University of Connecticut

Parity Redundancy Strategies in a Large Scale Distributed Storage System

John A. Chandy

john.chandy@uconn.edu

NASA/IEEE MSST 2004
12th NASA Goddard/21st IEEE Conference on
Mass Storage Systems & Technologies
The Inn and Conference Center
University of Maryland University College
Adelphi MD USA
April 13-16, 2004

IEEE COMPUTER SOCIETY

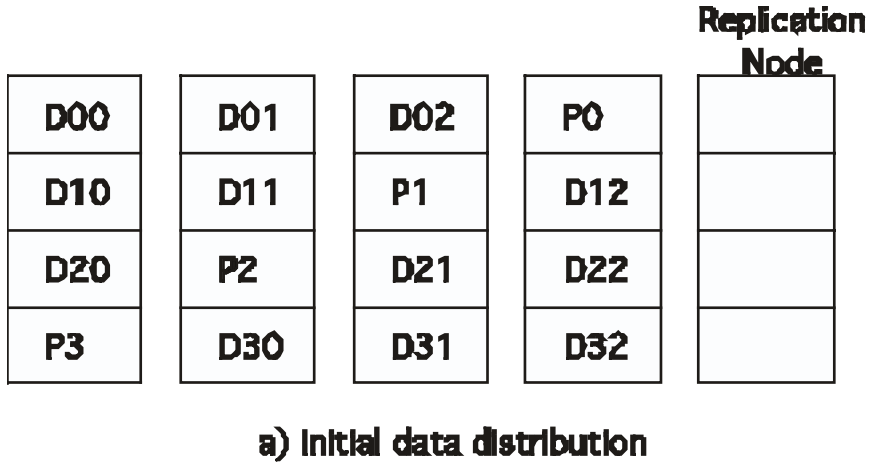
Parity Redundancy Strategies in a Large Scale Distributed Storage System

- Large scale distributed storage typically uses mirroring for redundancy
  - Easier to manage than RAID-5 parity style redundancy across a large number of nodes
  - Much better reliability than RAID-5
  - High cost in terms of redundancy overhead
- Use delayed parity instead
  - Mean time to data loss better than mirroring
  - Redundancy overhead is comparable to RAID5

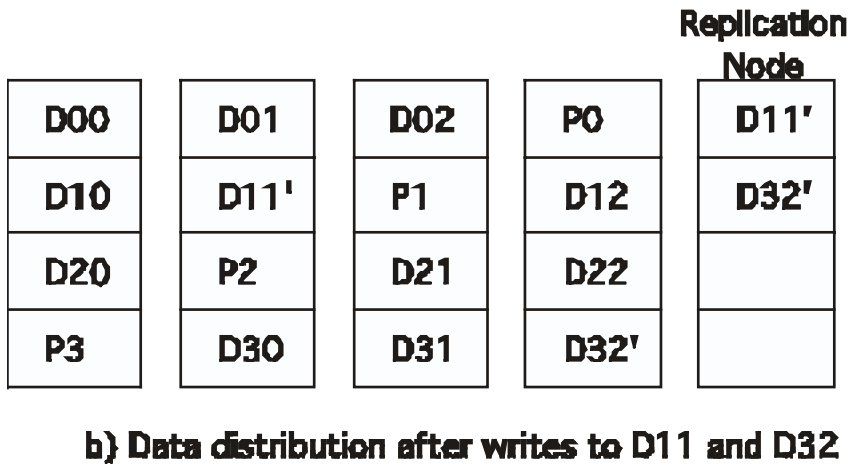
Delayed Parity Generation with Active Data Replication

- Mirror new data to a replication node
- Parity will be generated at a later time
- With the use of backups, can tolerate many double faults
- Active data replication node can be used to implement snapshots

DPGADR Data Distribution

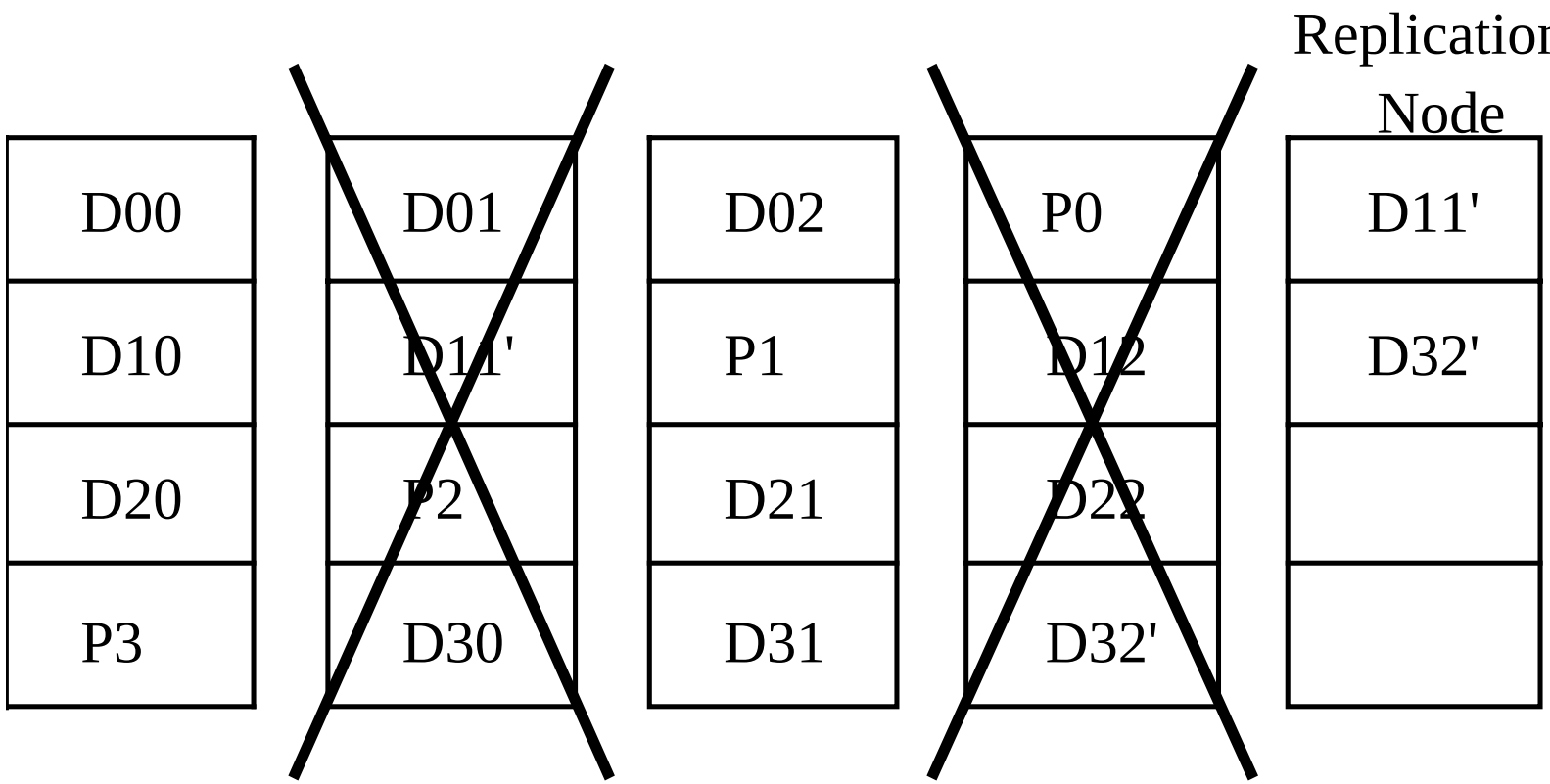


a) Initial data distribution



b) Data distribution after writes to D11 and D32

DPGADR with two failures



DPGADR comparison

- 1000 data nodes, MTTF=100,000 hours, MTTR=24 hours

Configuration	MTTDL	Overhead
RAID5 (d=5)	7.9 years	200 nodes
Mirroring	23.8 years	1000 nodes
DPGADR (n_G=4)	39.6 years	250 nodes