Depletable Storage Systems

Vijayan Prabhakaran
Mahesh Balakrishnan, John Davis, Ted Wobber

Microsoft Research, Silicon Valley

Depletable Storage Systems

- Traditional disk-based storage systems
  - Space is the primary resource constraint
  - E.g.: Quota on file servers, pricing model in cloud
- SSD-based storage systems
  - Also, limited by number of erasures
  - Space and write cycles
- Depletable storage systems
  - Limited lifetime
  - Measureable, predictable, relate to workload

Write-lifetime of an SSD

- Write-lifetime
  - Amount of data written during SSD's lifetime
  - Ideal: size x maximum erase cycles
  - E.g.: 80 GB x 5000 cycles = 400 TB

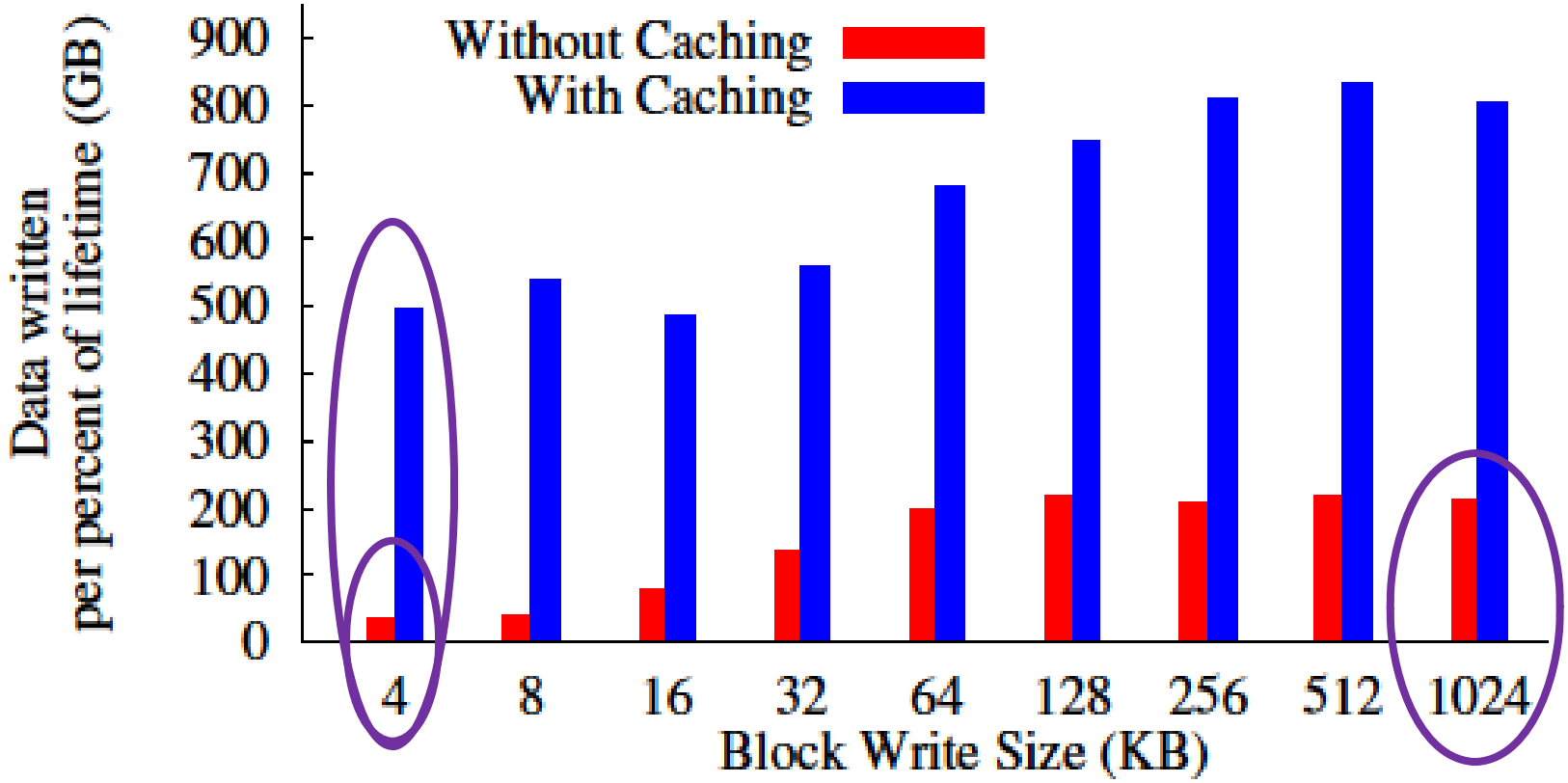
- Write-lifetime in practice
  - Affected by firmware inefficiencies
  - Write amplification
  - Cleaning and wear-leveling

Write-lifetime Metrics

- SSD manufacturers address write-lifetime
- SanDisk's Longterm Data Endurance (LDE)
  - Includes write amplification and wear-leveling
- Intel's media wearout indicator
  - Percentage of lifetime left
  - "Decreases from 100 to 1 as average erase cycles used increases to the rated maximum"

Write-lifetime Measurements

- Simple experiment
  - Create a 70 GB file in Intel X25-M
  - Write certain size to a random offset



Depletion-Aware Functionalities

- Predictable device replacement based on lifetime
  - E.g., proactive RAID reconstruction
- New pricing model
  - Charge users based on writes as well
- New axis for comparison
  - Compare designs that reduce depletion
- New attack models
  - Depletion of lifetime attack

New Mechanisms and Algorithms

- Mechanisms
  - Track the writes
  - Attribute writes to appropriate applications
  - Control the writes
- Depletion-aware resource management
  - New scheduling algorithms

Challenges

- Layers in software stack
  - VFS, caching, journaling, I/O schedulers, volume manager, software/hardware RAID, device
  - Impact writes: Delay, Amplify, Reduce
- Media heterogeneity
  - MLC or SLC
  - Different price-performance and erasure limits

Possible Solutions

- Track, Attribute, and Control
  - VM to isolate applications and their writes
  - Cloud already uses VM for isolation
- Beneath the VM
  - Minimize layers before issuing to SSD
  - Scheduling: allocate time quanta per VM [Argon]
  - May provide depletion isolation
- Heterogeneity: write credit
  - Ideal write-lifetime / price
  - Intel X25-M: (80GB x 5K erasures) / $220 = 1.78 TB per dollar
  - Intel X25-E: (64GB x 100K erasures)/ $745 = 8.59 TB per dollar

Conclusion

- SSD Focus: performance, reliability, lifetime
- Propose to treat SSDs as depletable storage systems
- Need new mechanisms and algorithms to enable new functionalities