Dealing with Disks: RAID and Failures

15-712

Today's Stars

- A Case for Redundant Arrays of Inexpensive Disks (RAID)
  - Patterson, Gibson, Katz, SIGMOD '88
- Disk Failures in the Real World: What does an MTBF of 1,000,000 hours mean to you?
  - Schroeder & Gibson, FAST 2007
- A lot changes in 19 years. A lot doesn't...

Birth of RAID

- CPUs are going along nicely
- Patterson worried: Amdahl's Law says CPU cycles wasted if disk doesn't keep up

$$S = \frac{1}{(1-f) + f/k}$$

S = Speedup
f = frac work faster
k = how much faster

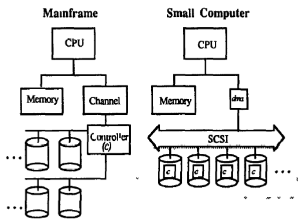
- e.g. if you make 30% of the system run 9x faster:
  - S = 1/((1-0.3) + 0.3/9) = speedup of 1.36x (bad...)

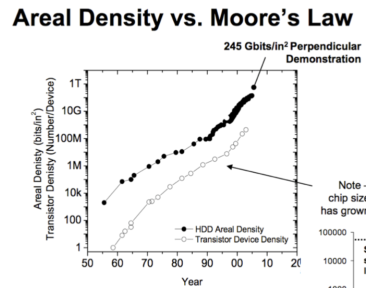
General problem!

- Balancing performance of components in computer system == eternal challenge
  - CPU speed
  - Memory cache speed (L1, L2, L3, ...)
  - Bus speed
  - Disk throughput
  - Disk IO operations / sec
  - Network throughput
  - Network latency
- Trying to substitute one for another == great fun, popular
  - Transistors for memory speed: prediction...
  - Spend local disk instead of network bw: Caching
  - Spend network bw instead of local disk: RDMA

Context

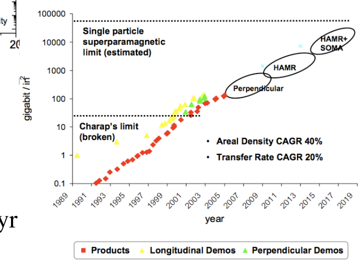
- IBM 3380 disk
  - 7.5GB (!!)
  - 4 arms
  - Size of a washing machine
- New PC industry…
  - Demands cost-effective drives
  - 3.5" form factor
  - Embedded SCSI controllers





Technology Trends

- Disk density tracks Moore's Law & no stalls in sight
- Transfer rate ~SQRT
- Random accesses ~5%/yr



Disk Specs

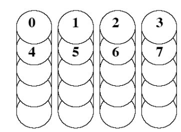
3.5 Inch Nearline	2005 (Longitudinal)	2009 (Perpendicular)	2013 (HAMR)
Drive Capacity (GB)	500	2,000	8,000
Number of Discs	3	3	3
Capacity (GB/disc)	168	670	2,670
Product Areal Density (Gbpsi)	120	500	1,800
Transfer Rate (Mb/sec)	995	2,000	5,000
RPM	7,200	7,200	10,000
Read Seek Time (ms)	8	7.2	6.5
2.5 Inch Enterprise	**2005**	**2009**	**2013**
Drive Capacity (GB)	75	300	1,000
Number of Discs	2	2	2
Capacity (GB/disc)	40	150	500
Product Areal Density (Gbpsi)	70	300	1,000
Transfer Rate (Mb/sec)	750	2,000	4,000
RPM	10,000	15,000	15,000
Read Seek Time (ms)	4.7	3.8	3.1
1.0 Inch Handheld	**2005**	**2009**	**2013**
Drive Capacity (GB)	8	30	100
[illegible]	1	1	1
[illegible]	8	30	100
[illegible] (Gbpsi)	80	300	1,200
[illegible]	110	200	600
[illegible]	3,600	4,200	5,400
[illegible]	16	14.5	13.1

Characteristics	*IBM 3380*	*Fujitsu M2361A*	*Conners CP3100*	*3380 v 3100 (>1 means 3100 is better)*	*2361 v 3100*
Disk diameter (inches)	14	10 5	3 5	4	3
Formatted Data Capacity (MB)	7500	600	100	01	2
Price/MB(controller incl)	$18-$10	$20-$17	$10-$7	1-2 5	1 7-3
MTTF Rated (hours)	30,000	20,000	30,000	1	1 5
MTTF in practice (hours)	100,000	?	?	?	?
No Actuators	4	1	1	2	1
Maximum I/O's/second/Actuator	50	40	30	6	8
Typical I/O's/second/Actuator	30	24	20	7	8
Maximum I/O's/second/box	200	40	30	2	8
Typical I/O's/second/box	120	24	20	2	8
Transfer Rate (MB/sec)	3	2 5	1	3	4
Power/box (W)	6,600	640	10	660	64
Volume (cu ft)	24	3 4	03	800	110

Table I *Comparison of IBM 3380 disk model AK4 for mainframe computers, the Fujitsu M2361A "Super Eagle" disk for minicomputers, and the Conners Peripherals CP 3100 disk for personal computers By*

- Diameter decrease driven by consumer formfactor
- Capacity up 1000X
- Transfer rate up 100X
- MTTF up 30X
- IO/sec up 5X

Striping: Read Throughput

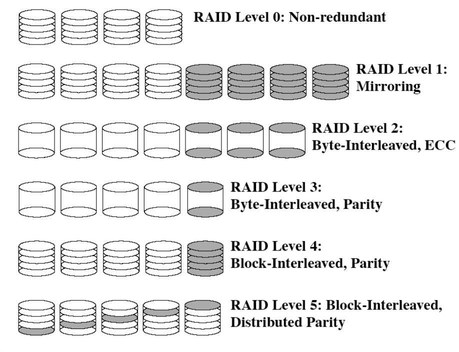


- Goals:
  - Load balance high-concurrency, small accesses across disks
  - Enable parallel transfers for low-concurrency large reads
- Striping to the rescue
  - Uniform load for small reads
    - If striping unit contains the whole object (e.g., small read is contained on one disk)
  - Parallelism for large reads
    - Stripe unit small enough to spread read across many disks

Wither RAID?

- Performance
  - IO ops/sec
  - Throughput
- Cost
  - "Inexpensive" disks (compare IBM to Conner...)
  - Somewhat less true today. 10x? range of prices
- Reliability

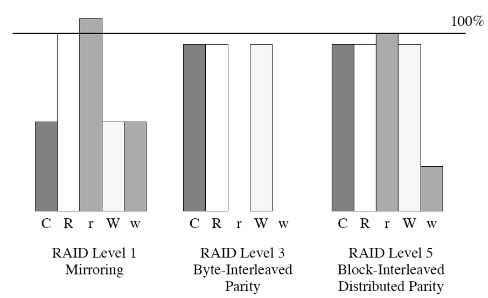
Synopsis of RAID levels



- RAID 0: Screamingly fast.
  - RAID 0 of 1000 drives: Screamingly *dead*...
- RAID 1: Mirroring
  - Really fast reads, *if* controller support
- RAID 2/3: Byte-interleaved (seems like bad idea)
  - Forced to access all disks for single read, even small
- RAID 4: Single parity disk
- RAID 5: Parity disk rotates
  - Difference not too huge in practice. Some major vendors use 4,

Tradeoffs of RAID levels

- Relative to non-redundant, 16-disk array



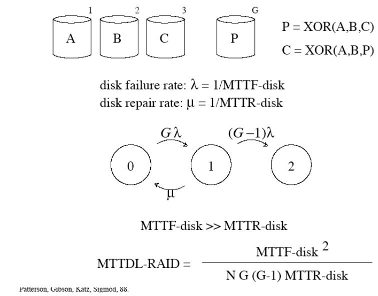
From "The Case for RAID"

RAID for Reliability

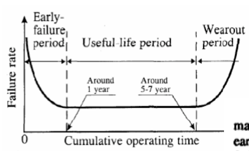
- Patterson88 focus: Performance
- Modern use: High availability + performance
  - More hard disk assemblies (HDA)s == more failures
- Note emphasis on "known" failure recovery
  - Disks have non-zero chances of undetected read/write errors -- data corruption.
  - *Mostly* get around with sector checksums/etc.
  - But requires careful integration of RAID+disk

Disk Reliability Model

- Exponential lifetime and repair



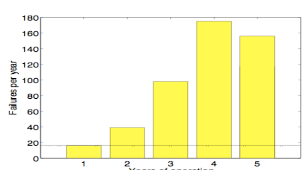
- G=#disks (notation change), N=# groups
- ex: G=15, N=5, MTTF=1M, MTTR=1h = 186x10^9 (!)



Naive modeling

As mentioned above we make the same assumptions that disk manufacturers make--that failures are exponential and independent (An earthquake or power surge is a situation where an array of disks might not fail independently) Since these reliability predictions will be very high, we want to emphasize that the reliability is only of the the disk-head assemblies with this failure model, and not the whole software and electronic system In addition, in our view the pace of technology means extremely high MTTF are "overkill"--for, independent of expected lifetime, users will replace obsolete disks After all, how many people are still using 20 year old disks?

- If failure rates are constant at 1/MTTF then Prob(fail in [t,t+dt]|live at t) ~= dt/MTTF
- Expected loss per month linear in failure rate
- But failure rates may not be constant :-(



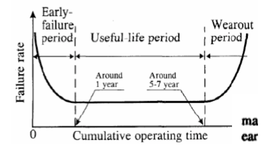
How reliable are they?

- Disks not as reliable as specs say
  - 3+% annual return rate
    - "Return" about as good as failure - if you yank the disk, you have to recover...
  - MTTF doesn't capture reality

Sources of error

- Real failure rates != spec sheets
  - Specs derived from "accelerated testing" & running many disks for shorter periods of time
  - Failures may depend on environment
    - Heat, bad power, small metallic particles in air from construction, excess vibration from AC, ...
- Correlated failures
  - RAID reliability assumptions assumed independent
  - If P(2nd failure | first failure) >> P(first failure), your RAID has a bit of a problem
  - P(2nd failure | first failure) _is_ > P(first failure)

1 step towards reality: Traditional "bathtub model"

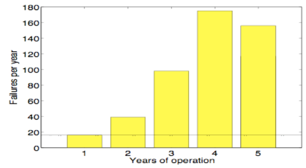


- Infant mortality: undetected faults in assembly, manuf., etc.
- Useful life - things work pretty well
- Wearout: mechanical components begin to break down
  - 30x more likely to fail when 5 years old (paper)

Missing Tub

- Schroeder07 didn't see start of bathtub curve
  - Good testing? Errors eliminated during burn-in?

- And "useful life" period wasn't static
  - Increasing failure P over life of disk
  - Big increase after 5 years...



Correlated Failures

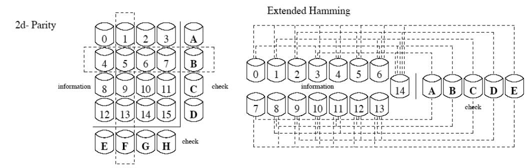
- Environmental correlation, same age, etc
- Hurst parameter 0.6-0.8
  - autocorrelation drops off slowly
- Expected time to next replacement: 4 days. (!!)
  -

Beyond RAID0-5

- Failure rates + sizes such that the probability of a failure during reconstruction is non-trivial
  - Contrast w/back of the envelope #s from Patterson88...
  - MAD grew faster than disk xfer rate
    - *Longer recovery times - several hours*
  - # of drives in an array got huge: 1000+ drive arrays
- RAID5 isn't particularly speedy
- Also note:
  - Disks 2.5x more likely to fail than CPU; 2x more likely than motherboard
  - So just making disks super-reliable isn't enough (Amdahl)
  - Few lectures from now: Cluster storage

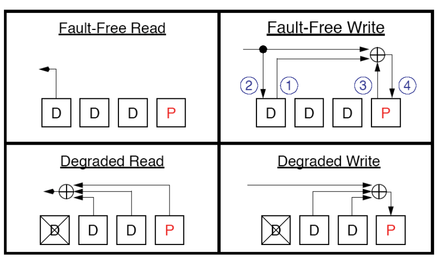
Double Correcting

- Borrow from earlier approaches
  - Orthogonal parity groups (lec focus: 2D parity)
  - Double error-detecting codes from mem. systems



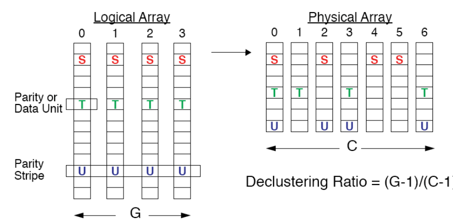
  - Overhead: check space vs. check update time
    - 2d parity: small time overhead (3), space = sqrt()
    - Hamming: lower space, higher avg time overhead

Performance During Recovery



- Per-disk load increase in degraded mode: 1 + r + 0.25w
- 50% throughput wall; long resp. time; long recovery

Reducing Load Increase: Parity Declustering

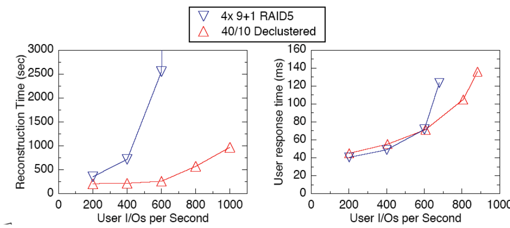


- **Per-disk failure-induced workload increase reduced**
- **Entire array bandwidth available for reconstruction**
- **Allows fault-free utilization > ~50%**
- **Map parity groups using** ***Balanced Incomplete Block Designs*** **or** ***Random Selection of Permutations***

Comparing to Multiple RAID Level 5 Groups

RAID5:	**4 groups of 9+1**	**⇒ 40 disks, 10% ovhd**
Declustered:	**1 group of C=40/G=10**	**⇒ 40 disks, 10% ovhd**

Performance during reconstruction



Eval

- RAID paper was survey/taxonomy, back of envelope
  - Patterson asked Gibson to learn about I/O & teach UCB architects about state of the art in high perf storage
  - Wrong on MTTDL value: P(DL this month) = K/MTTDL
  - Terms like "S" weak, but no experience to judge by
    - Most trends correct, but MIPS increase replaced by sum(cores) or clusters; MAD varied from 60%-100% per year, now about 40-50%. DRAM speed hasn't kept up -- big problem for architects. Still choices in disks (SATA vs FC - 2x IOs/sec)
- Key contributions:
  - Taxonomy by cost/perf - basic RAID 0,1, 5 still used in $15B market
  - A catchy name! (Patterson advice... RAID, ROC, NOW, ...)

Eval 2

- Schroeder paper
  - Brought more rigorous statistical methods to bear
  - 19 years later, had lots of data (comparatively)
    - likely that NetApp, EMC, Sun, etc. have more, but they don't talk about it
  - Challenges "data-sheet" numbers; seems to more closely match reality and experience