Network-based Storage Architecture for Exa-scale Computing Systems

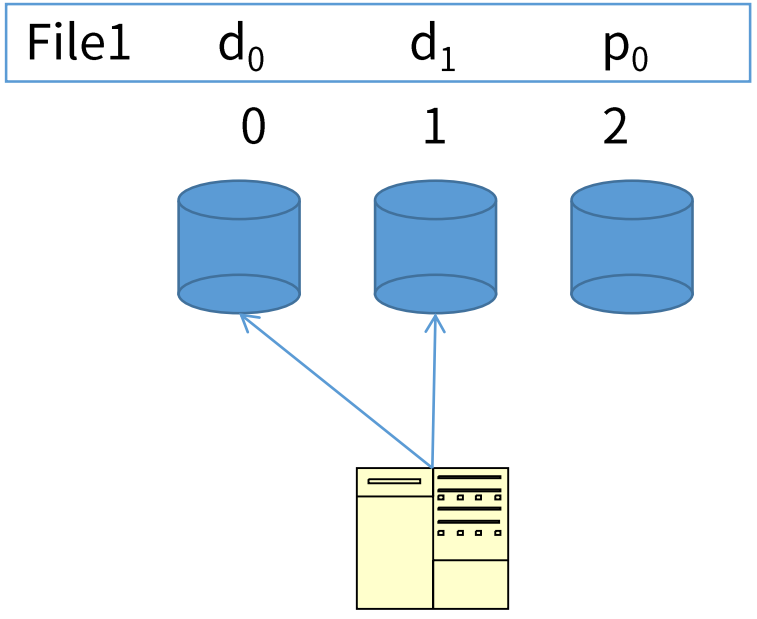
Hiroki Ohtsuji and Osamu Tatebe
University of Tsukuba, Japan

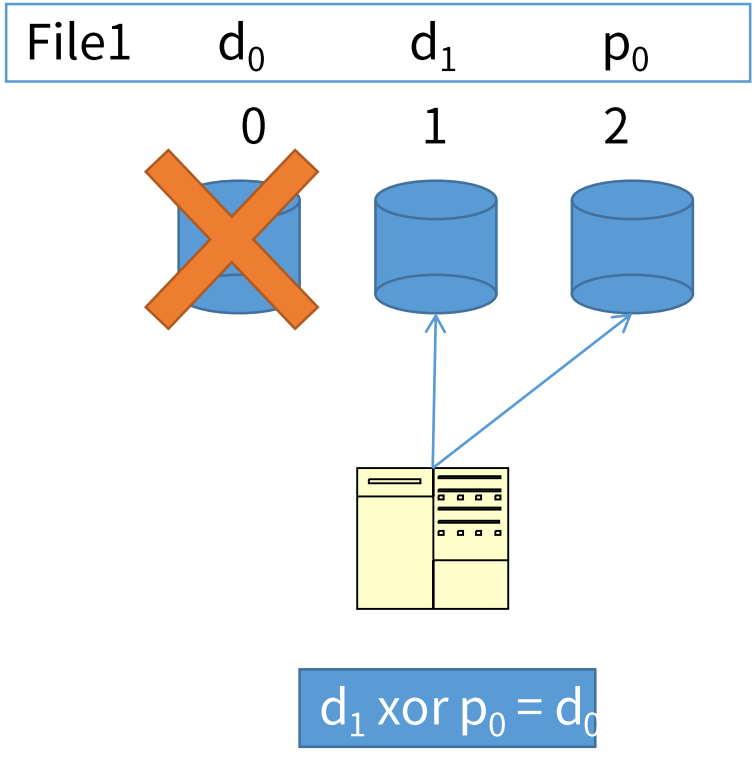
Abstract

- Network-based storage architecture for Exa-scale computing
  - Erasure coding
    - Special efficiency (cost)
    - Reliability
  - High-speed storage device / network
    - Performance
      - Latency
      - Bandwidth
- Implementation of efficient RAID-like network storage system

RAID and Cluster-wide RAID

- Ability to reconstruct the original data from one striped block and parity

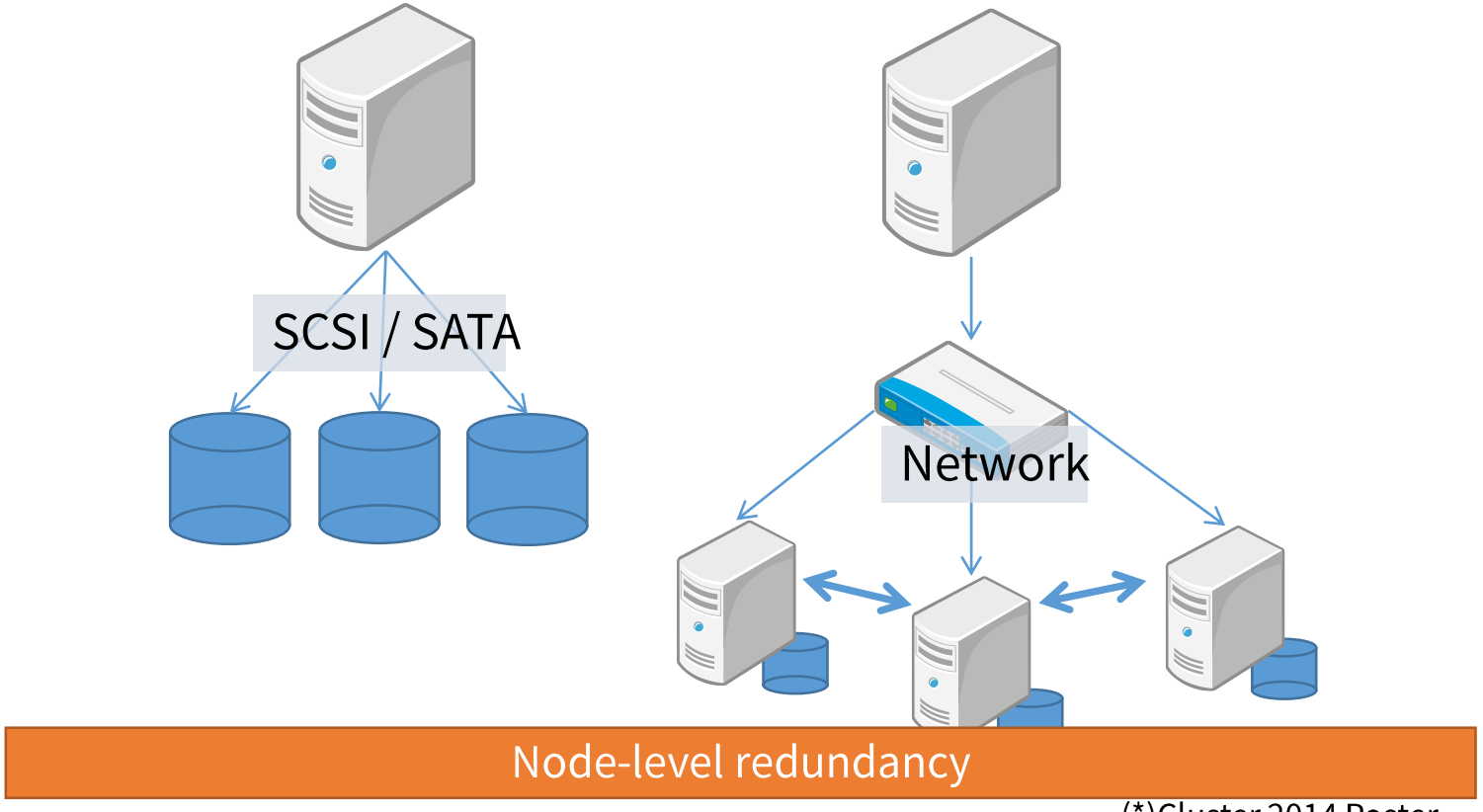




(*)Cluster 2014 Poster

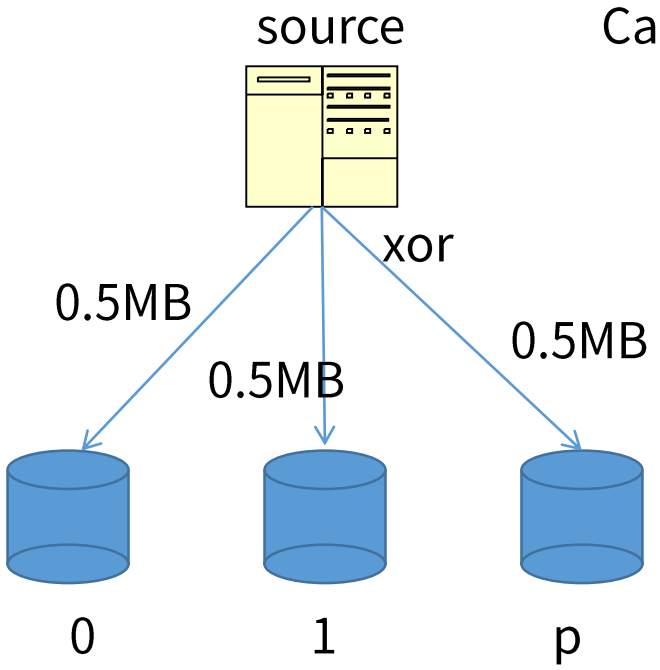
Cluster-wide RAID

- Disks are replaced with storage nodes



(*)Cluster 2014 Poster

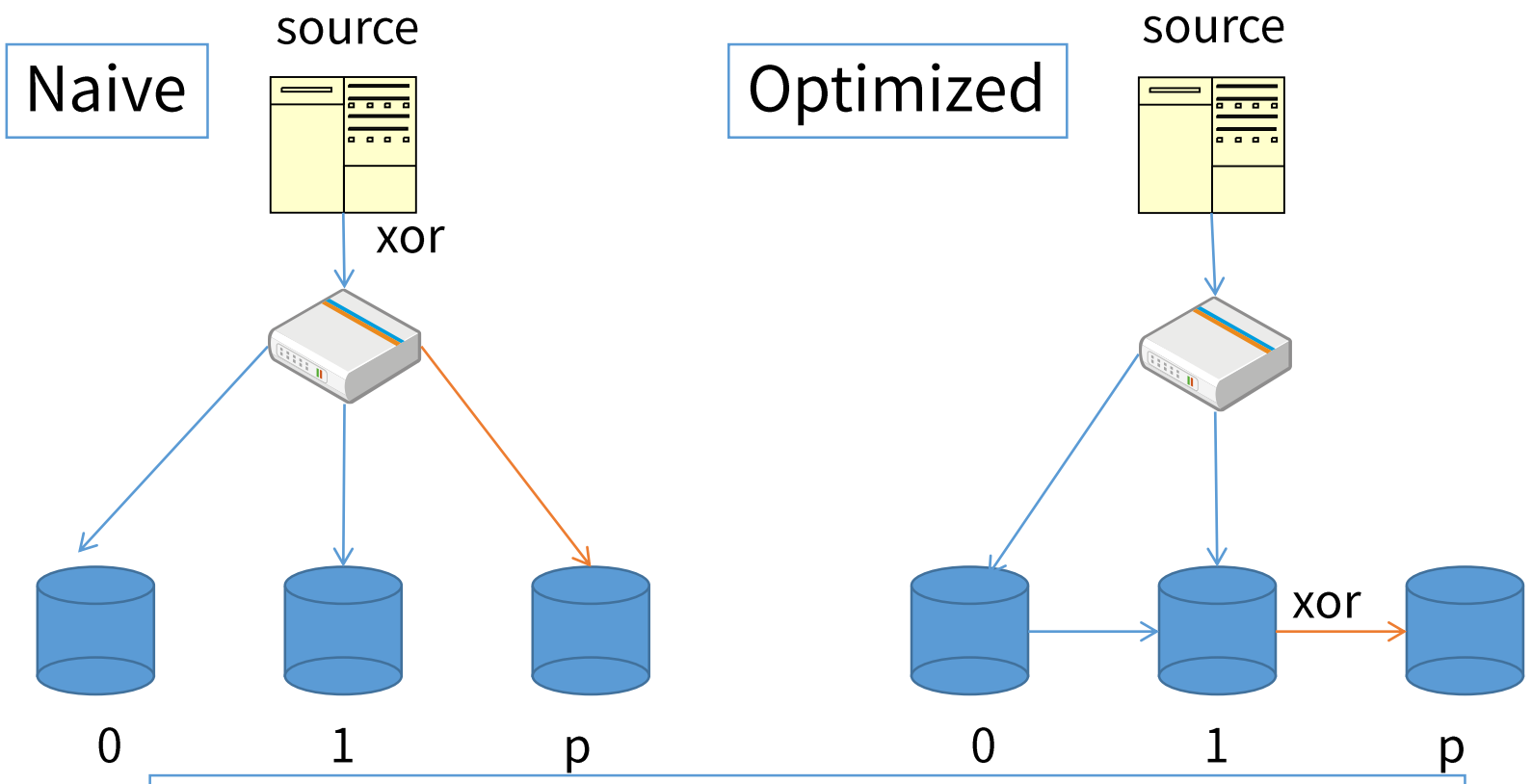
Increased traffic of Cluster-wide RAID-4 (write phase)



Traffic: Increased by 50% (compared to original data)

(*)Cluster 2014 Poster

Our proposal:



Traffic (in/out) of each storage nodes when the source node writes a 1MB file

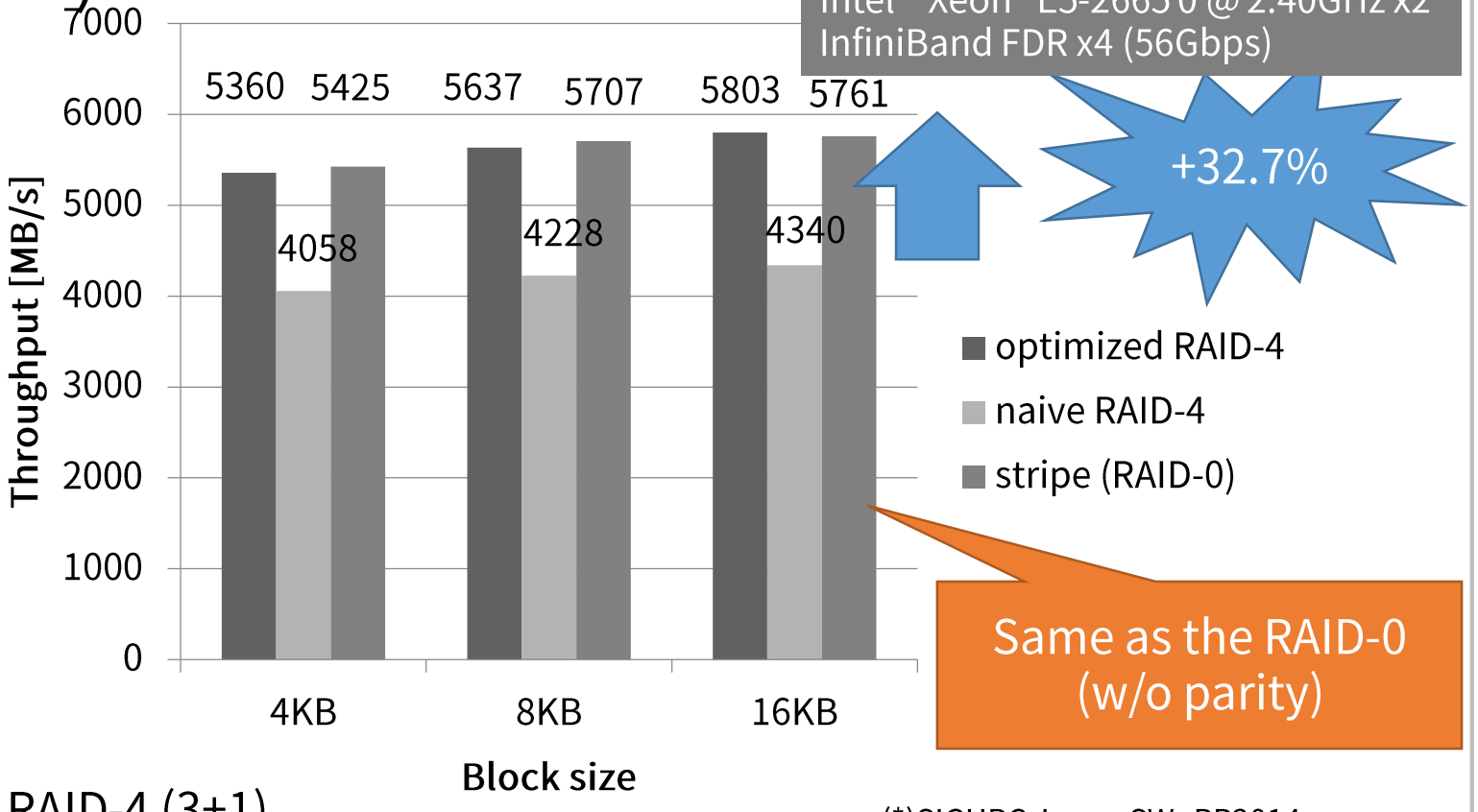
	client	0	1	P
in	0	0.5MB	0.5MB	0.5MB
out	1.5MB	0	0	0

	client	0	1	P
in	0	0.5MB	1MB	0.5MB
out	1MB	0.5MB	0.5MB	0

(*)SIGHPC Japan SWoPP2014

Preliminary Performance Evaluation

- Comparison: naïve, optimized and stripe (RAID-0)



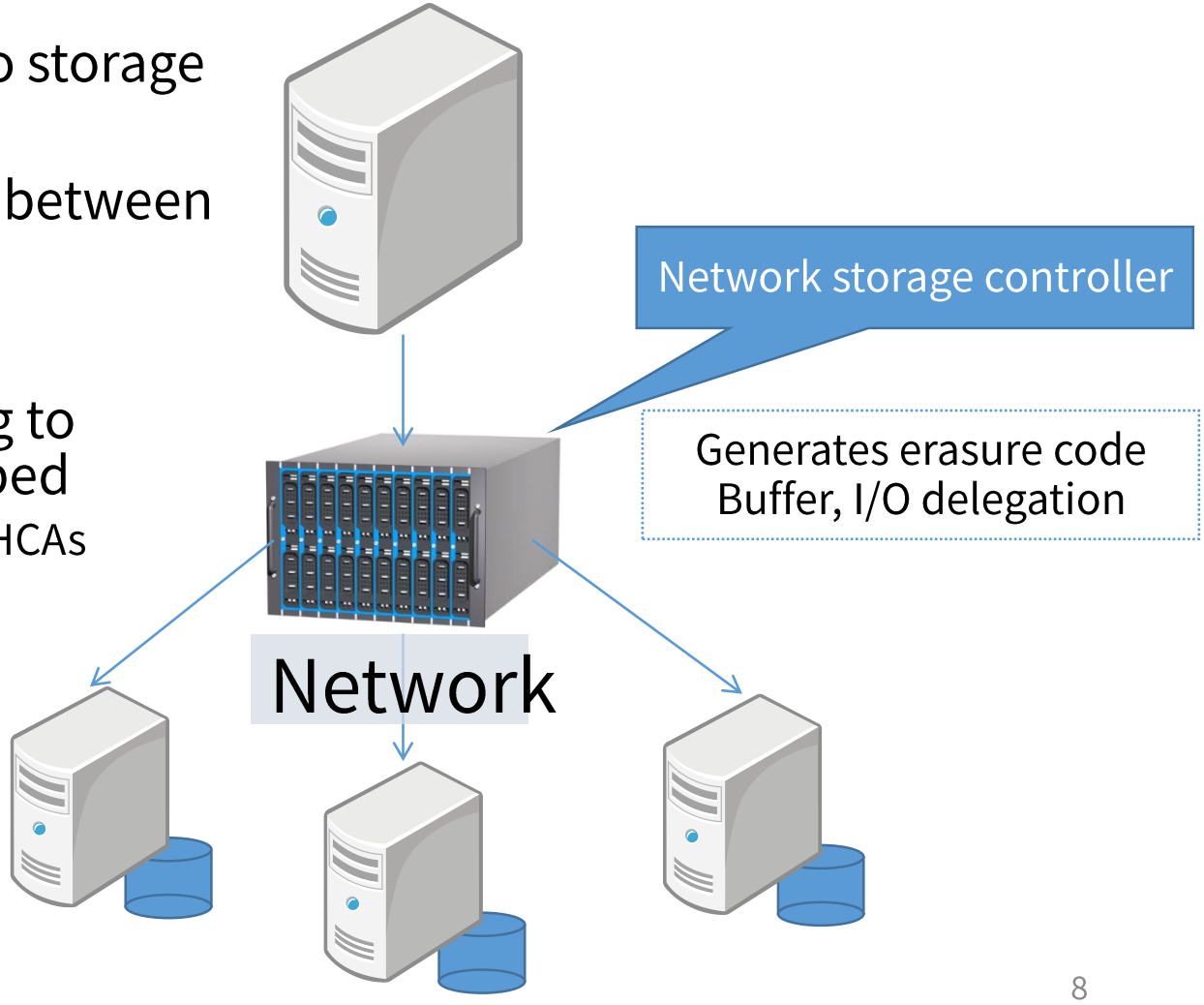
Cluster-wide RAID-4 (3+1)

(*)SIGHPC Japan SWoPP2014

The test implementation does not write to the actual disks due to lack of disk's performance.

New: Network storage controller

- Add a controller to storage system
- Reduce the traffic between storage nodes
- I/O optimization
- Currently, working to implement a testbed
  - Many InfiniBand HCAs



Conclusion

- Introduced the optimized implementation of Cluster-wide RAID-4
  - Performance gain was **31.5%** compared to naïve method
  - The performance of an optimized Cluster-wide RAID-4 is the same as RAID-0 (without parity data)
    - **Zero-overhead**
- Showed a plan for different implementation
  - Network storage controller

Acknowledgement

This works is supported by

- JST CREST “System Software for Post Petascale Data Intensive Science”,
- JST CREST “Extreme Big Data (EBD) Next Generation Big Data Infrastructure Technologies Towards Yottabyte/Year”
- JSPS KAKENHI Grant-in-Aid for JSPS Fellows