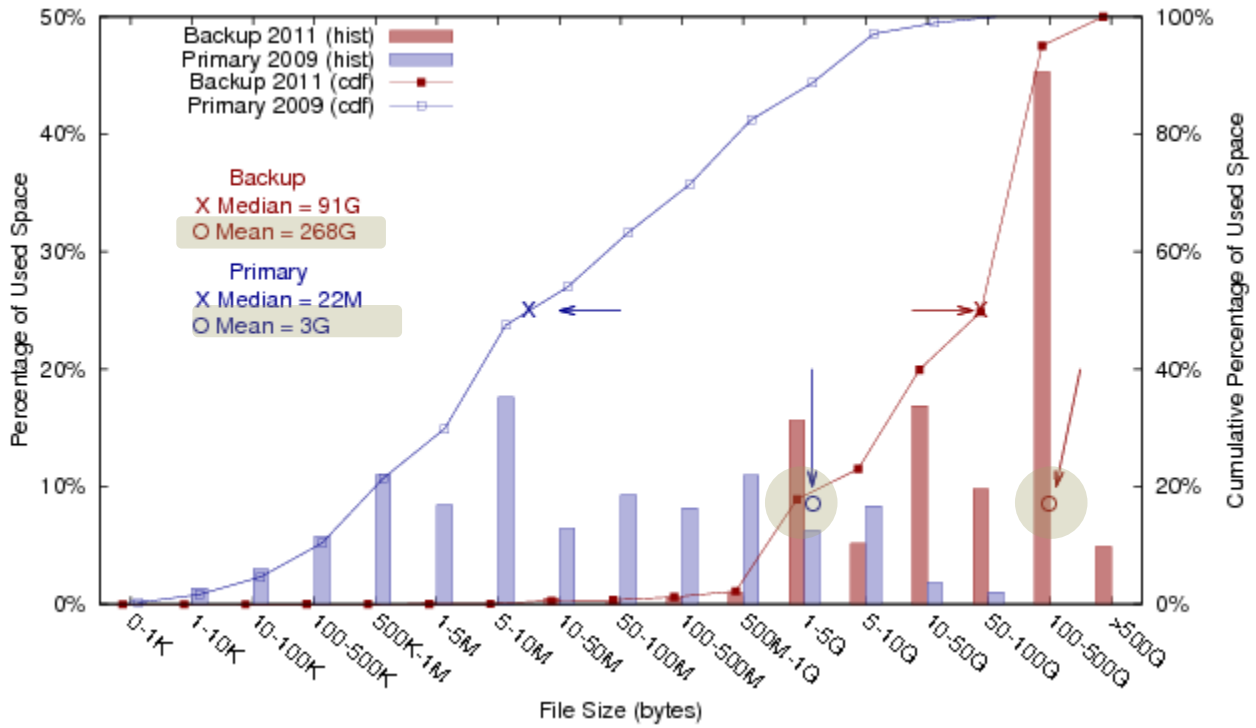
File Size by Space

- How do backup files compare to primary storage files?



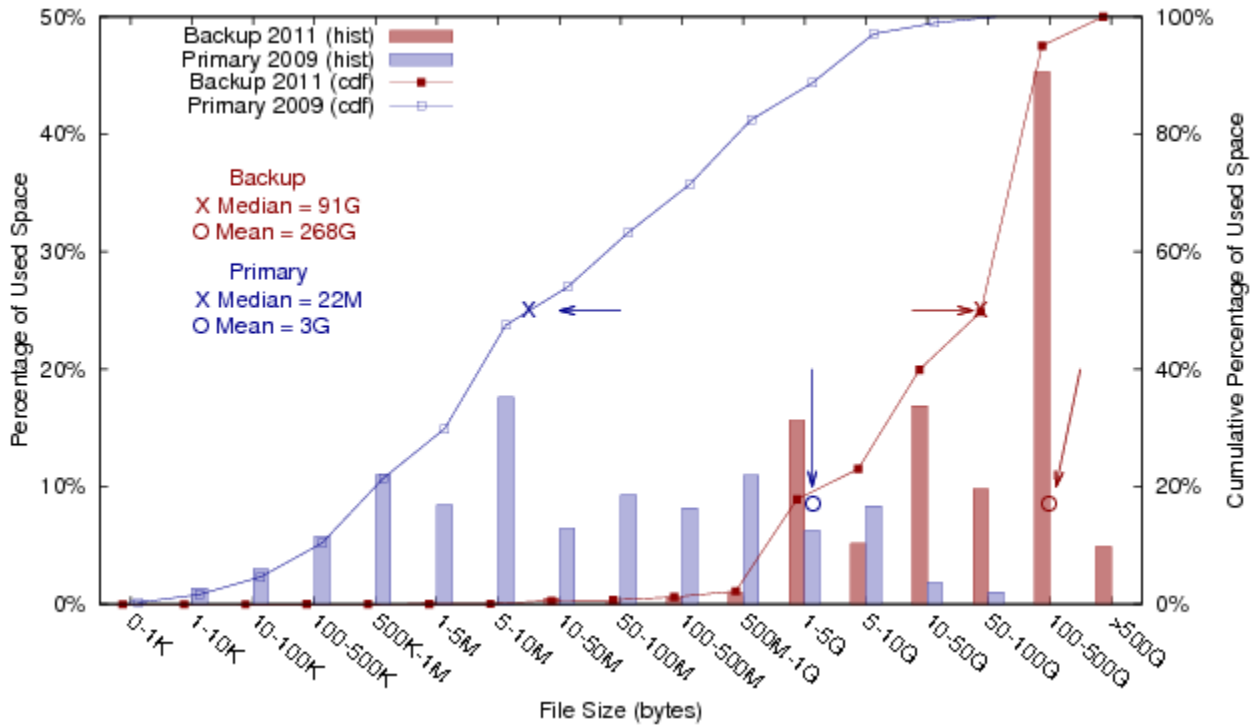
File Size by Space

- How do backup files compare to primary storage files?



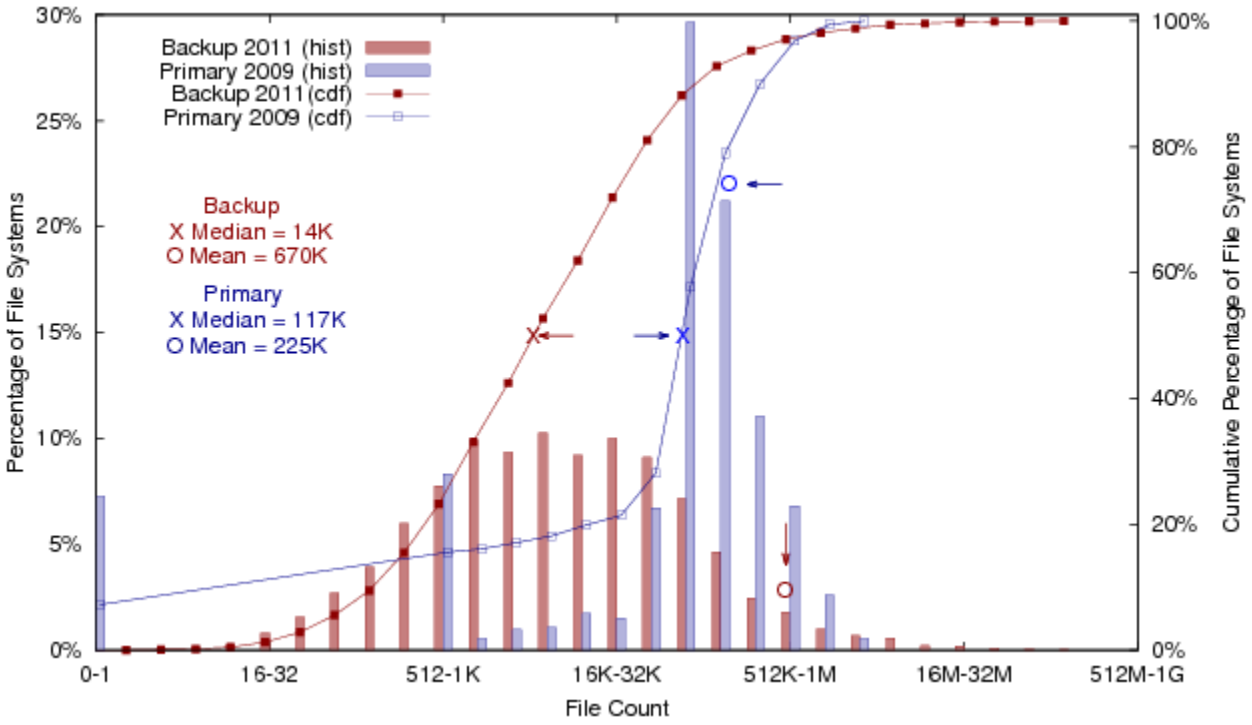
File Size by Space

- Backup files are orders of magnitude larger than primary storage files
- Small-file optimizations, e.g., embedding data in inodes, don't work for backup
- Use large allocation units



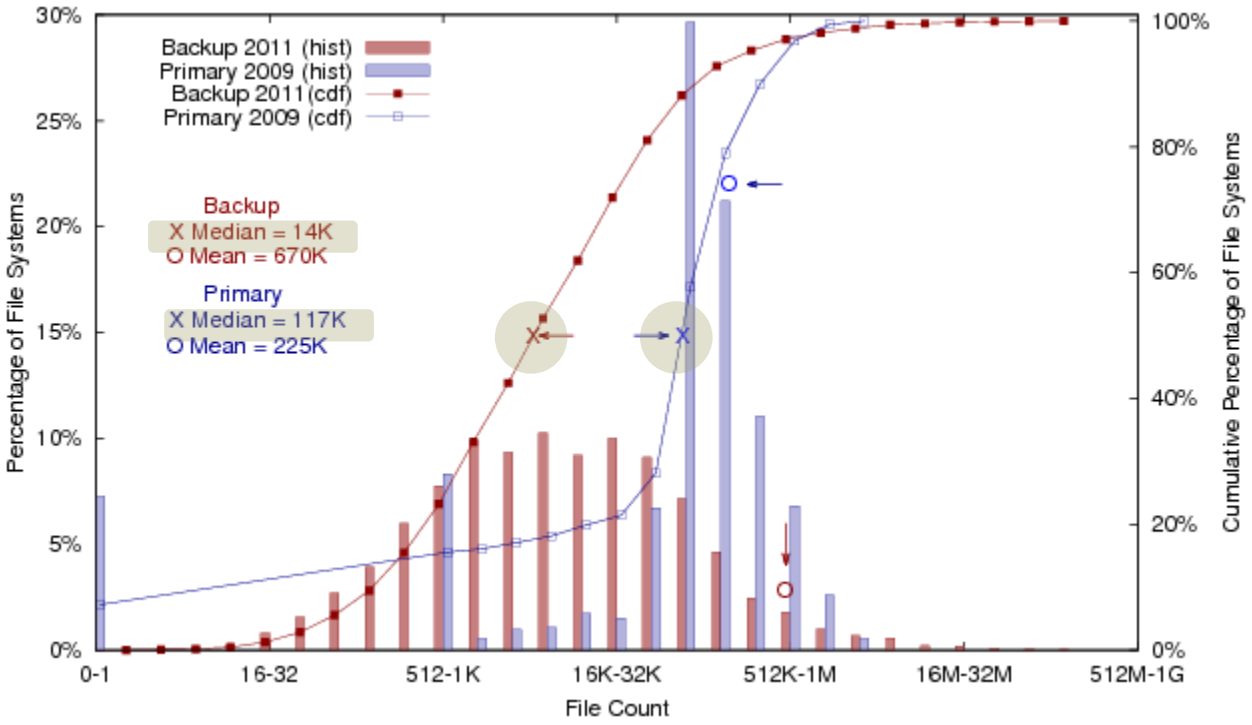
File and Directory Counts

- How sparse is the file system hierarchy?



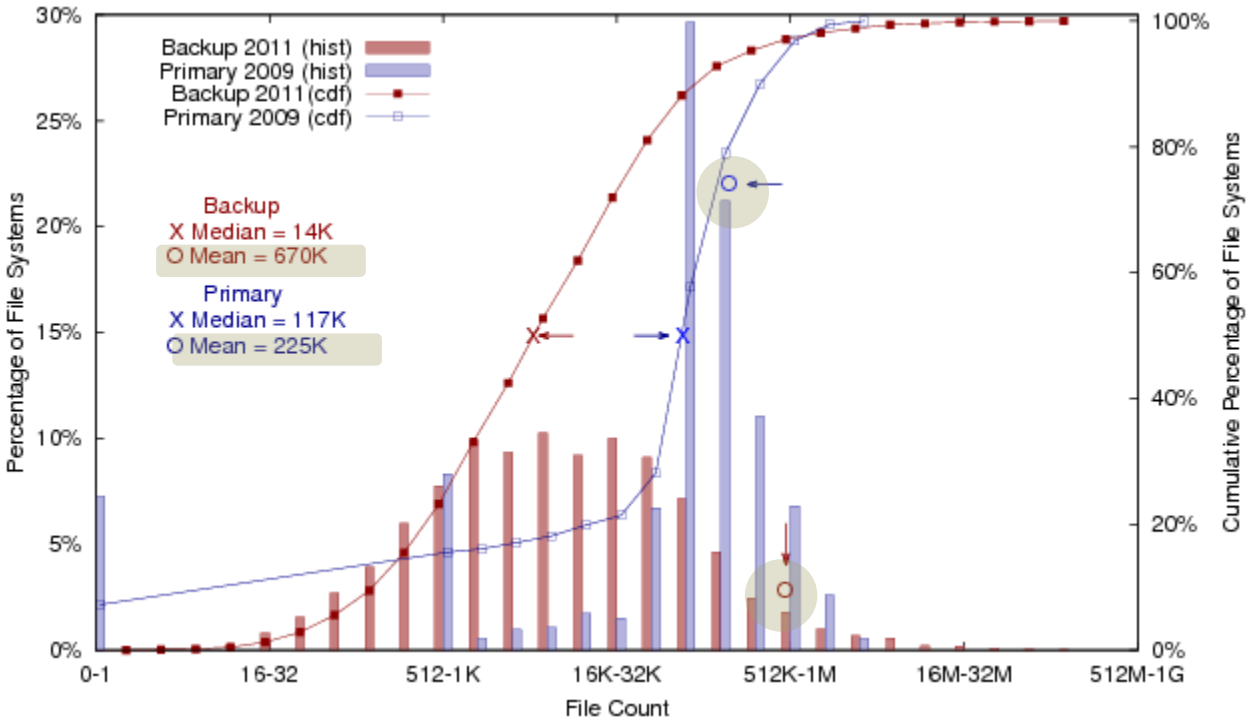
File and Directory Counts

- How sparse is the file system hierarchy?



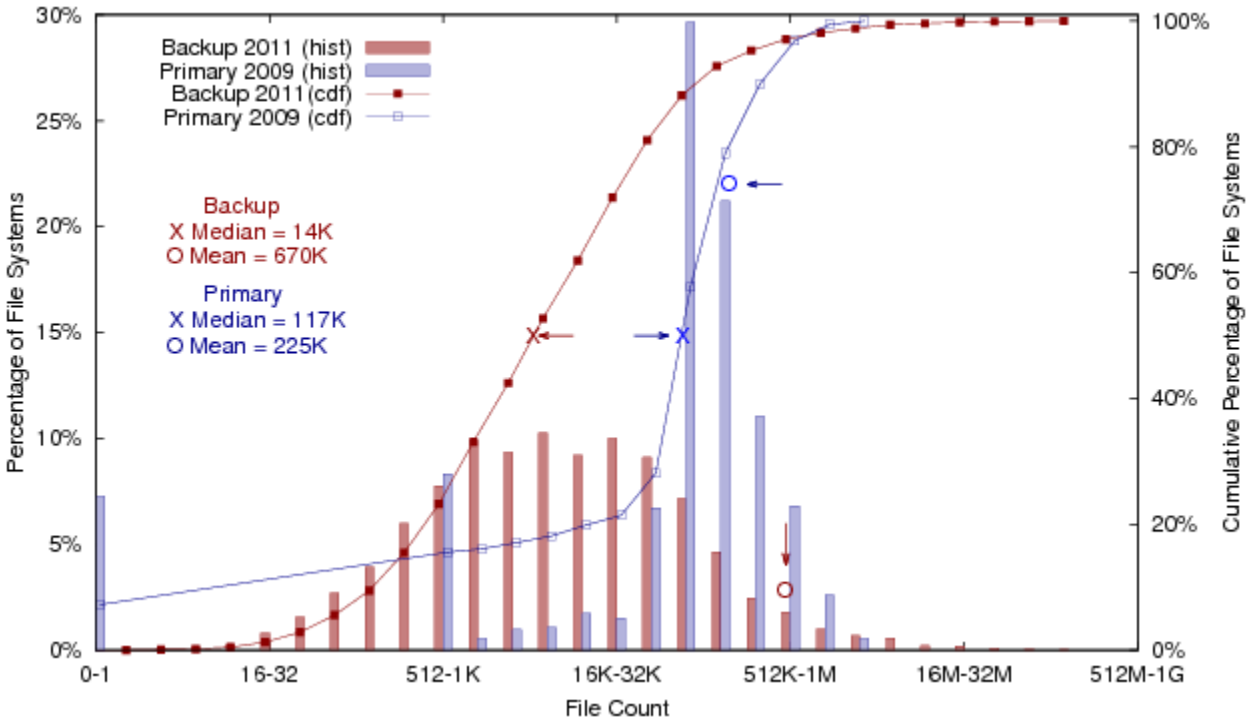
File and Directory Counts

- How sparse is the file system hierarchy?



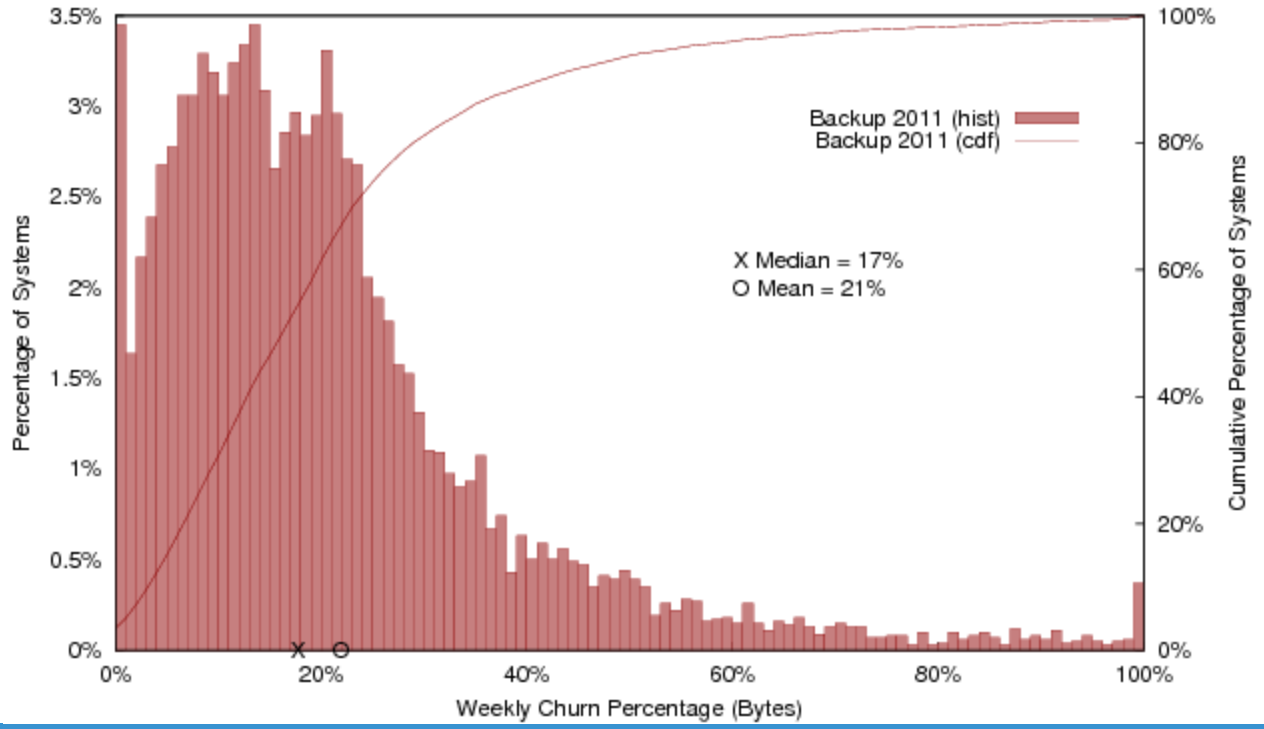
File and Directory Counts

- Many fewer files & directories on backup systems than primary
- Flat hierarchy for backup: many files per directory
- Backup software uses catalog – doesn't organize files the way humans do



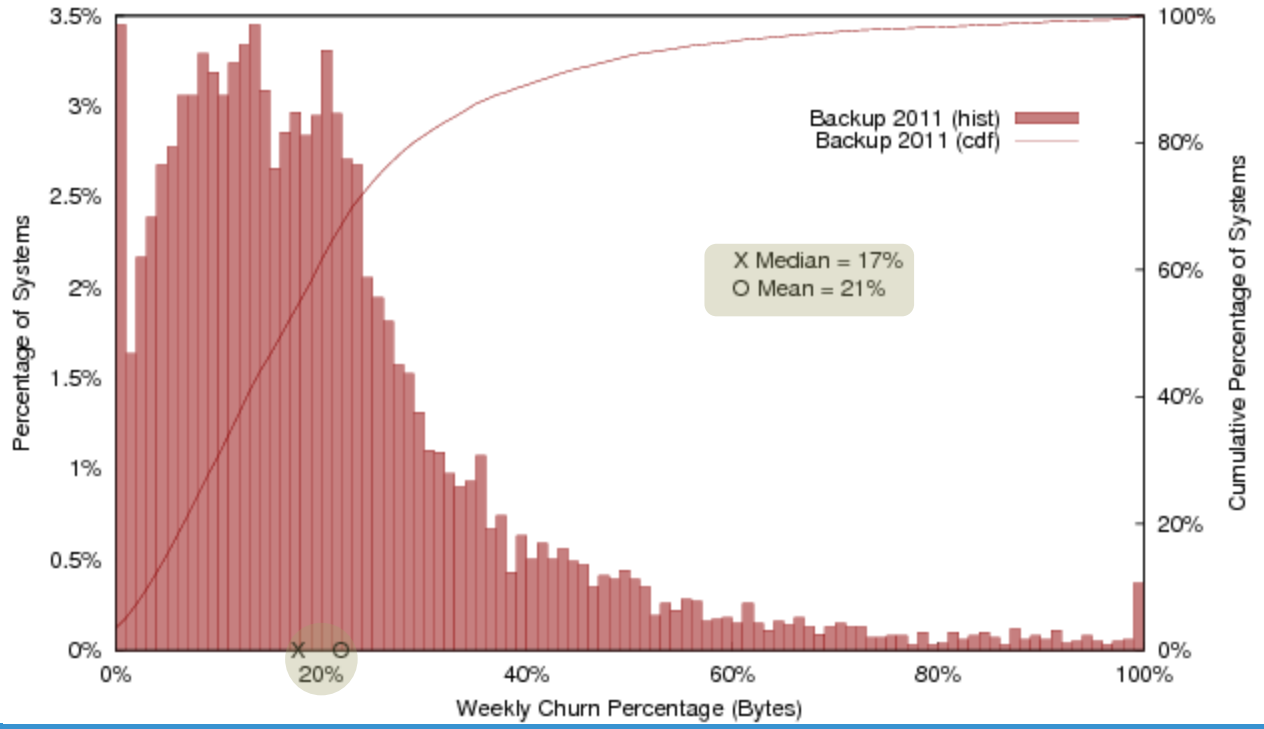
Weekly Churn

- How quickly does data get replaced in the backup appliance?
  - Logical churn: backup files deleted and added



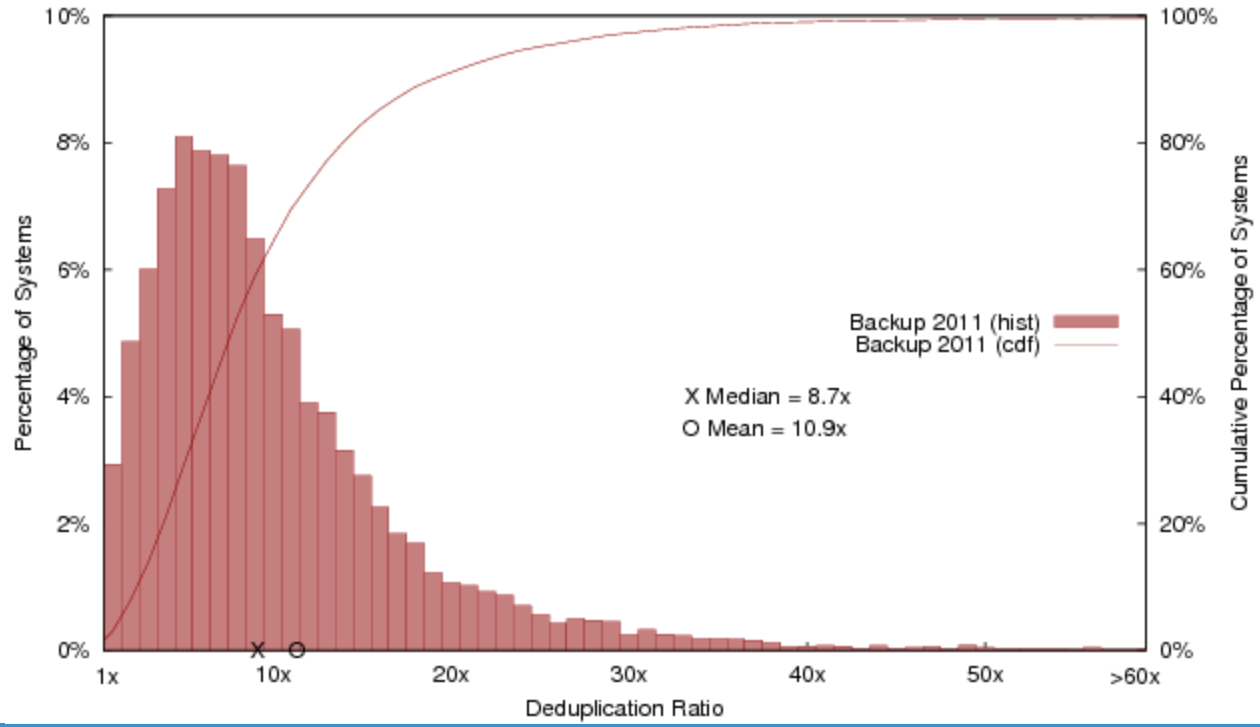
Weekly Churn

- On average, ~20% of total stored data freed & written per week
- System needs to be able to reclaim huge amounts of data on a regular basis
- Deduplication helps, since one physical copy can be retained over time



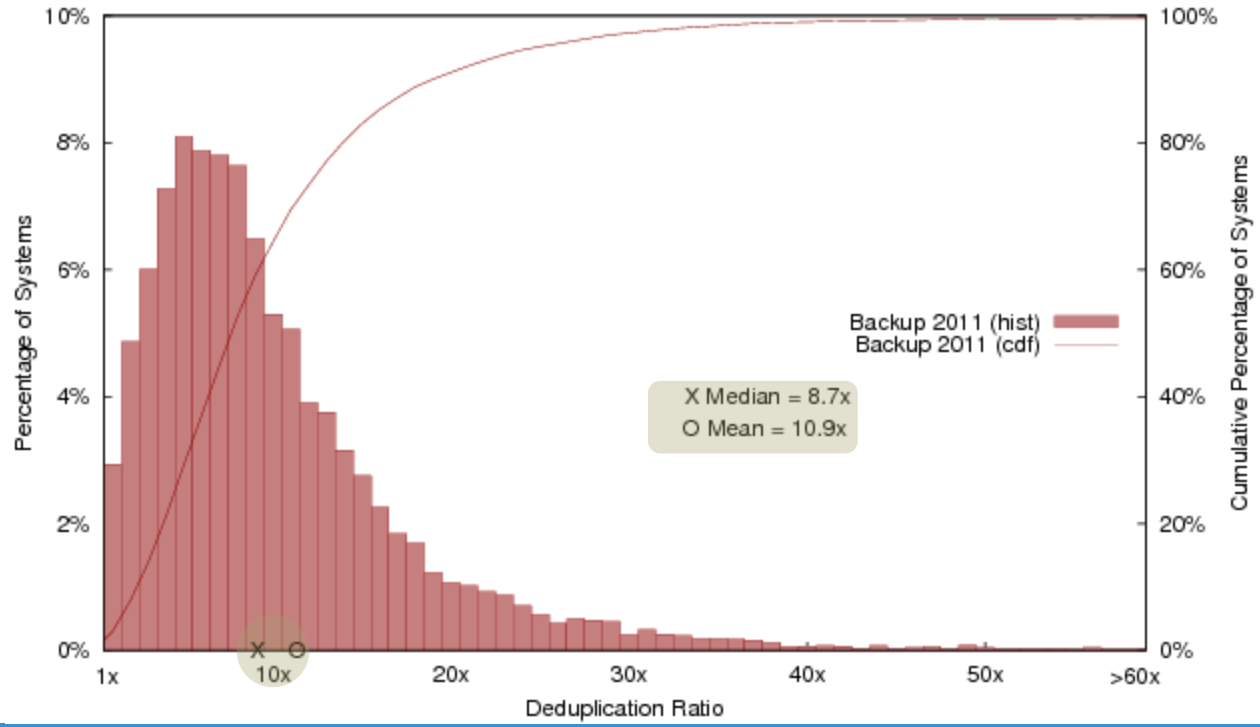
Deduplication

- How much deduplication do backup systems get?
  - Microsoft primary study was a single aggregate across many systems
  - Not directly comparable, but ~3X (cross-system), ~6X (4 weekly fulls)



Deduplication

- Long tail: some systems with 60x + dedupe!
  - Max dedupe seen is 384x!
- Much higher than primary workloads



Sensitivity Analyses

Content Metadata Analysis

Goals

- Assess impact of chunk size
  - What is the right size for a system?
  - Can we evaluate without the full data snapshot?
- Compare alternatives for caching
  - What is the right cache unit?

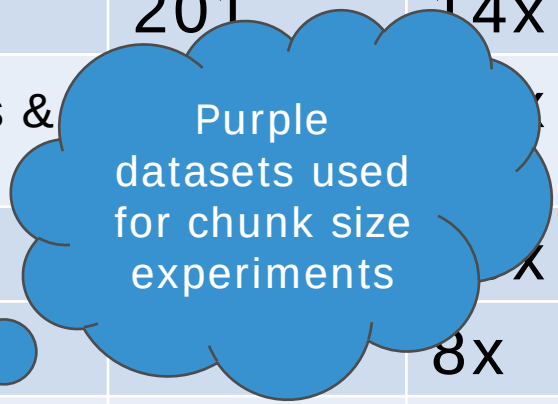
Data Set Characteristics

Dataset	Size (TB)	Deduplication	Median Age (weeks)
Homedirs	201	14x	3.5
Mixed2 (Workstations & Servers)	43	11x	9.4
Email	146	10x	1.4
Workstations	5	8x	13.6
Fileservers (Exchange, DB)	60	6x	5.8
Mixed1 (NAS)	47	6x	3.2
Database1	177	5x	2.2
Database2	4	2x	0.2

EMC²

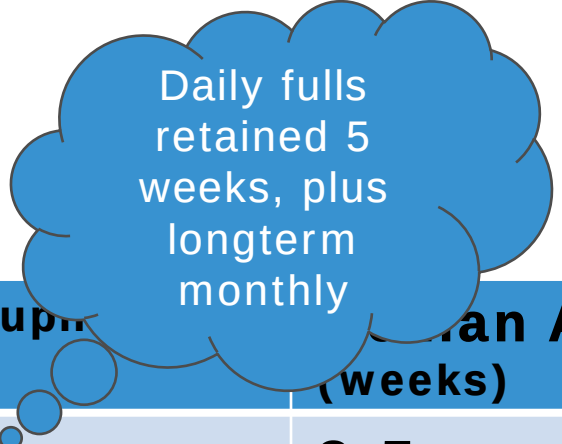
Data Set Characteristics

Dataset	Size (TB)	Deduplication	Median Age (weeks)
Homedirs	201	14x	3.5
Mixed2 (Workstations & Servers)			9.4
Email		x	1.4
Workstations		8x	13.6
Fileservers (Exchange, DB)	60	6x	5.8
Mixed1 (NAS)	47	6x	3.2
Database1	177	5x	2.2
Database2	4	2x	0.2



EMC²

Data Set Characteristics



Dataset	Size (TB)	Dedupl…	…an Age (weeks)
Homedirs	201	14x	3.5
Mixed2 (Workstations & Servers)	43	11x	9.4
Email	146	10x	1.4
Workstations	5	8x	13.6
Fileservers (Exchange, DB)	60	6x	5.8
Mixed1 (NAS)	47	6x	3.2
Database1	177	5x	2.2
Database2	4	2x	0.2

EMC²

Data Set Characteristics

Dataset	Size (TB)	Deduplication	Median Age (weeks)
Homedirs	201	14x	3.5
Mixed2 (Workstations & Servers)	43	11x	9.4
Email	146	10x	
Workstations	5	8x	
Fileservers (Exchange, DB)	60	6x	
Mixed1 (NAS)	47	6x	3.2
Database1	177	5x	2.2
Database2	4	2x	0.2

EMC²

Merging Chunks

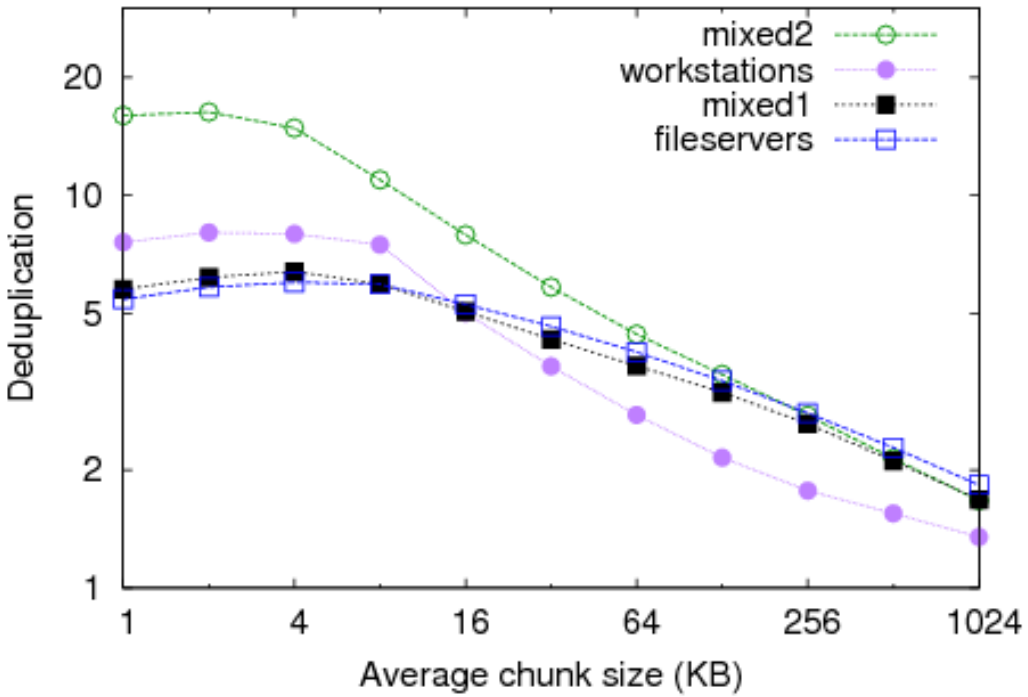
- Goal
  - Analyze deduplication rates across range of chunk sizes *without having access to the contents*
- Methodology
  - Collect fingerprints and sizes at standard 8K chunk size and of 1K sub-chunks
  - Merge 1K into 2K and 4K, and 8K into 16K+
    - Content-defined merging technique to make merges repeatable when content repeats
  - Consider impact of metadata overheads
    - Per-chunk overhead decreases effective deduplication on disk and adds to memory overheads
    - Greater relative overhead with higher deduplication, smaller chunks

EMC²

Impact of Chunk Size

- A rule of thumb is 15% better deduplication for each smaller power of 2 in chunk size, but about 2x the metadata
- Best deduplication is 4KB, but also considering cost to maintain data-structures and cleaning, 8KB is often a sweet spot
- A given dataset with small interspersed changes will see much more improvement

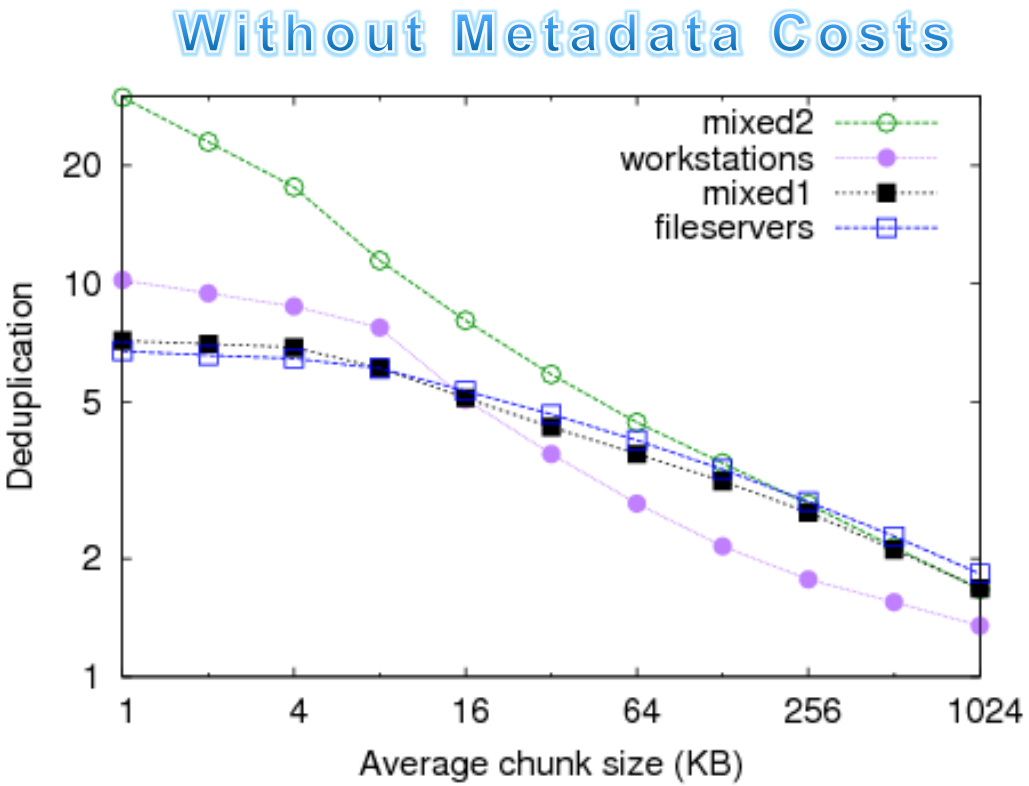
- For small chunks, metadata overhead dominates increased deduplication



Impact of Chunk Size

- A rule of thumb is 15% better deduplication for each smaller power of 2 in chunk size, but about 2x the metadata
- Best deduplication is 4KB, but also considering cost to maintain data-structures and cleaning, 8KB is often a sweet spot
- A given dataset with small interspersed changes will see much more improvement

- For small chunks, metadata overhead dominates increased deduplication



Impact of Chunk Size

- A rule of thumb is 15% better deduplication for each smaller power of 2 in chunk size, but about 2x the metadata
- Best deduplication is 4KB, but also considering cost to maintain data-structures and cleaning, 8KB is often a sweet spot
- A given dataset with small interspersed changes will see much more improvement
- For small chunks, metadata overhead dominates increased deduplication
- Microsoft study found whole-file deduplication got 87% of block-level deduplication for backups
  - Works for individual files but not when files are aggregated before backup

EMC²

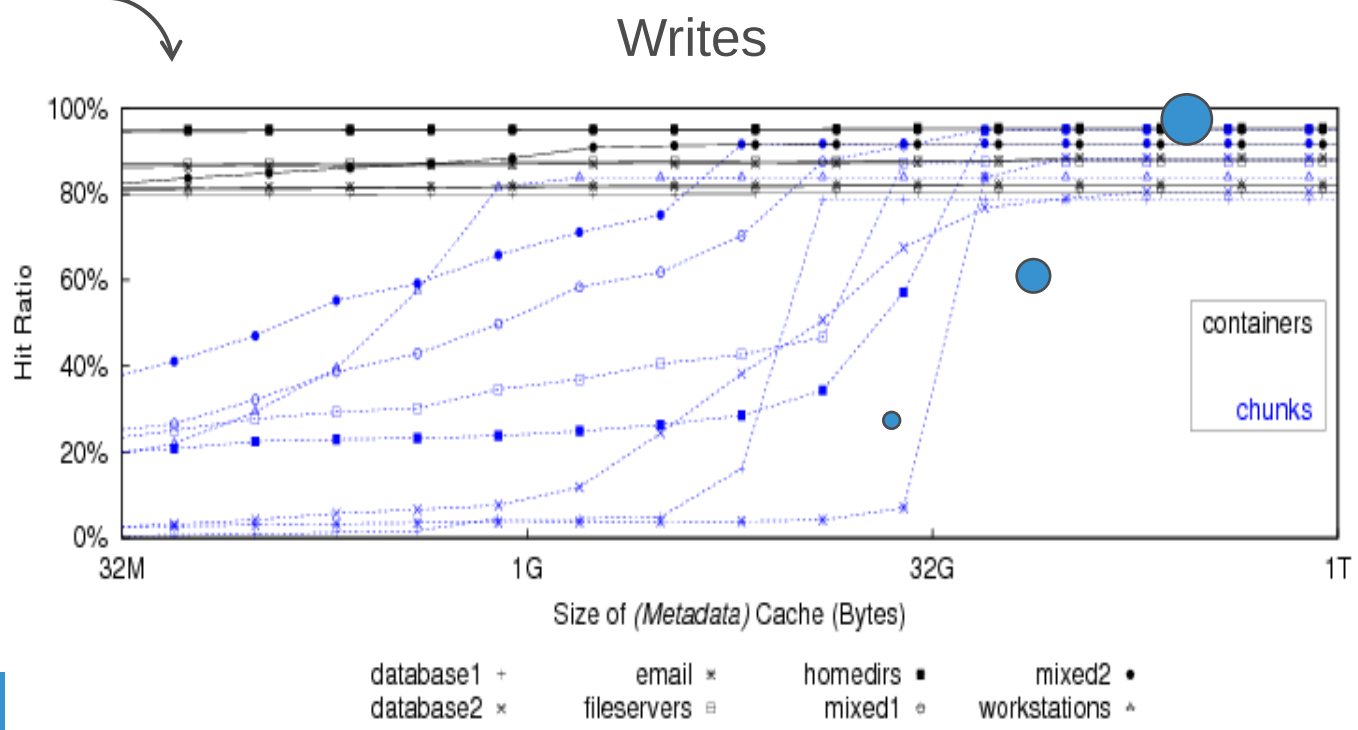
Caching

- Memory cache to avoid disk accesses
  - For writes, need to cache metadata so we know which chunks are duplicates
  - For reads, want to also cache the data
- Granularity (possibly using stream locality hints)
  - Chunks: if you access a chunk, keep its metadata (or data) around
  - Compression regions: for reads, keep chunks that are compressed together as a group
  - Containers: cache all chunks in a SISL container together [Zhu08]
- Methodology
  - Replay trace with varying cache sizes
  - Report on the last Nth of the data (warm cache)
    - Where N is the deduplication ratio, so it approximates one full backup

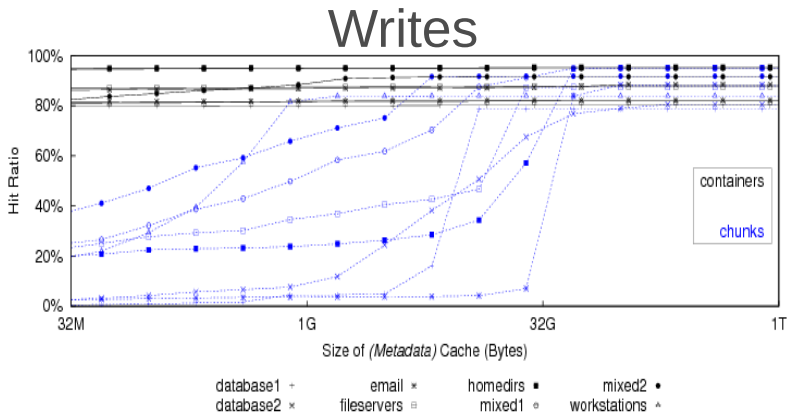
Caching Results

- Chunk-level LRU caching needs large cache to be effective for writes
  - Fit a full backup's metadata into cache
- Container-level LRU caching works well
  - Compulsory misses a function of deduplication rate

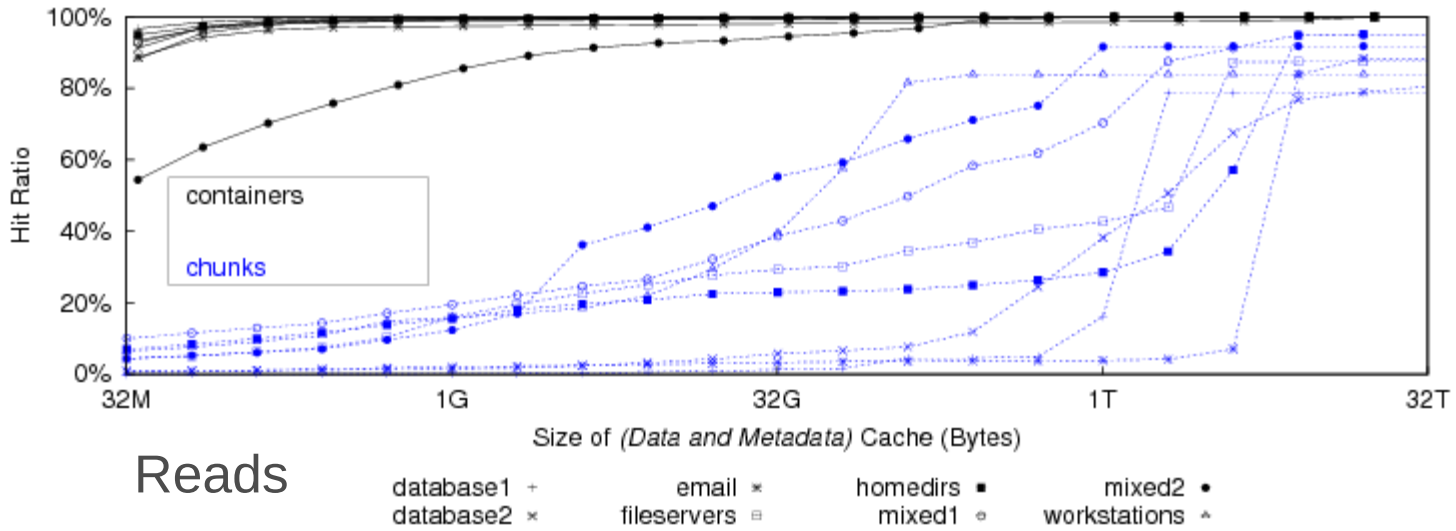
Sharp knee in some curves



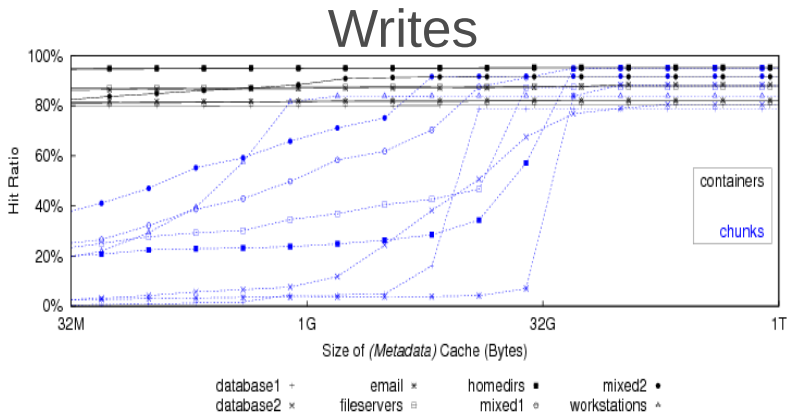
Caching Results



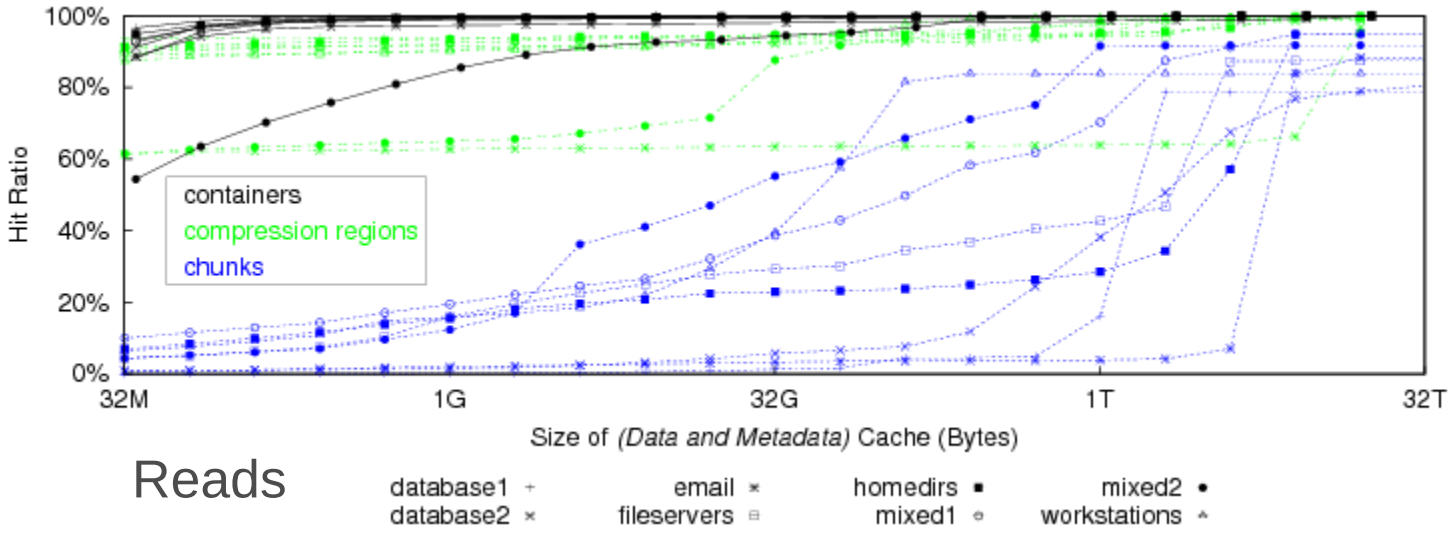
- Read cache behavior similar to writes but for much larger cache due to data caching
  - No compulsory misses beyond one access per container; fragmentation effects



Caching Results



- Read cache behavior similar to writes but for much larger cache due to data caching
  - No compulsory misses beyond one access per container; fragmentation effects
- Consider compression region caching
  - CRs usually close to container caching; some need very large caches



Related Work

- Deduplication
  - Windows 2000 (whole file), Venti (fixed blocks), many variable chunks including LBFS
  - Performance optimizations: SISL, Sparse Indexing, HydraSTOR, …
  - Bimodal Chunking for picking between two chunk sizes depending on deduplication effectiveness
- Data Characterization
  - Numerous primary storage studies, including Microsoft 2011 FAST study emphasizing deduplication
  - Univ. Minnesota backup deduplication characterization (limited datasets)

Conclusions

- High churn means throughput must scale with primary storage capacity growth
- Backup systems tend to have fewer, larger, and shorter-lived files than primary
- High locality and deduplication necessary for hit rates and high performance
- 8KB chunks are a “sweet spot” for backup deduplication

Backup != Primary

EMC²

Questions?

EMC²®