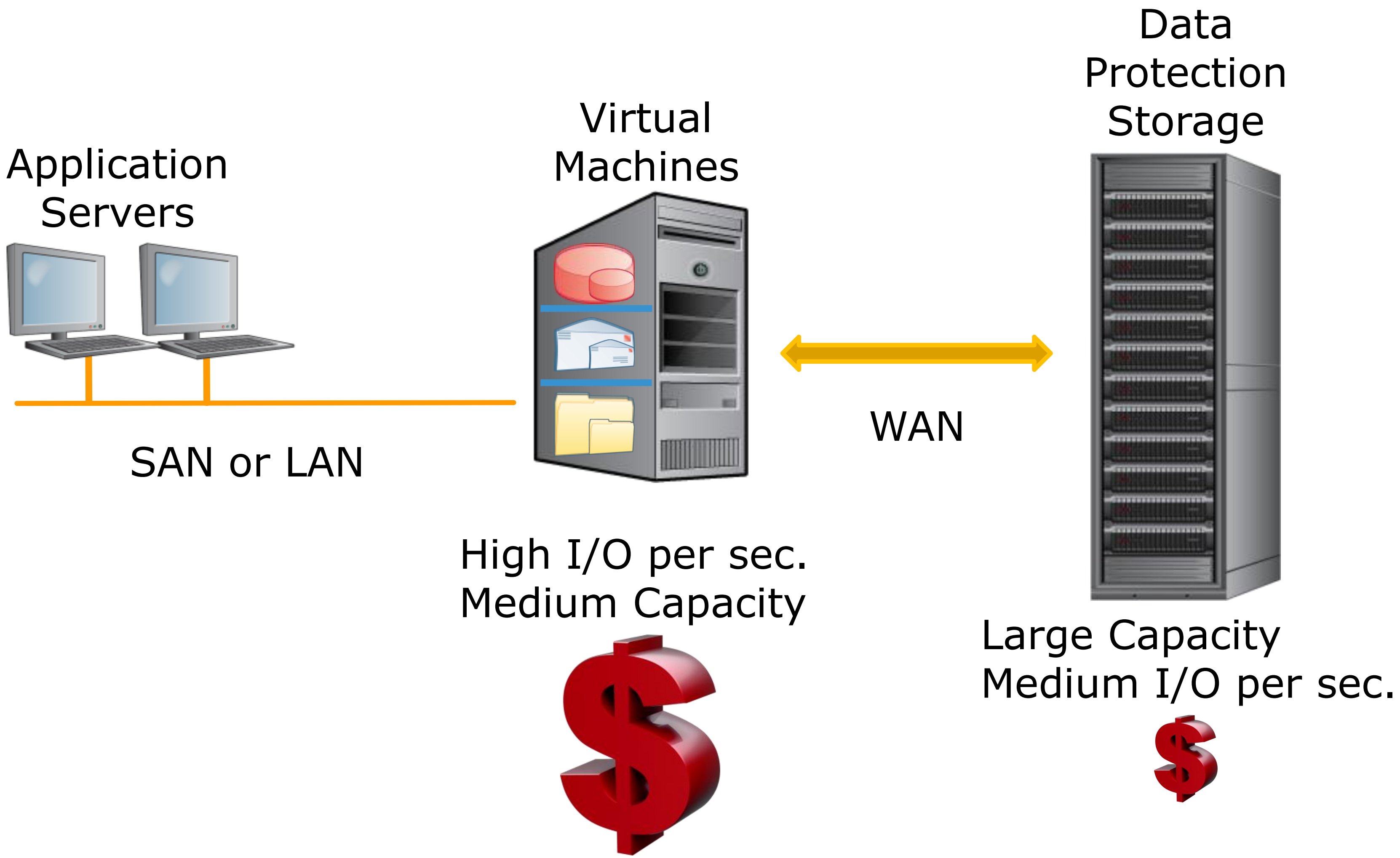
Characterization of Incremental Data Changes for Efficient Data Protection

Hyong Shim, Philip Shilane, & Windsor Hsu

Backup Recovery Systems Division
EMC Corporation

EMC²

Data Protection Environment



Contributions

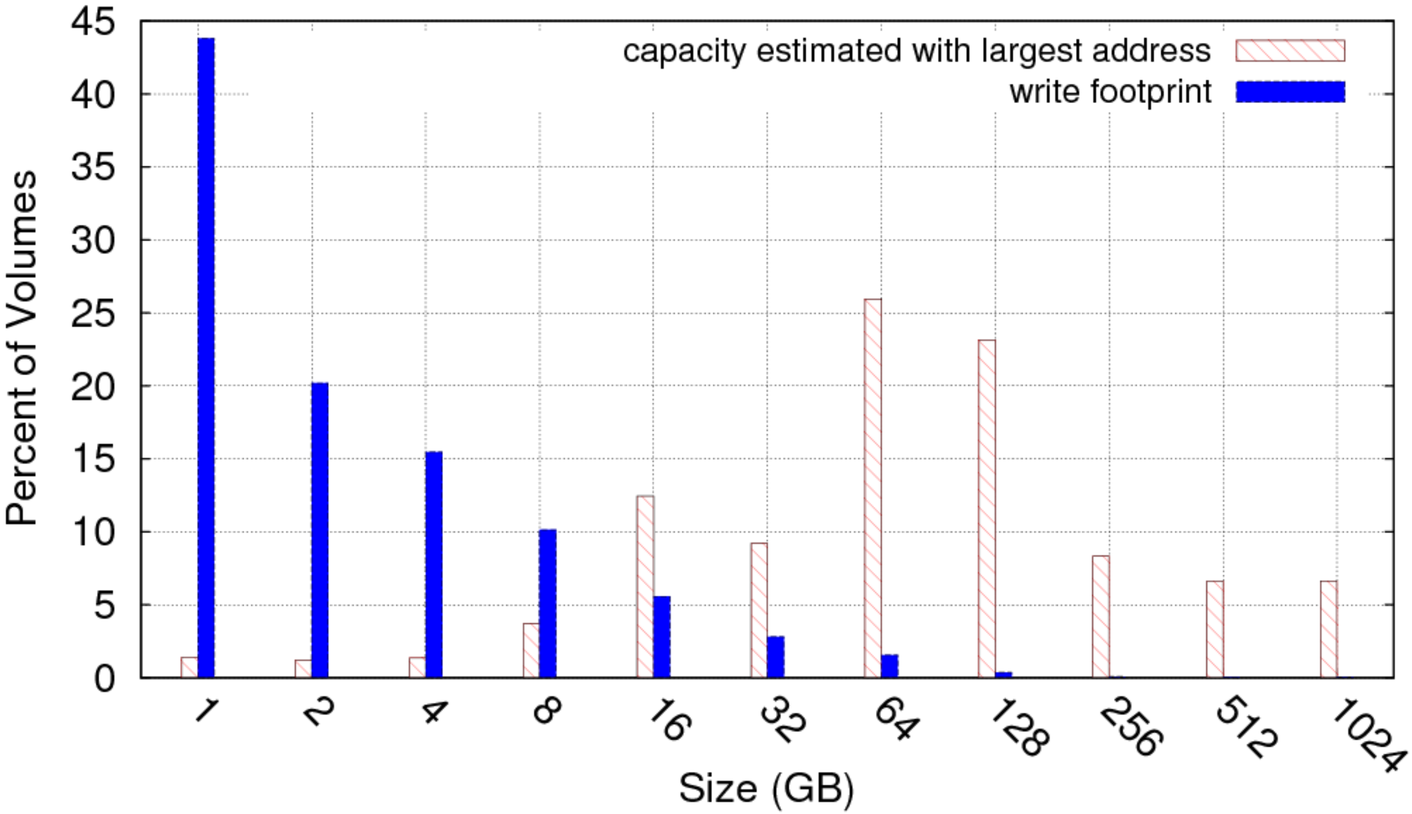
- Detailed analysis of data change characteristics from enterprise customers
- Design for replication snapshots to lower overheads on primary storage.
- Evaluation of overheads on data protection storage
- Rules-of-thumb for storage engineers and administrators

EMC Symmetrix VMAX Traces

- Collected from enterprise customer sites

Trace Set	#Volume	# Storage Systems	Duration hrs	Estimated Capacity (GB)
1hr_1Wrt	109,263	125	30.4 [78.3]	71 [203]
1hr_1GBWrt	16,100	120	7.7 [6.7]	132 [262]
24hr_1GBWrt	508	13	24.4 [1.2]	318 [439]

Capacity and Write Footprint



- Analysis for 1hr_1GBWrit
- Not collected: applications using each volume

EMC²

I/O Properties

Trace Set	#Write reqs (1000s)	Write size (GB)	#Read reqs (1000s)	Read size (GB)
1hr_1Wrt	72 [510]	2 [31]	167 [1963]	5 [66]
1hr_1GBWrt	429 [1270]	11 [80]	796 [4987]	25 [166]
24hr_1GBWrt	1803 [4839]	51 [338]	7824 [23875]	242 [763]

- 1.9-4.3X more read I/Os than write I/Os
- 2.3-4.7X more GB read than written
- High variability
- More analysis in the paper

Sequential vs. Random Write I/O

Trace Timeline (w = Write I/O, r = Read I/O)

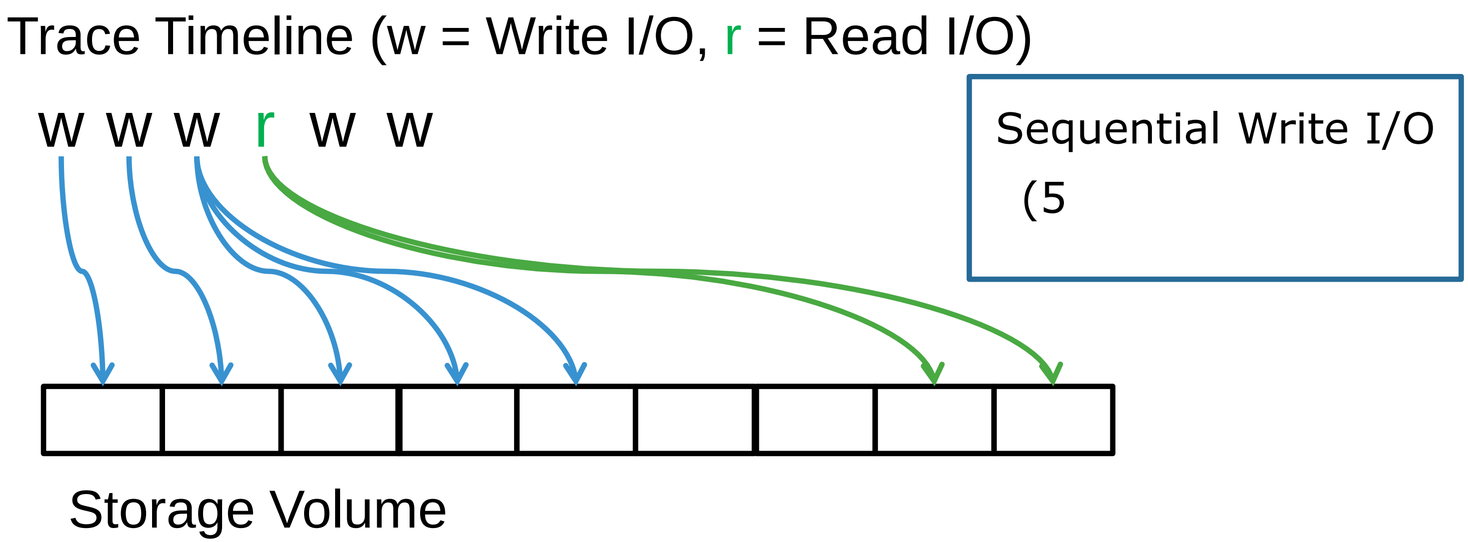
w w w r w w

Storage Volume

- We measure how much data are written, on average, after seeking to a non-consecutive sector.
- Selected most sequential and most random for analysis

EMC²

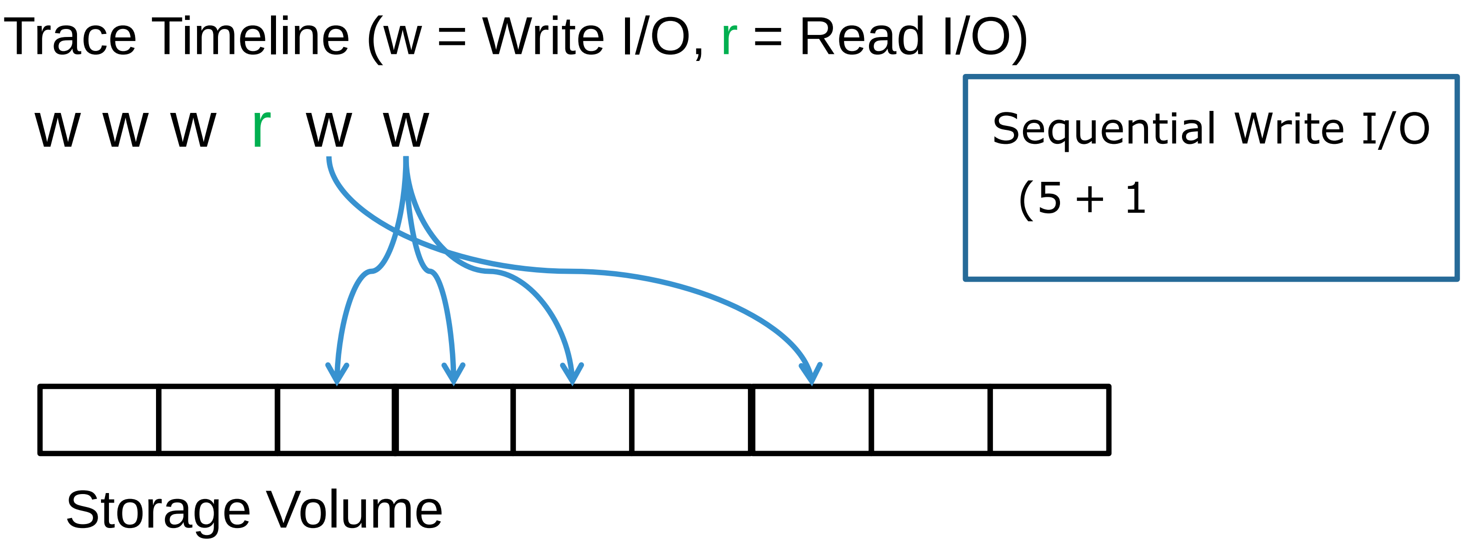
Sequential vs. Random Write I/O



- We measure how much data are written, on average, after seeking to a non-consecutive sector.
- Selected most sequential and most random for analysis

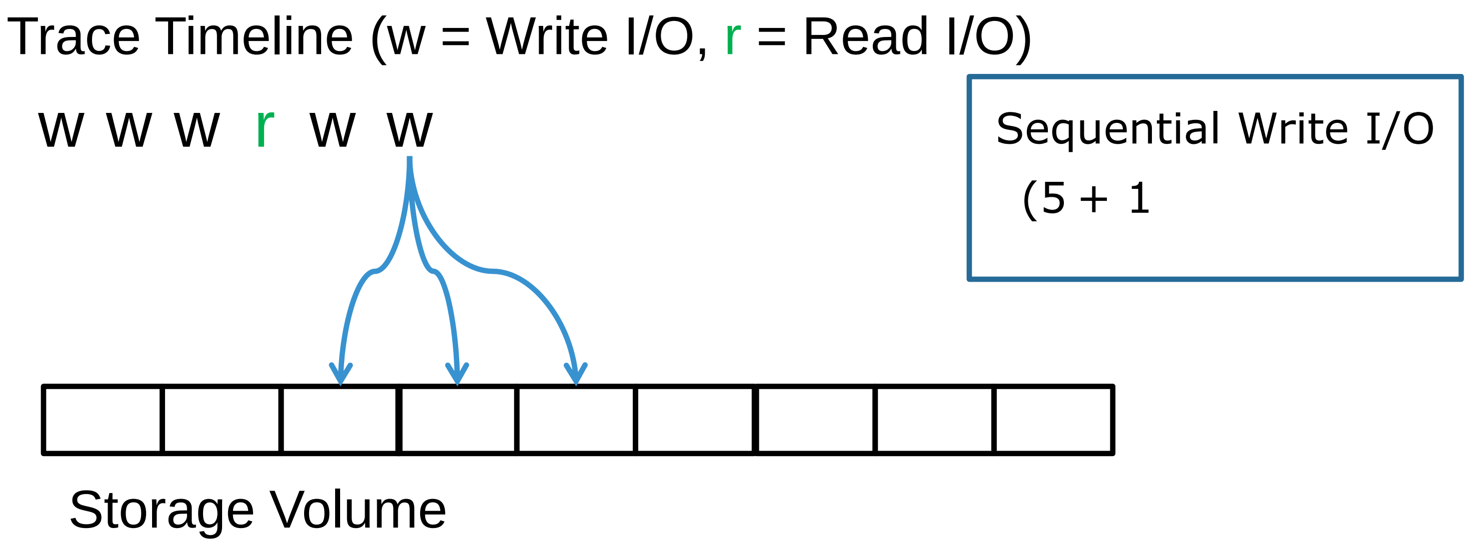
EMC²

Sequential vs. Random Write I/O



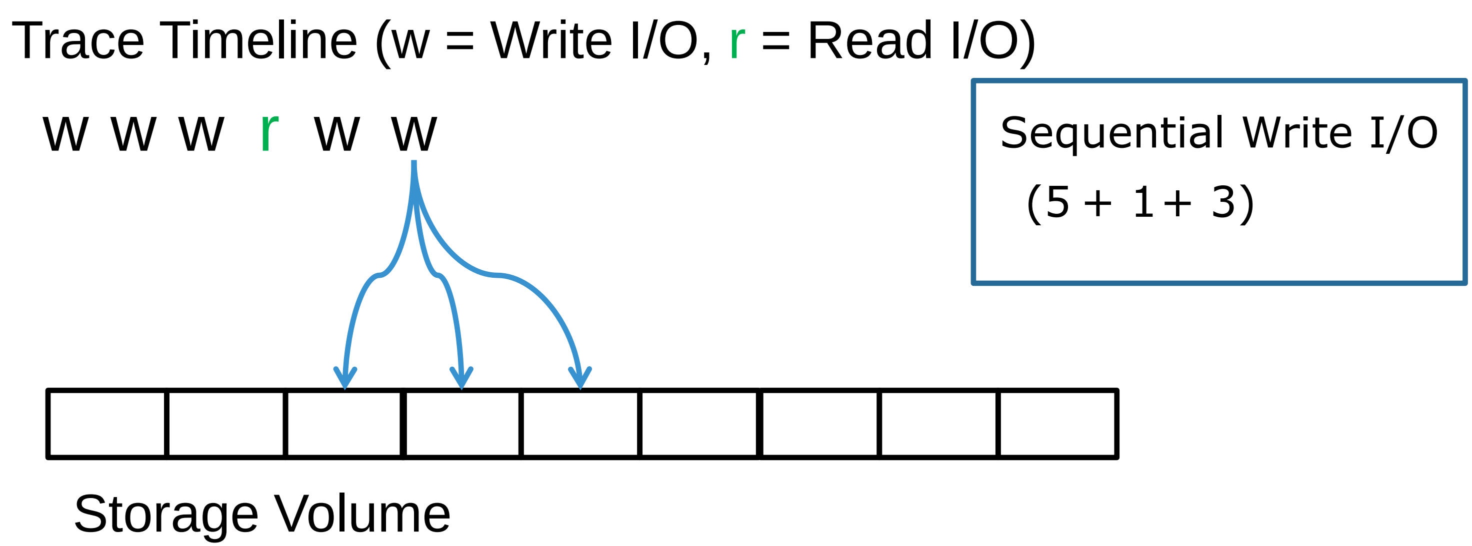
- We measure how much data are written, on average, after seeking to a non-consecutive sector.
- Selected most sequential and most random for analysis

Sequential vs. Random Write I/O



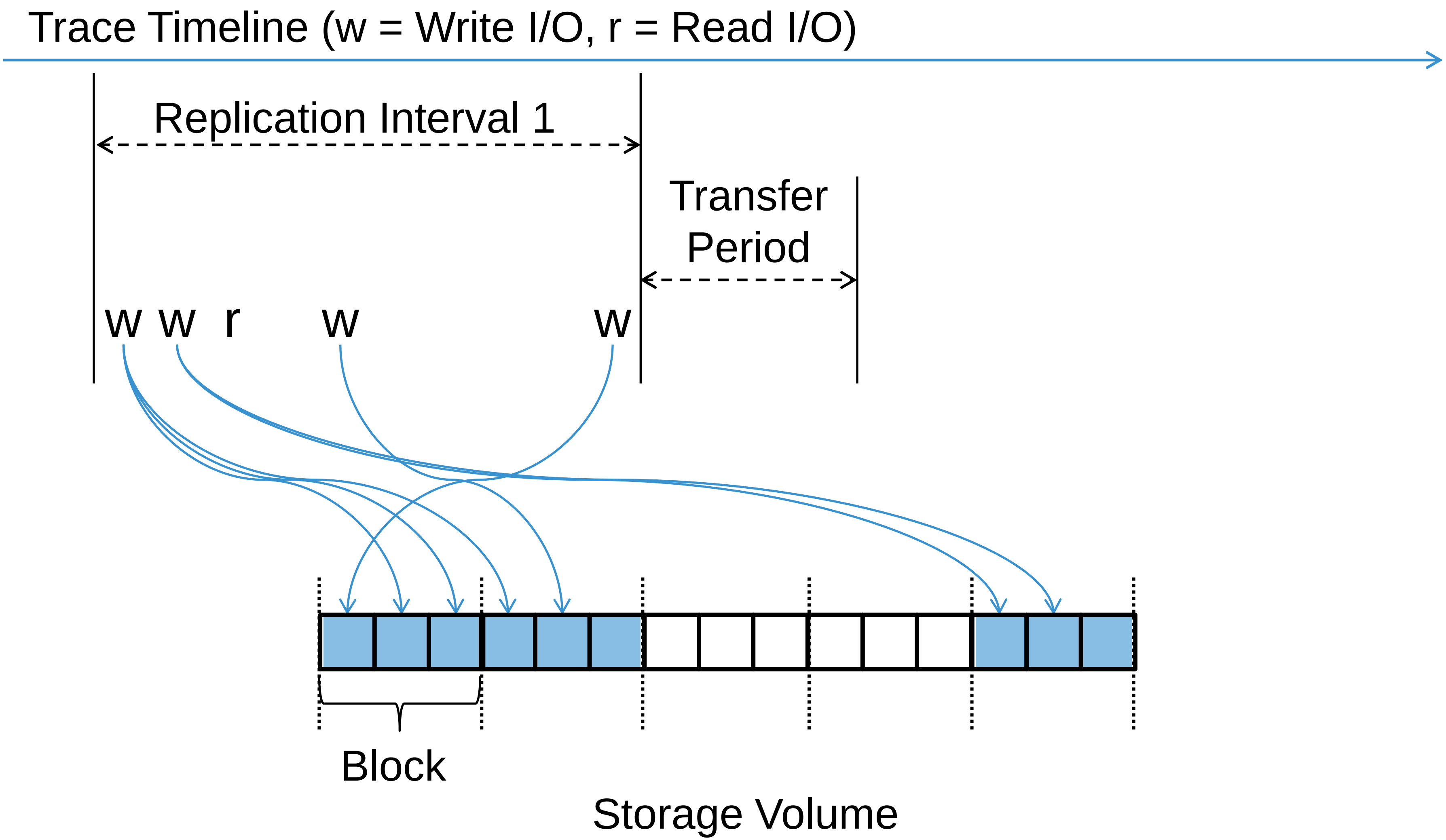
- We measure how much data are written, on average, after seeking to a non-consecutive sector.
- Selected most sequential and most random for analysis

Sequential vs. Random Write I/O

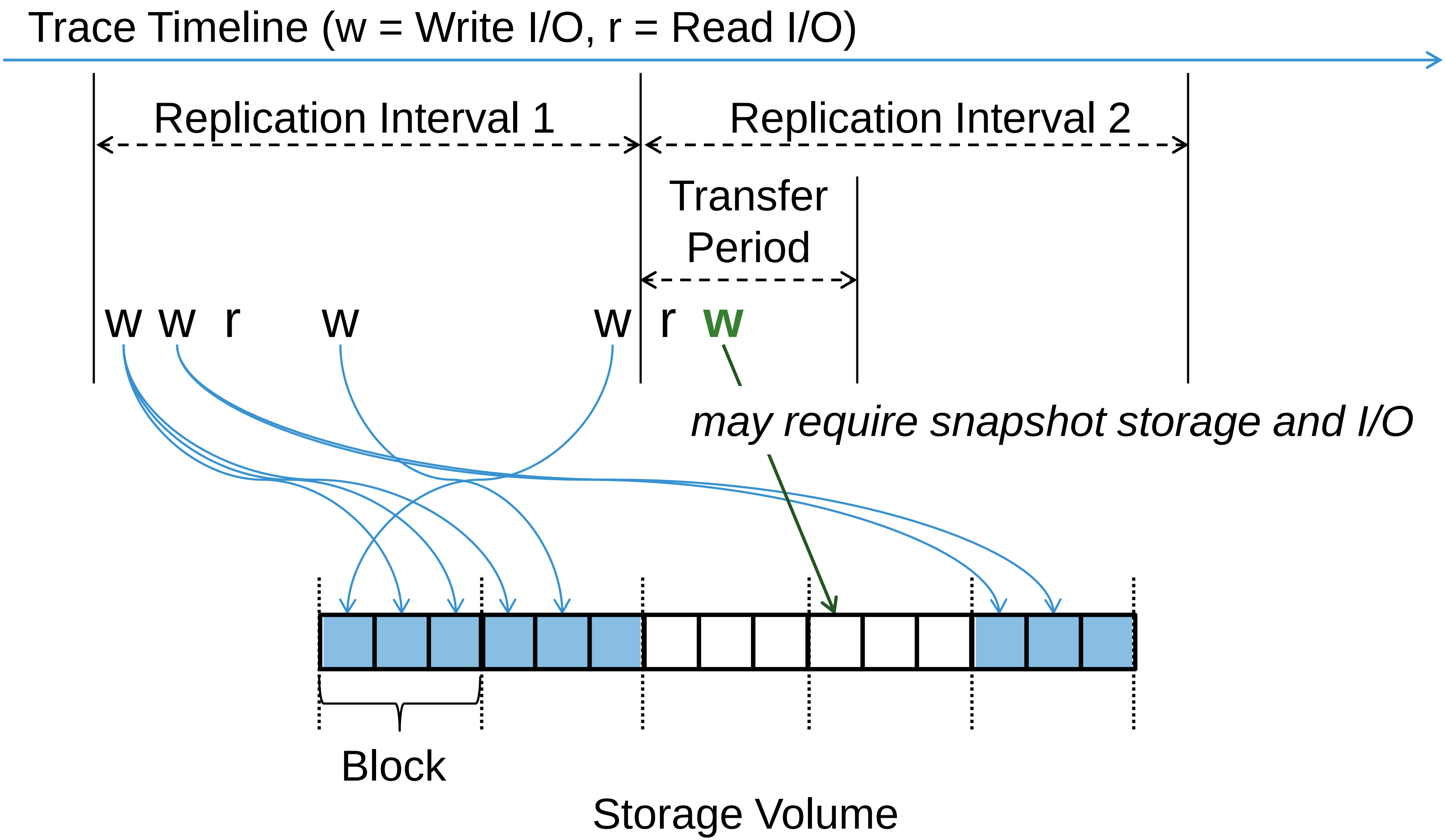


- We measure how much data are written, on average, after seeking to a non-consecutive sector.
- Selected most sequential and most random for analysis

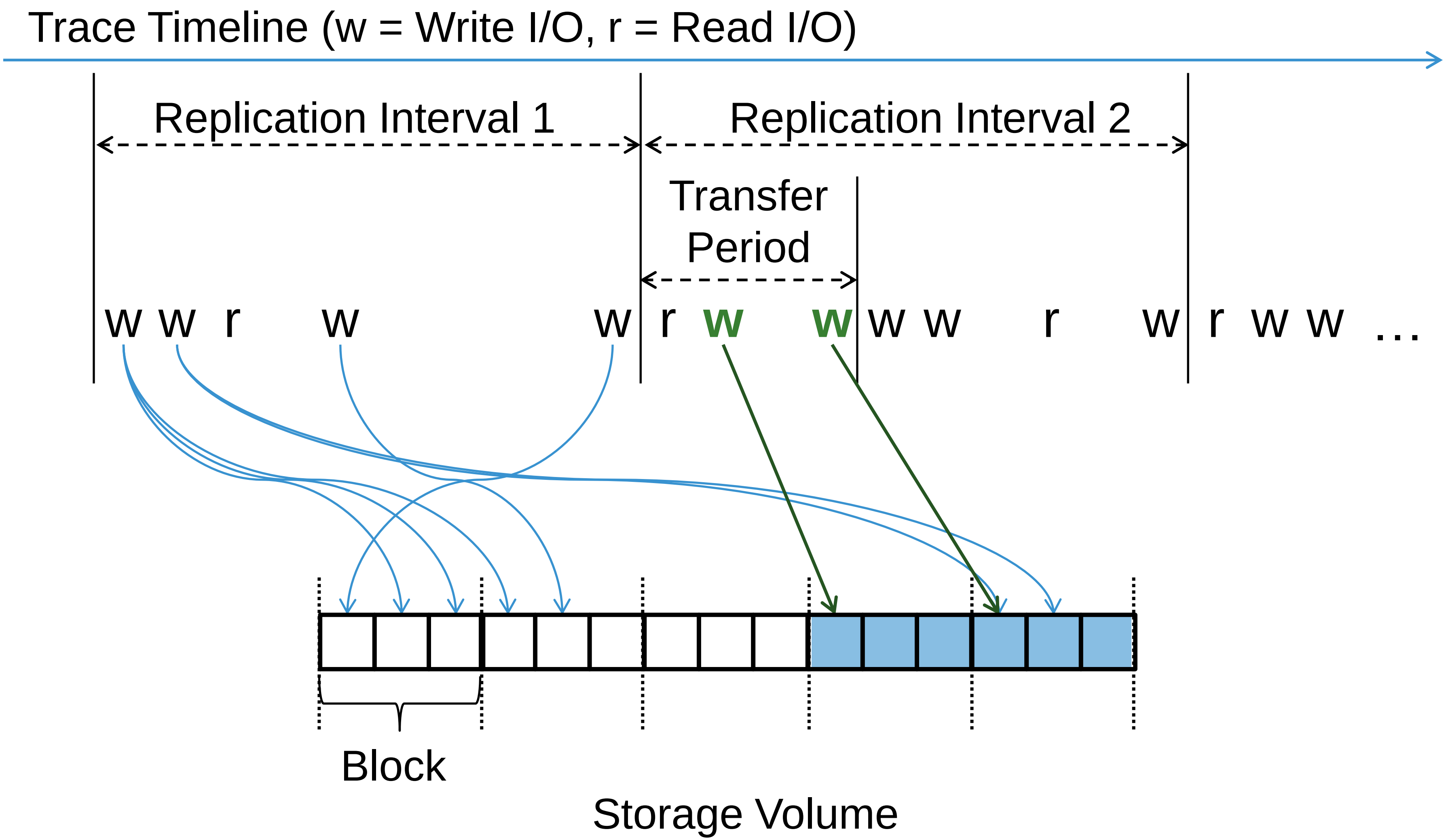
Trace Analysis Methodology



Trace Analysis Methodology

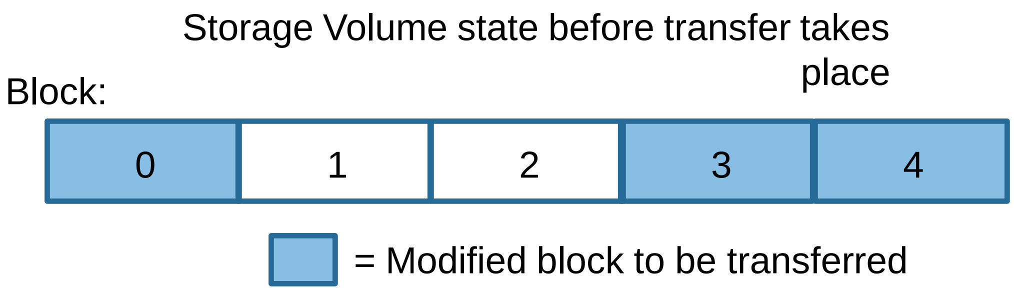


Trace Analysis Methodology



Replication Snapshot

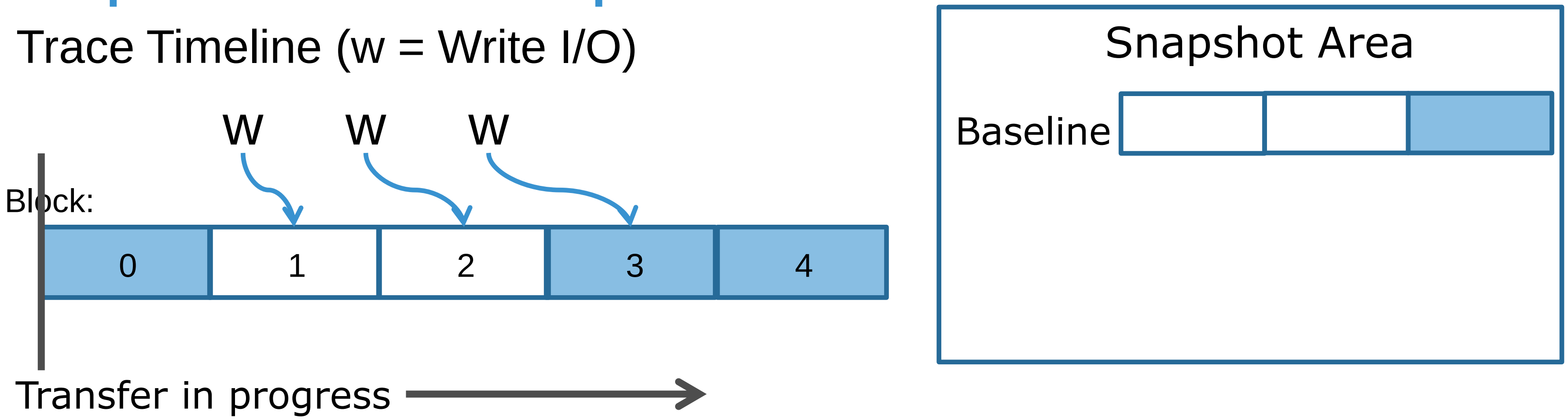
Trace Timeline (w = Write I/O)



- **Goal:** Create a snapshot technique that is integrated with replication that decreases overheads on primary storage
- **Change block tracking** records modified blocks for next replication interval, possibly with a bit vector.
- A **snapshot** has to maintain block values against overwrites.

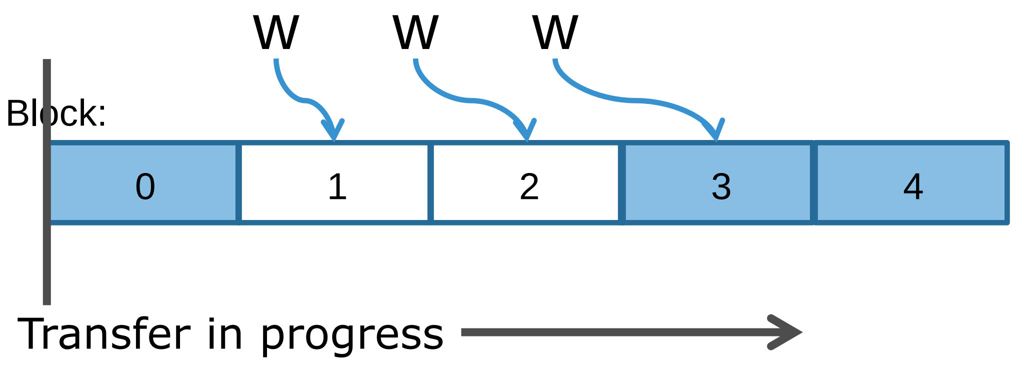
EMC²

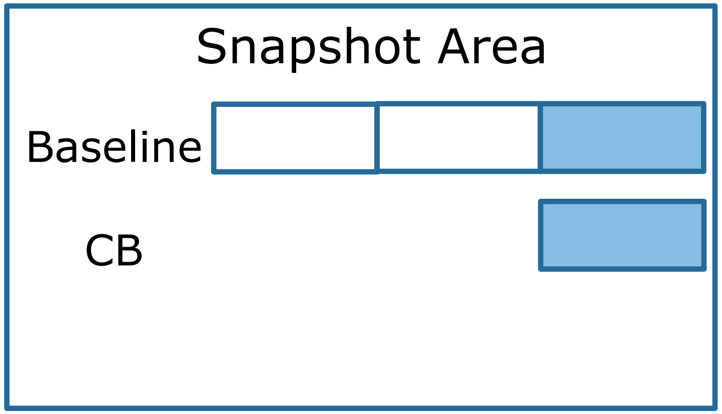
Replication Snapshot



- **Baseline Snapshot:** All writes cause copy-on-write

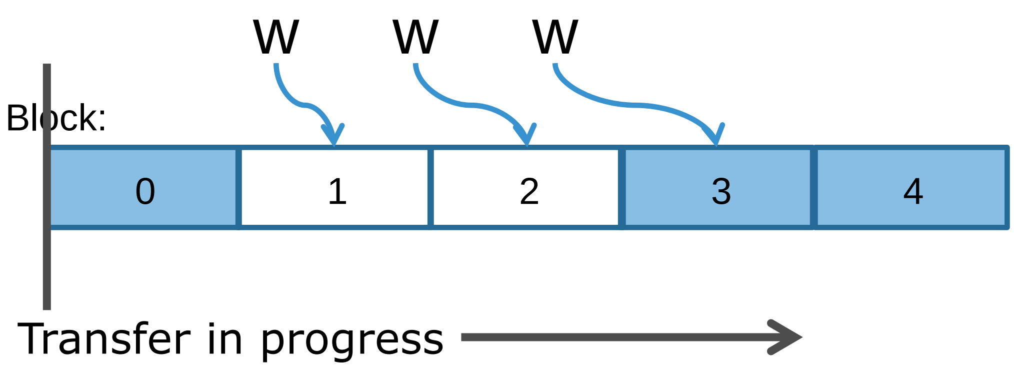
Replication Snapshot

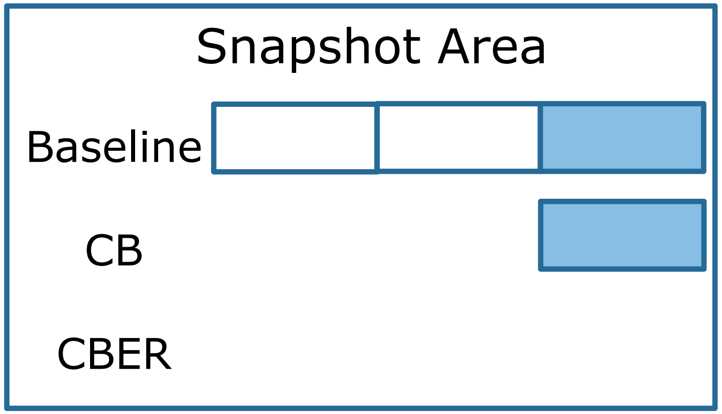




- **Changed Block Replication Snapshot (CB):** Only writes to tracked blocks cause copy-on-write

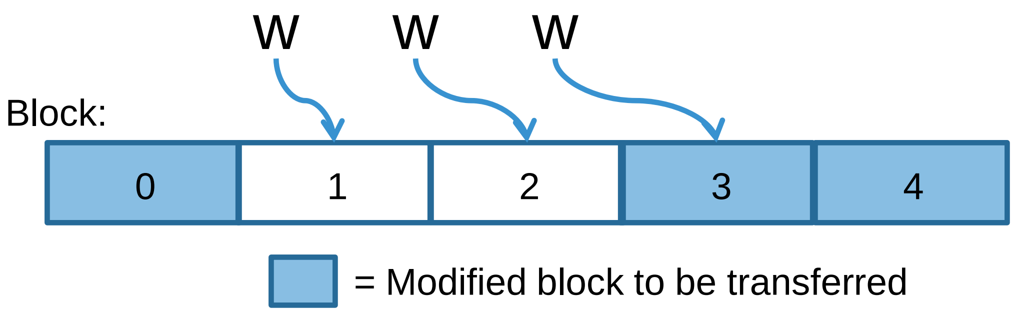
Replication Snapshot

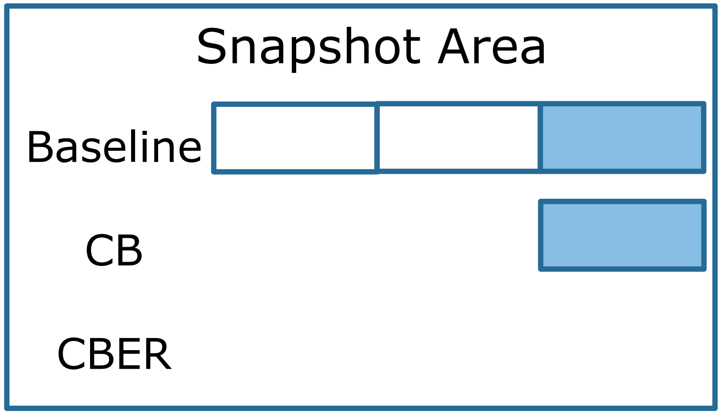




- **Changed Block with Early Release Replication Snapshot (CBER):** Only writes to tracked blocks cause copy-on-write, and blocks are released once transferred

Replication Snapshot

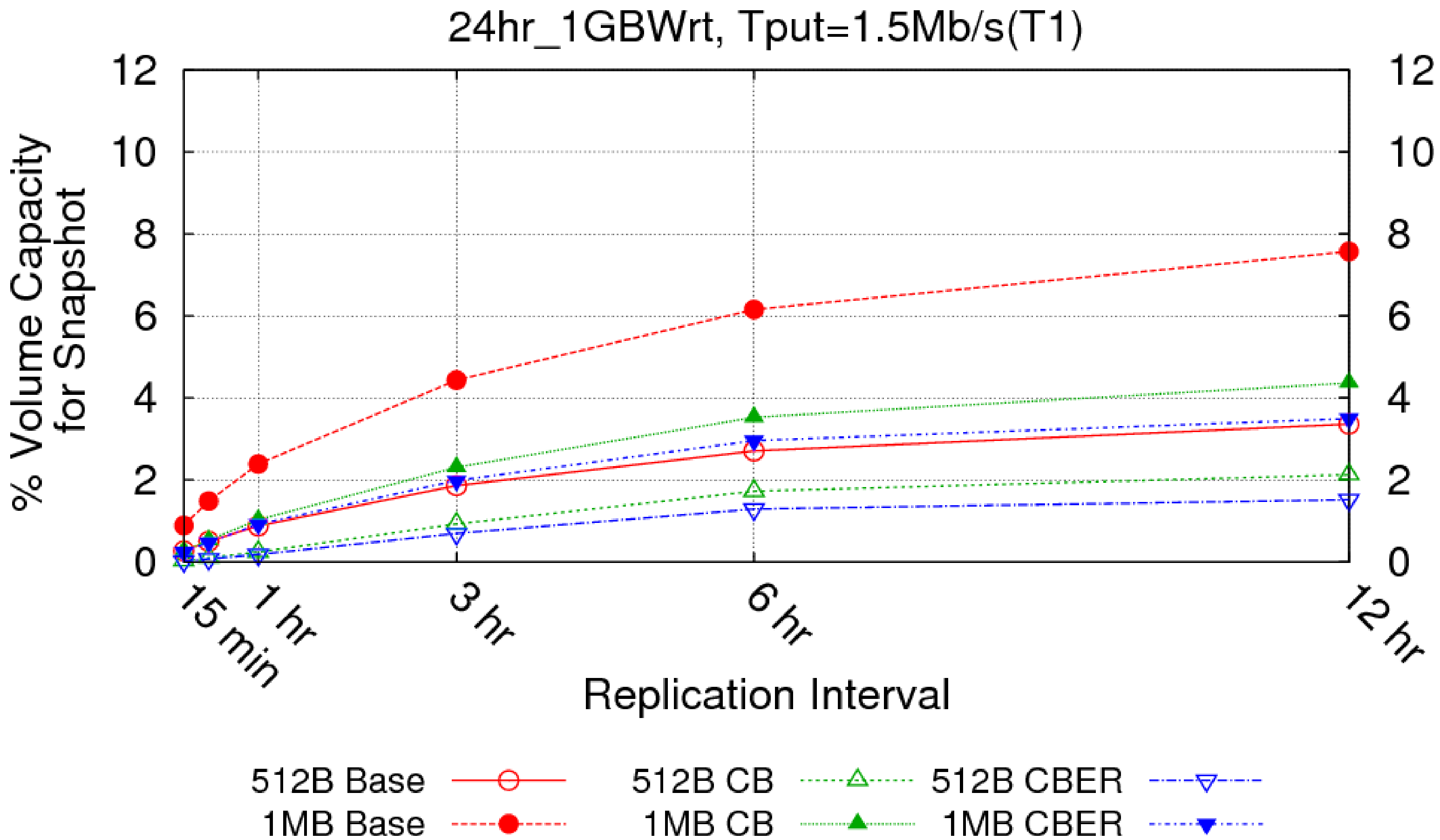




- **Baseline Snapshot:** All writes cause copy-on-write
- **Changed Block Replication Snapshot (CB):** Only writes to tracked blocks cause copy-on-write
- **Changed Block with Early Release Replication Snapshot (CBER):** Only writes to tracked blocks cause copy-on-write, and blocks are released once transferred

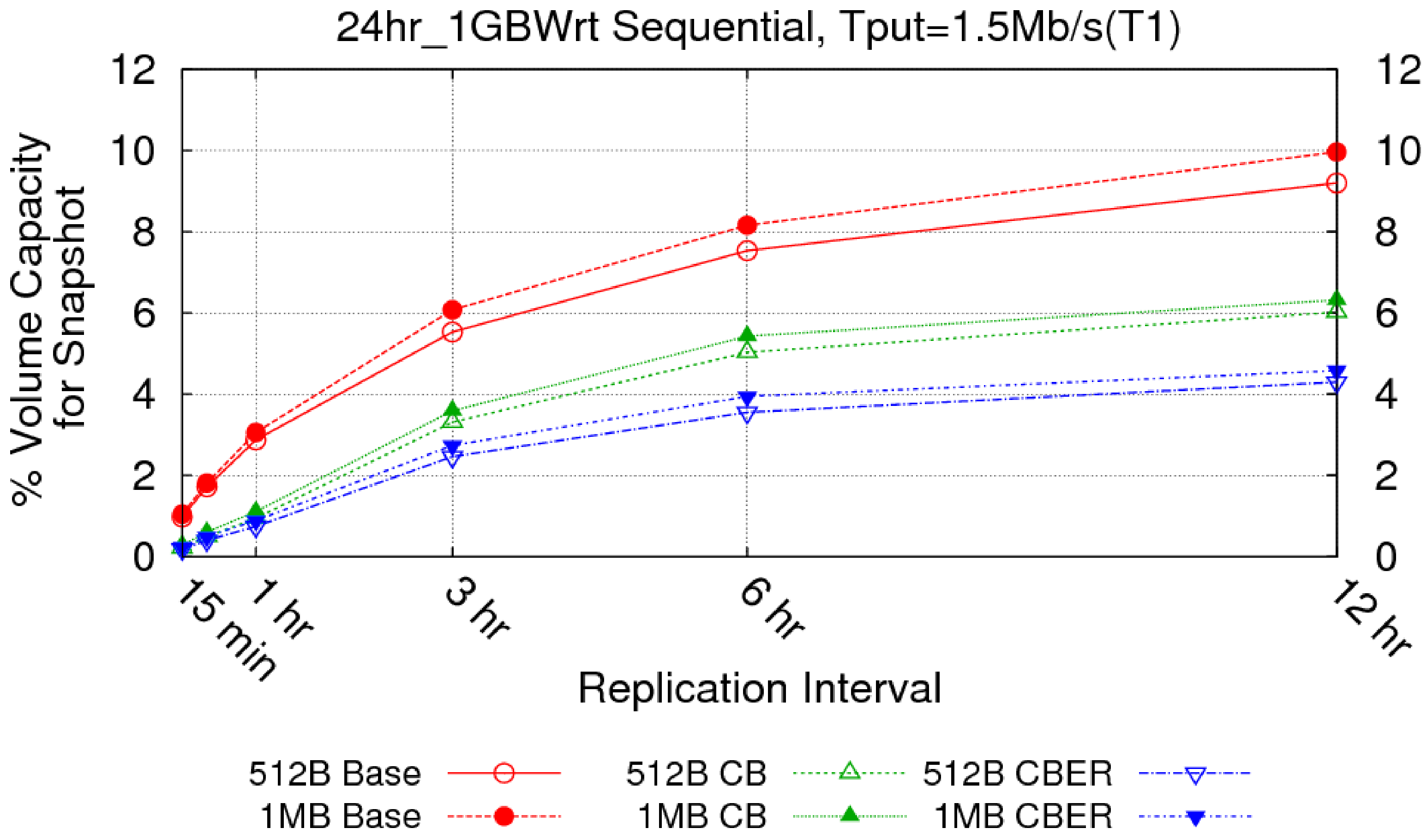
Snapshot Storage Overheads

Rule-of-thumb: Over-provision primary capacity by 8% for snapshots



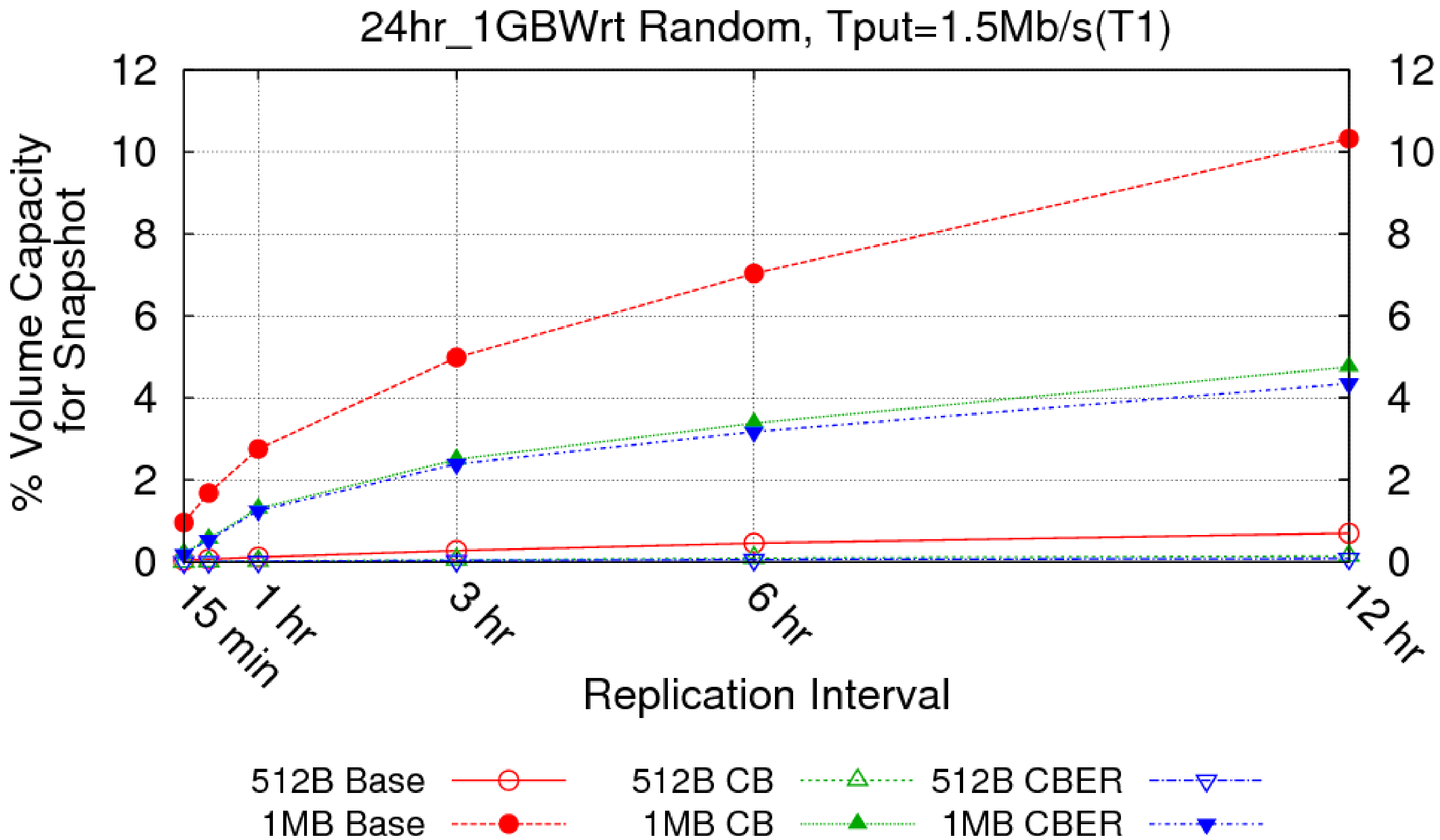
Snapshot Storage Overheads

Rule-of-thumb: Over-provision primary capacity by 8% for snapshots



Snapshot Storage Overheads

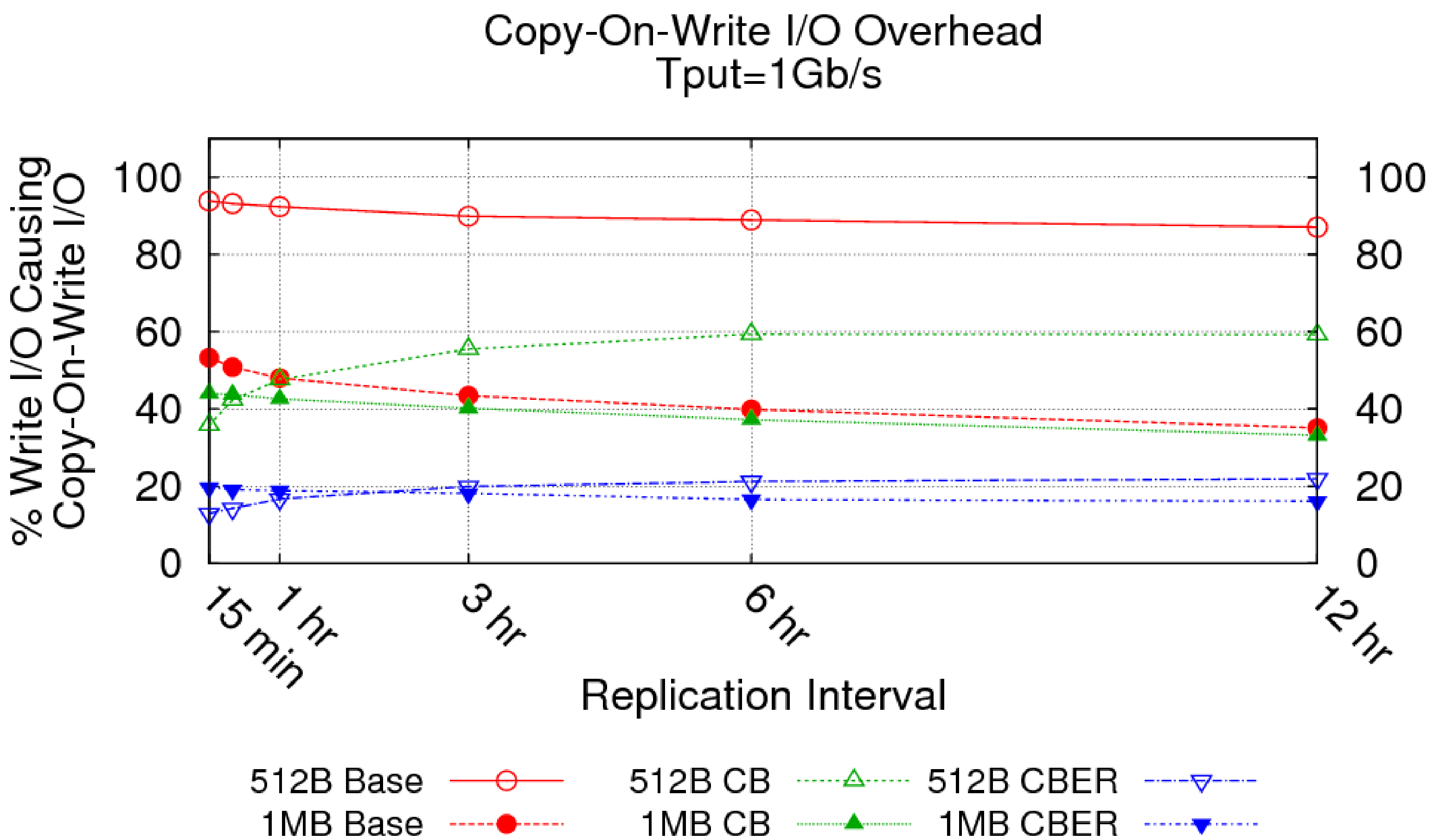
Rule-of-thumb: Over-provision primary capacity by 8% for snapshots



EMC2

Snapshot I/O Overheads

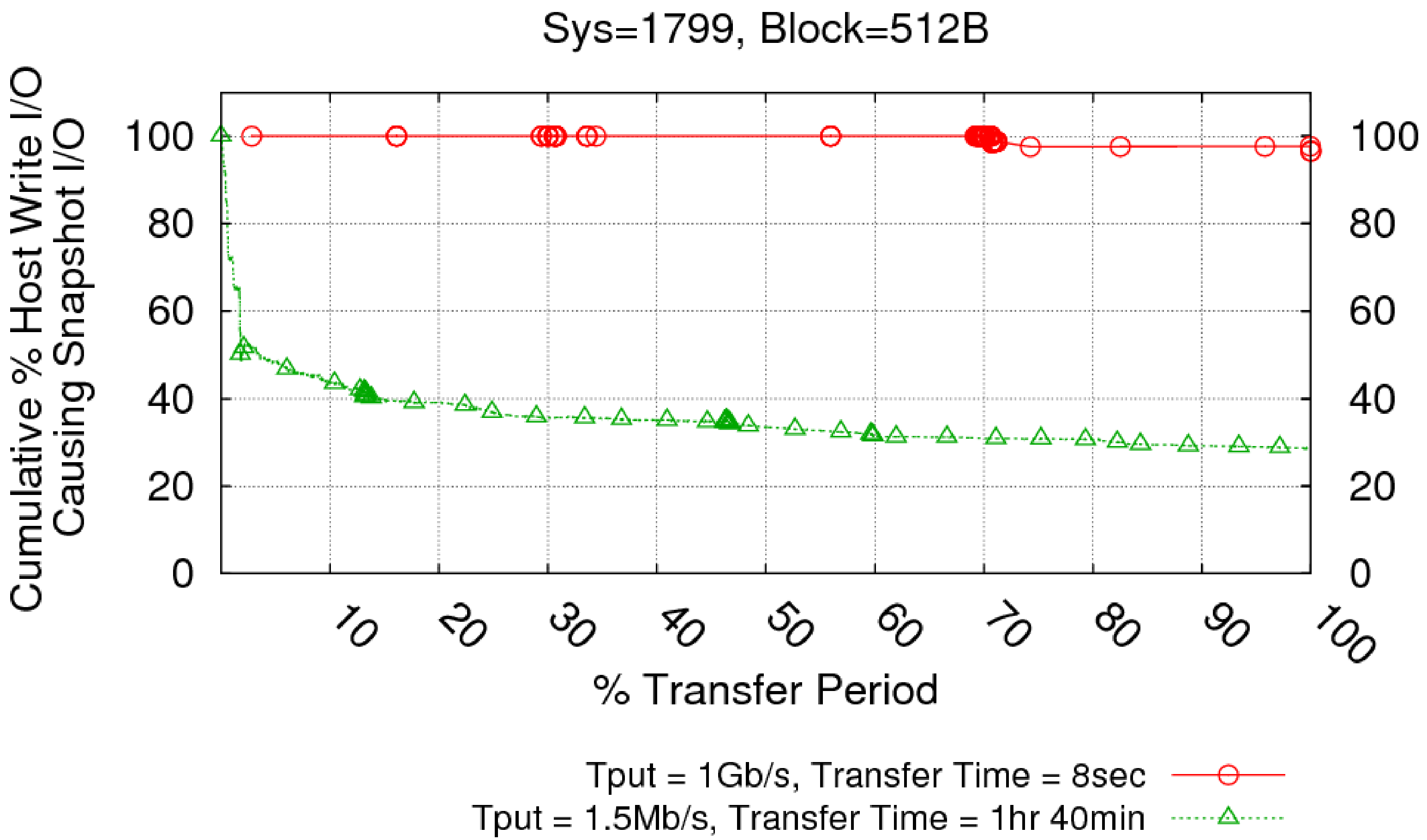
Rule-of-thumb: Over-provision primary I/O by 100% to support copy-on-write related write-amplification



EMC²

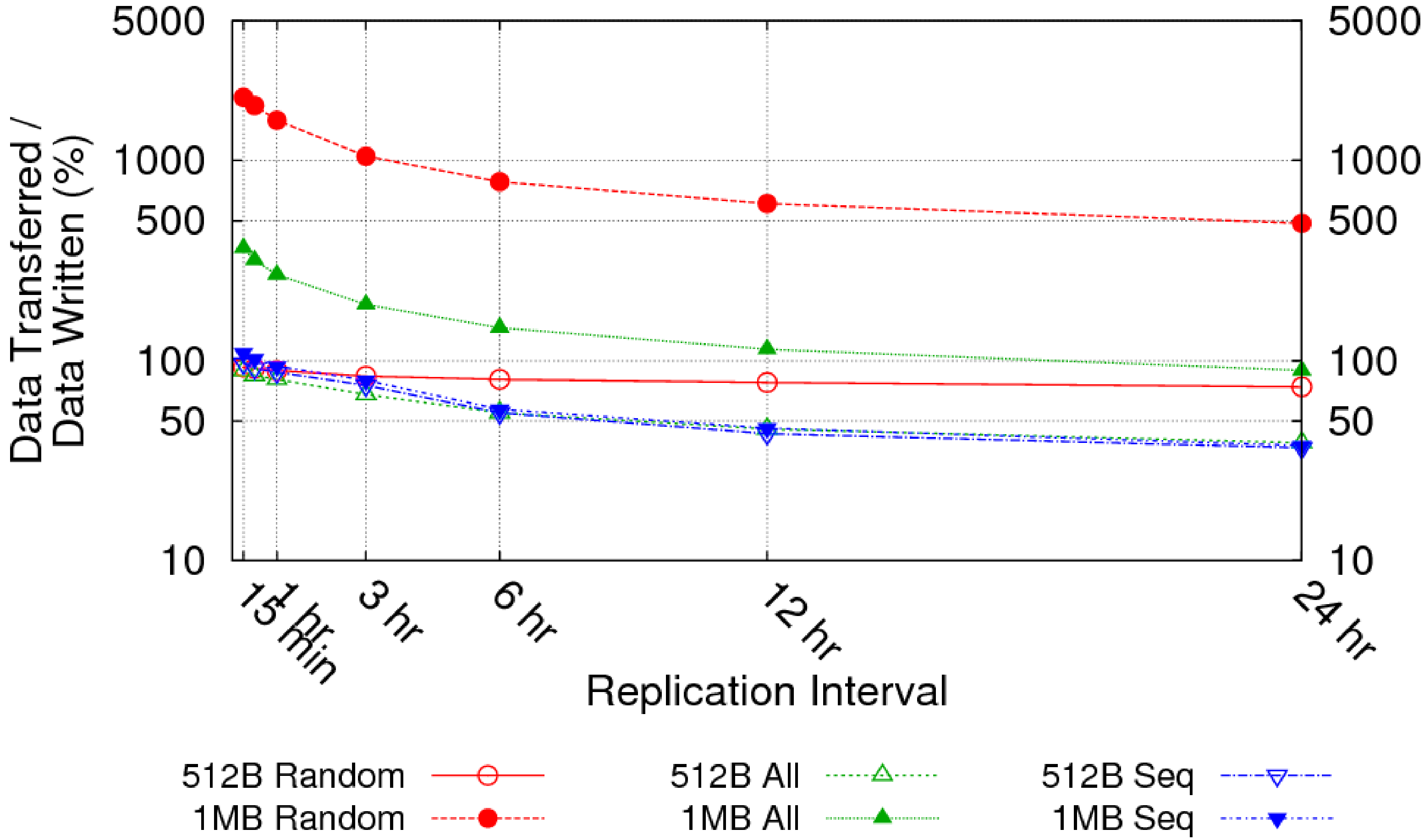
Snapshot I/O Overheads

Rule-of-thumb: Over-provision primary I/O by 100% to support copy-on-write related write-amplification



Transfer Size to Protection Storage

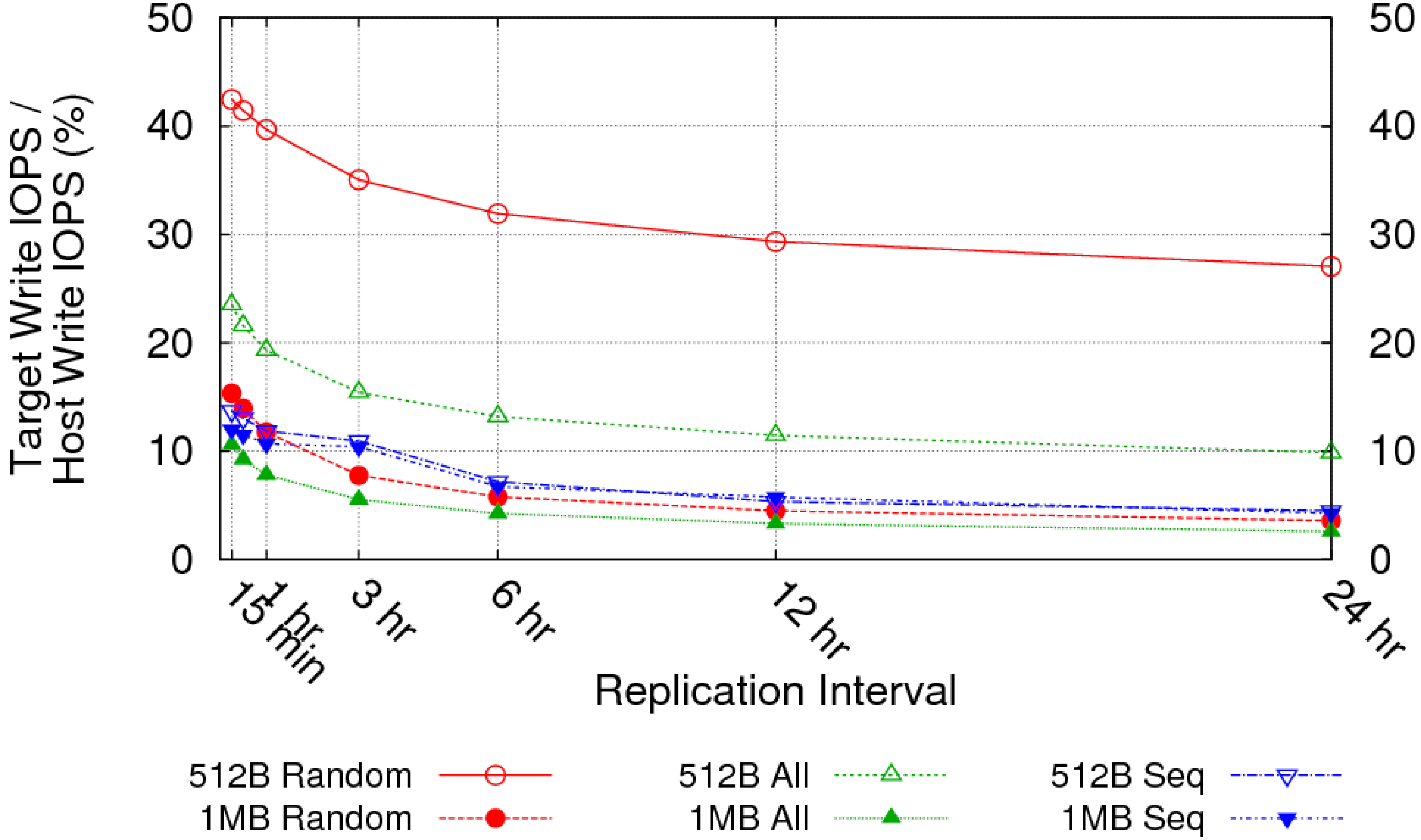
Rule-of-thumb: 40% of written bytes are transferred to protection storage



EMC²

IOPS Requirements for Protection Storage

Rule-of-thumb: Protection storage must support 20% of the I/O per second capabilities of primary storage



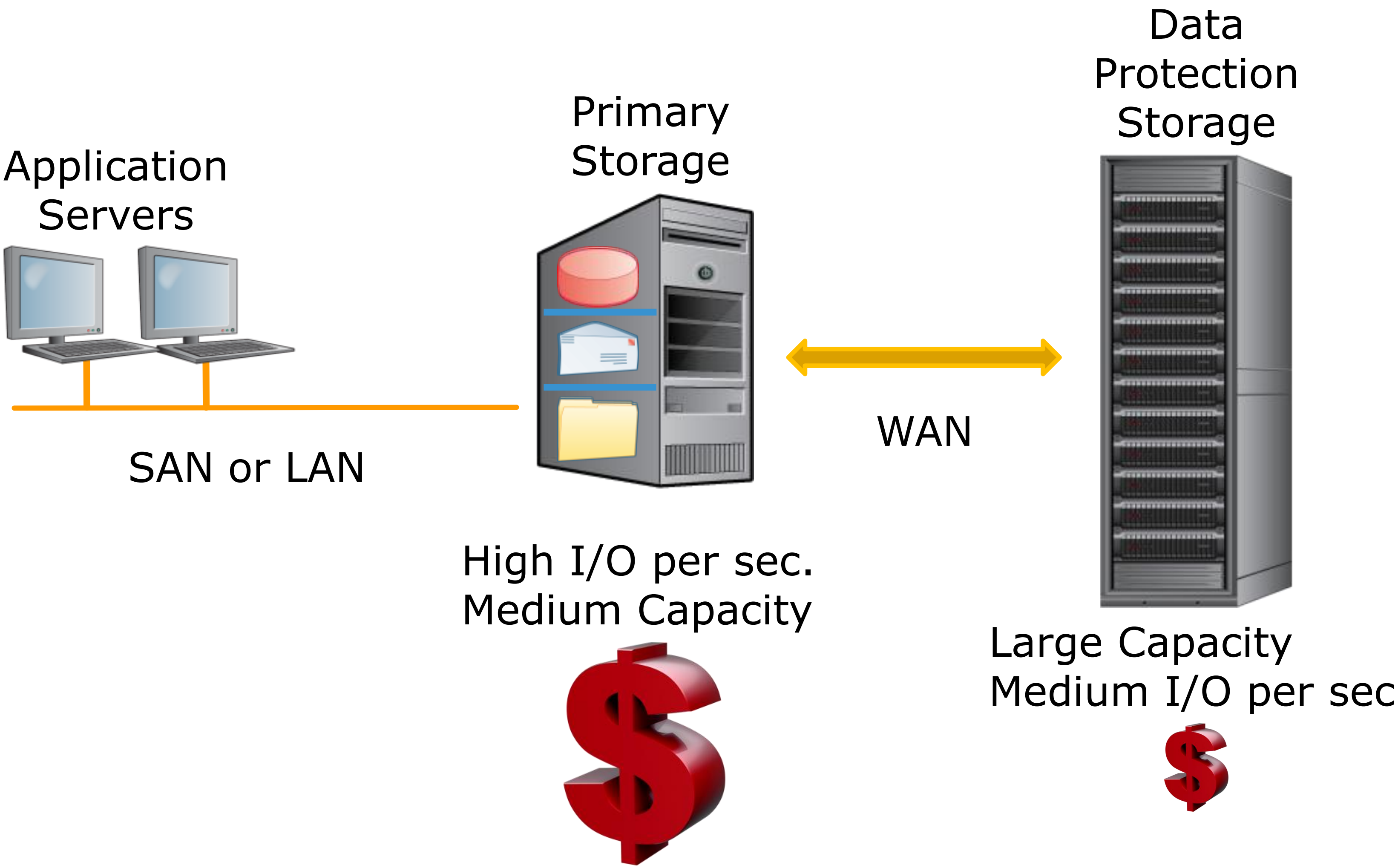
EMC²

Related Work

- Trace analysis
  - Numerous publications
    Most closely related is Patterson [2002]
- Snapshots
  - Common paradigm for storage but rarely integrated with incremental transfer techniques
  - Storage overheads Azagury [2002] and Shah [2006]
- Synchronous Mirroring
  - Effective when change rates are low and geographic distance is small
  - We are focused on periodic, asynchronous replication

EMC²

Conclusion


Application Servers
SAN or LAN
Primary Storage
High I/O per sec.
Medium Capacity
WAN
Data Protection Storage
Large Capacity
Medium I/O per sec.

Conclusion

- Trace analysis shows diversity of storage characteristics
- Snapshot overheads on primary storage can be decreased by improved integration with network transfer
- Sequential versus random access patterns affect incremental change patterns on both primary and protection storage

Rules-of-Thumb

- Over-provision primary capacity by 8% for snapshots
- Over-provision primary I/O by 100% to support copy-on-write related write-amplification
- A write buffer decreases snapshot I/O overheads but has little impact on storage overheads
- 40% of written bytes are transferred to protection storage
- Schedule at least 6 hours between transfers to minimize clean data in transferred blocks
- Schedule at least 12 hours between transfers to minimize peak network bandwidth requirements
- Protection storage must support 20% of the I/O per second capabilities of primary storage

EMC²

Questions?

EMC^2®