COPAN SYSTEMS

MAID for Active Archive Data

Aloke Guha

CTO, COPAN Systems

aloke.guha@copansys.com

Panel on Emerging Technologies

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

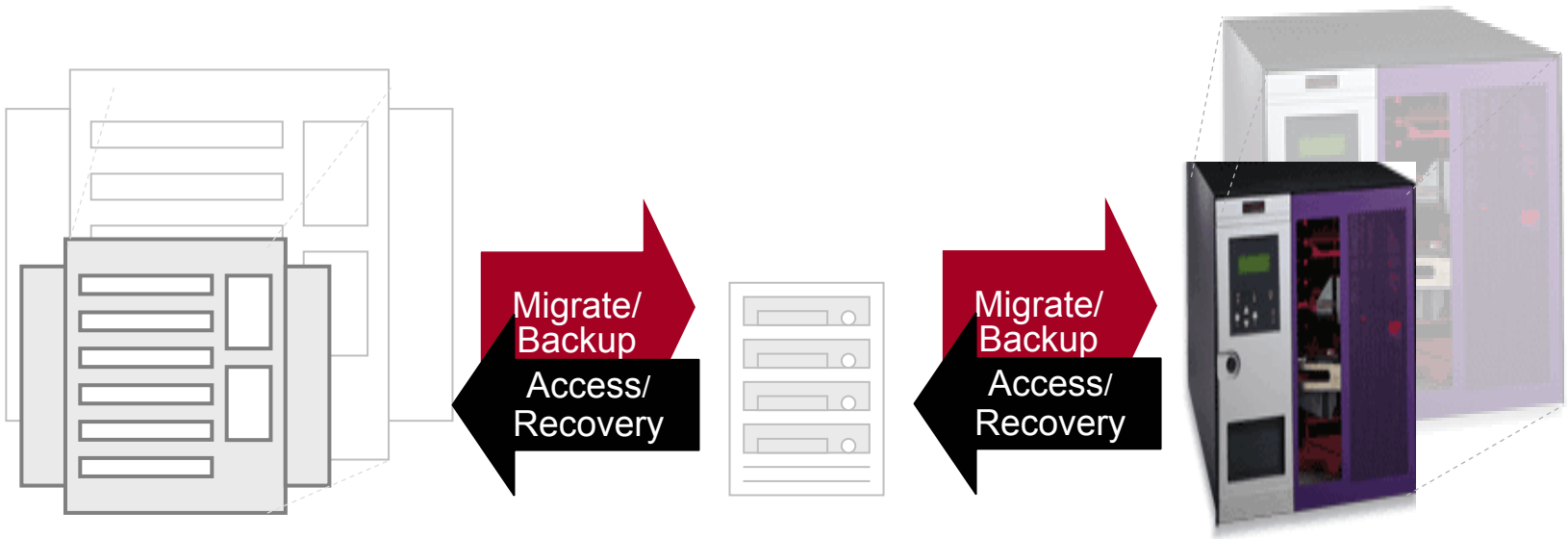
Adelphi MD USA

April 13-16, 2004

NASA

IEEE COMPUTER SOCIETY

Today's Hierarchy: Go Slow or Pay Up



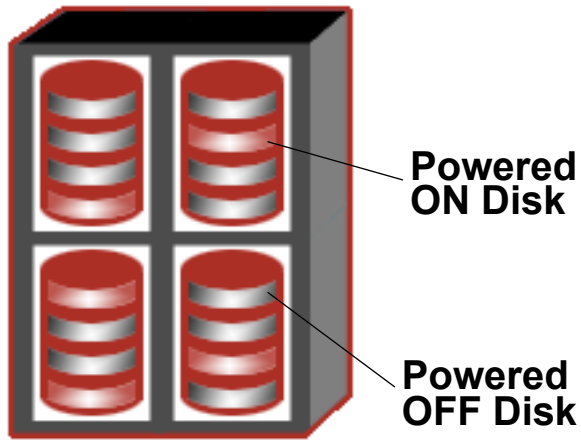
Primary Storage

Secondary Storage
Disk Cache

Tape Library
Bulk storage

MAID . . . . Power-Managed Disk

- Large number of power-managed drives
  - More than 50% drives powered OFF
  - Power-cycling by policy for application
- Benefits
  - Scale
  - Cost
  - Service life
  - Energy
- Ref: Colarelli and Grunwald, FAST 2002, SC 2002
  - Tradeoff of disk power vs. performance
  - Virtualization of ON (cache) and OFF drives, RAID-0
  - Caching not beneficial for archive workload

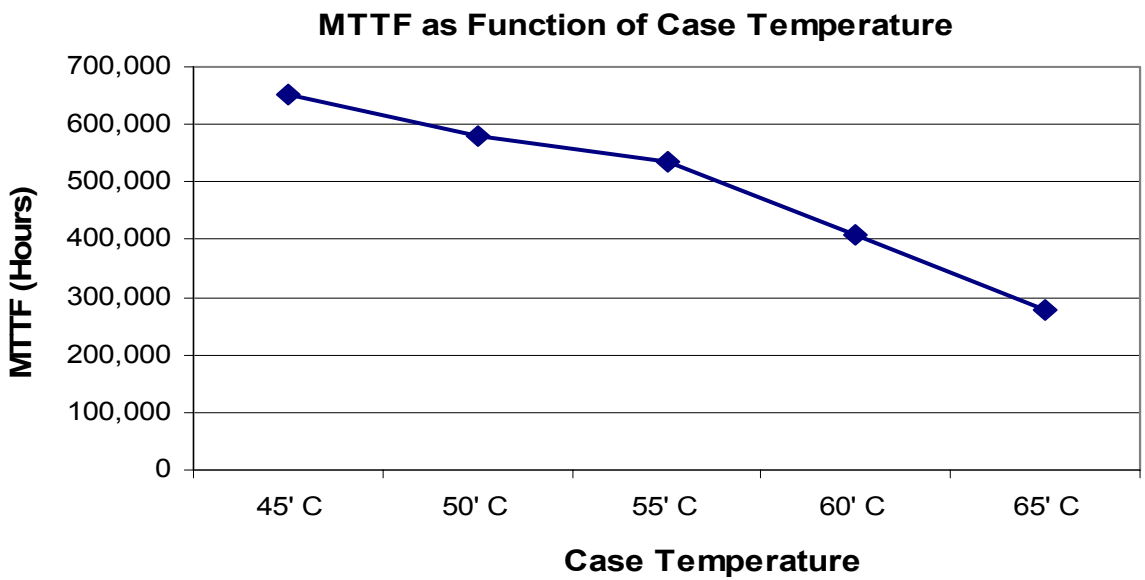


MAID Applicability

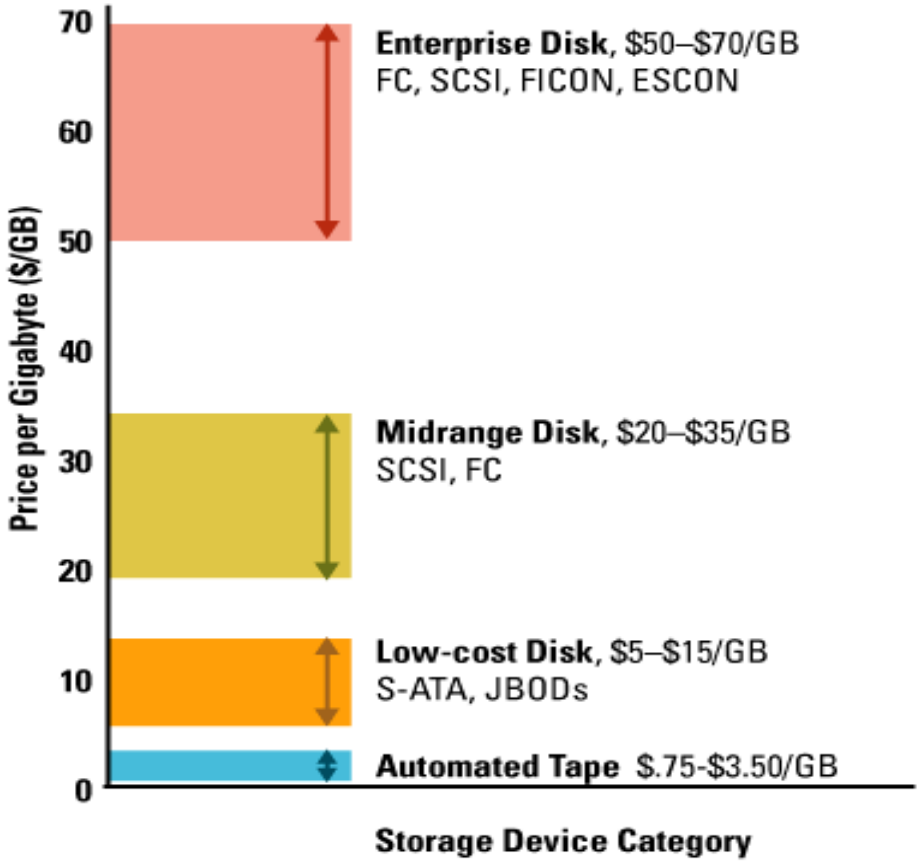
Content Type	Write	Update	Read	Technology	Metric
Dynamic	Many	Yes	Many	Disk	IOPs
Active Archive	Once	No	0 to n	MAID	Bandwidth
Deep Archive	Once	No	Rarely	Tape	Cost

Scale: Storage Capacity

- **Benefit**
  - –Large Number of Drives/Single System Footprint
  - – O(1000) drives ⇒ 250TB – 400TB
- **Needs**
  - – High-Density Interconnect Architecture
  - – Manage environmental conditions



Disk and Tape Pricing Guidelines



> The price per gigabyte decreases as the ratio of cartridges to drives increases, diverging from disk costs

> Disk prices are for working subsystems

> Automated tape prices include drives, media and library

> Tape cartridge capacity growing faster than disk drive capacity

> Automated tape nominally about 1/15 to 1/20 the price of disk for Unix, Linux, Win2K; 1/25 or less for mainframes.

Source: Horison Information Strategies

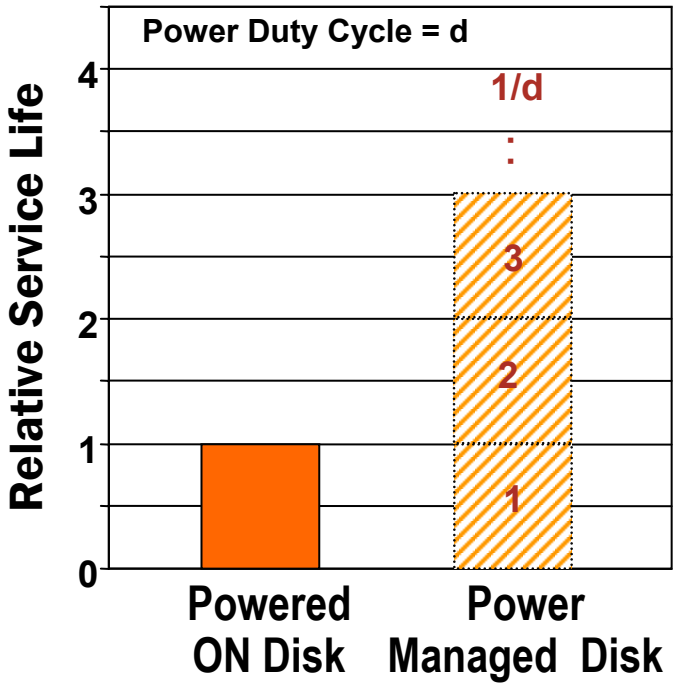
Understanding Storage Costs

- Non-media storage cost far exceeds media $/GB
- Cost Efficiency: $\frac{\text{Media Cost}}{\text{Storage System Cost}}$
  - Disk Example
    - 250GB SATA ~$1/GB vs Storage: $5-$15/GB
    - Cost Efficiency: 0.07 – 0.2
  - Tape Example
    - 200GB LTO2 ≤ $0.5/GB vs Storage: $0.75-$3.5/GB*
    - Cost Efficiency: 0.14 – 0.67
- Traditional disk systems 3x or more cost of tape*
- MAID levels playing field between Disk and Tape!

*Native uncompressed capacity: cost/GB depends on ratio of cartridges to drives, typ. 20:1 – 80:1
Assume same compression applied to data on disk or tape

Reliability: Drive Service Life

- Effective drive service life
  - MTBF ∝ 1/Annual Failure Rate
  - AFR ∝ Power On Hours
- Service life ∝ 1/(Power Duty Cycle)
- Data Rel. ∝ 1/(Power Duty Cycle)^2
- Needs
  - Manage start stops ≤ 50K
  - Data protection and integrity



Power duty cycle = # of powered-ON drives/# of powered-OFF drives

Managing Start Stops

- Bandwidth/capacity limits SS to 3% of max[1]
- Archive #mounts/volume limits SS to <5 % of max[1, 2]

	Industry	Volumes Used	Daily #Mounts/Volumes Used		
			Average	Max	Median
1	Telco	373	0.0	0.1	0.0
2	Telco	1,015	0.1	0.4	0.0
3	Telco	688	0.1	0.5	0.0
4	Telco	1,189	0.0	0.0	0.0
.	.	.	.	.	.
33	Utility	278	2.6	4.9	3.5
34	Govt	3,393	0.4	0.7	0.5
35	Govt	84	0.1	0.4	0.1
	Average	1,122	0.6	1.1	0.6

Ave # Start-Stops over 5 yr. ops

	Capacity (TB)		
Bandwidth (TB/hr)	150	200	250
2	584	438	350
3	876	657	526
4	1168	876	701

Typical Specified Limit: 50K

[1]Over 5-year period
[2]Source: SW Vendor - data from 43 archives on tape: Volumes_Used excludes tape volumes not allocated in ATL

Performance Implications

- Bandwidth increases as power duty cycle
- Access time depends drive state, power duty cycle

- Needs
  - Limit duty cycle, else increase device failure rates
  - Optimize overall architecture for performance wrt cost

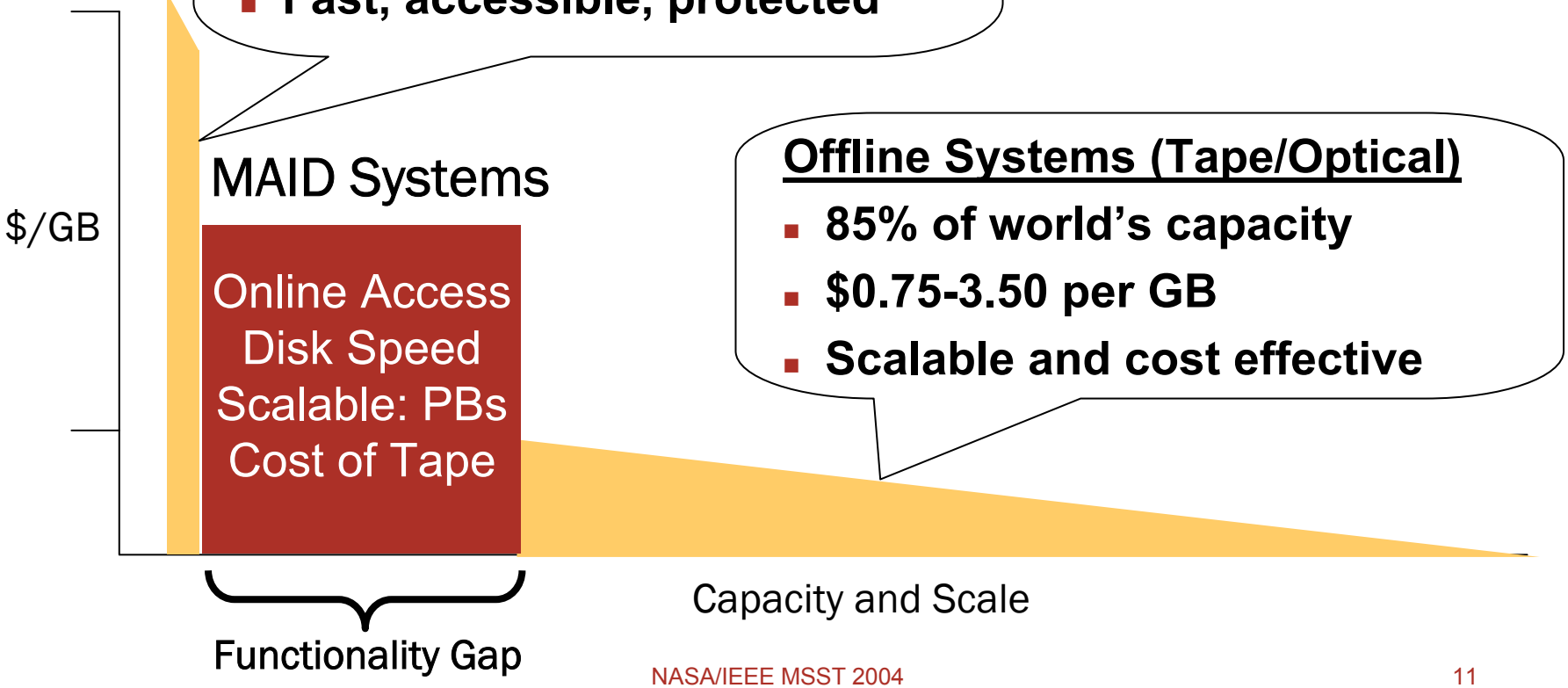
Device Type	Load (secs)	First Byte (secs)	File Access (secs)	Unload (secs)	Total/File* (secs)
Single Drive	<10	0.1s	10-12	0.1	12
RAID(n)	<10*n	4s	14-15	0.1	≥15
Tape Drive	18	41	59	18	77

* Transfer time depends on size of RAID set

Optimized MAID Fills the Gap in Hierarchy

Primary disk systems

- 15% of world's capacity
- $10-100 per GB
- Fast, accessible, protected



Offline Systems (Tape/Optical)

- 85% of world's capacity
- $0.75-3.50 per GB
- Scalable and cost effective