HOW RELIABLE IS A RAID?[1]

Martin Schulze[2], Garth Gibson, Randy Katz, David Patterson

Computer Science Division
Electrical Engineering and Computer Science Department
University of California, Berkeley
Berkeley, CA 94720

ABSTRACT: Disk arrays provide the promise of greatly increased transfer bandwidth at low cost. But without additional data redundancy, an array can suffer from significantly degraded reliability. In this paper we more closely examine the reliability of RAID systems and find that although phenomenal reliability disk arrays can not be attained with data redundancy alone, RAID system reliability can be made better than conventional large disks with little extra hardware.

1. Introduction

A critical challenge for future computer systems is to match the predicted advances in processor performance with comparable improvements in the I/O system. Arrays of disk drives hold forth the promise of such improvements, by dramatically improving transfer bandwidth through the parallelism inherent in many disk arms with data spread across multiple drives. Small format drives are particularly attractive candidates for building disk arrays, because of their low cost and their high volumetric (MB/ft^3) and power (MB/watt) efficiencies [VASU88].

The one significant drawback to using large numbers of disks to improve I/O performance is the impact it has on data reliability. More disks mean more disk failures and an increase in the probability of data loss. This observation is captured in the simple equation that relates the Mean Time To Failure (MTTF) of an array to the MTTFs of its component disks, assuming independent failures and a constant rate of failure:

$$MTTF_{disk\ array} = \frac{MTTF_{single\ disk}}{Number\ of\ Disks\ in\ the\ Array}$$

A disk array built from 49 small disks (ample capacity to replace conventional disks), each with an MTTF of 40,000 hours, would yield an overall MTTF of only 816 hours. This is significantly worse than the MTTF of a conventional mainframe disk with the same capacity. These conventional large disks have MTTFs of about 50,000 hours [BALA88].

Thus, arrays must be *redundant*, i.e., some capacity and bandwidth are sacrificed in order to redundantly store data, so that it is possible to reconstruct data lost in a failure. We have called such arrays **RAIDs**, for Redundant Arrays of Inexpensive Disks. This approach does not make array components more reliable, but, by making arrays tolerate component failures, it provides data availability comparable to conventional large disks. We view data availability as the reliability of the data. For simplicity, we call this the reliability of the array and assuming a constant failure rate, measure it with $MTTF_{RAID}$.

A RAID array is broken into a number of reliability *groups*, each containing extra "check" disks containing redundant data. This redundancy is intended to allow RAID systems to obtain a level of reliability at least equivalent to the conventional disk systems they are to replace. The following MTTF equation for a single error correcting RAID (i.e., an array that can tolerant a single failed disk in any reliability group, but not two failed disks) is reproduced from [PATT88]. It describes the reliability of a disk array (mean time to a failure that results in data loss) of n_G groups each with G data disks and 1 check disk when only failures in the disk devices are considered.

[1]Research Supported by National Science Foundation Grant # MIP-8715235 and the California MICRO Program with matching industrial support from SUN Microsystems, Inc.

[2]Current Address: Digital Equipment Corporation, Colorado Springs, CO.

$$MTTF_{RAID} = \frac{(MTTF_{disk})^2}{n_G\, G\, (G+1)\, MTTR_{disk}}$$

In this equation $MTTR_{disk}$ is the disk Mean Time To Repair and disk failure rates are assumed to be constant (ie., disk lifetime is exponentially distributed). Following our earlier example of 49 data disks, this equation predicts that an array of 7 groups each with 7 data disks that have a 40,000 hour $MTTF_{disk}$ and a 2 hour $MTTR_{disk}$ will have an overall $MTTF_{RAID}$ of 2,040,816 hours. Fast repair is facilitated by onsite spare (inexpensive) disks and operating system directed online replacement and reconstruction. Unfortunately, this phenomenal $MTTF_{RAID}$, 239 years, is a very optimistic estimate of reliability because failures of disk support hardware have been ignored.

The purpose of this paper is to develop a better understanding of RAID reliability than the simple MTTF calculation just introduced. The reliability of other components, such as power supplies, controller electronics, cables, and fans, should also be considered for they affect the overall reliability of the array. Our goal is not "non-stop" reliability, but rather a level of reliability for large numbers of disks that is comparable to that of a single conventional disk. However, if failure rates of disks and support hardware are constant, we shall see that seemingly excessively large MTTFs may be desirable to secure a low probability of data loss.

The rest of this paper is organized as follows. In the next section we briefly review basic reliability definitions and examine sources of non-catastrophic disk failures. We next examine the exponential failure model and conclude that users' perceptions of MTTFs are not really what MTTFs mean. We then develop a model of reliability for array systems that includes support hardware. This leads us to hardware reliability groups, which contain redundant hardware components to further improve the array reliability.

2. How Disks Fail

We adopt definitions from the fault-tolerant research community [MAXI88]. A *failure* is a detectable physical change to hardware. Failures may be repaired by the replacement of a physical component. A *fault* is an event which interferes with normal operation and can be either *soft (transient)*, i.e., not readily repeatable, or *hard*, i.e., repeatable with high probability. Hard faults may be caused by failures, while soft faults are more likely caused by environmental factors or insufficient design margins. An *error* is a manifestation of a fault by an incorrect value. Errors, therefore, can be either soft or hard.[3] In this paper we are mainly concerned with catastrophic failures; failures that render a device module inoperable (such as head crashes or read/write electronics failures), but we begin by exonerating non-catastrophic failures and faults.

Disk drive manufacturers have identified a few types of errors associated with the servo system (positioning the heads) and the read/write system as critical to customer satisfaction. Typical specifications for the occurrence rates of these types of errors are shown in Table I [CDC 88, QUAN87].

A recoverable seek error is a seek in which the drive does not locate the desired cylinder on the first try, but is successful during a retry (if it is never successful then a catastrophic failure has occurred). A data error is defined as one sector read incorrectly, as detected by an Error Correcting Code (ECC). Random recoverable data errors are soft errors usually related

Table I -- Disk Drive Error Rates

Type of Error	Average Error Rate	Recovery	Consequences
Recoverable Seek Error	<1 error in 10^6 seeks	retry	none
Transient Recoverable Data Error	<1 error in 10^{10} bits read	retry or ECC	none
Repeatable Recoverable Data Error	< 1 error in 10^{12} bits read	ECC	Data rewritten to relocated sector.
Unrecoverable Data Error	< 1 error in 10^{14} bits read	none	One sector's data lost. Sector marked bad.
Miscorrected Data Error	< 1 error in 10^{21} bits read	none	One sector's data incorrectly read.

[3] In some systems recoverable errors are called *soft* and unrecoverable errors are called *hard*. This differs from our usage mainly through repeatable errors that are recoverable by error correcting codes.

to the signal-to-noise ratio of the system. Repeatable recoverable errors are hard errors, most often due to media defects, that can be corrected by ECC. Unrecoverable data errors lose data because the sector is detectably too damaged to recover by ECC. Miscorrected data errors occur when ECC was incorrectly not invoked or has resulted in incorrect data.

To get a feeling for the magnitude of these error rates, consider a disk that is performing 50 seeks/sec and reading 512KB/sec sustained. The mean time to next failure for each of these error types is: for recoverable seek errors, 5.6 hours; for random recoverable data errors, 40 minutes; for repeatable recoverable data errors, 2.8 days; for unrecoverable data errors, 276 days; and for miscorrected data errors, 7.6 million years. The first three of these are recoverable without user intervention and the last is negligible when it is compared to each disk's 5 year mean time to catastrophic failure. Only unrecoverable data errors appear to pose a problem, but these can be dealt with in the same manner as a catastrophic drive failure – the sector can be reconstructed from redundant data in the array. We now turn our attention to just how low we should try to make a disk array's catastrophic failure MTTF.

3. Exponential Failure Model

Many disk drives offer catastrophic failure reliabilities specified as MTTF = 30,000 to 50,000 hours of normal usage. Disk users tend to interpret these specifications as implying that their disks will fail after this many hours of operation. Unfortunately, this is not the case. If disk lifetimes are truly exponential, i.e., failure rates are constant, with mean lifetime equal to the MTTF, then there is a large probability that the disk will fail before the MTTF is reached. Table II summarizes this observation. The actual formula is:

$$\text{Prob(exactly k failures in MTTF)} = \frac{1}{(e\ k!)}$$

where e is the base of the natural logarithm.

Table II -- Exponential Lifetimes compared to MTTF

Percent of All Disks	# of Failures Experienced Within MTTF	Cumulative Percent	# of Failures Experienced Within MTTF
36.8	0	100.0	0 or more
36.8	1	63.2	1 or more
18.4	2	26.4	2 or more
6.1	3	8.0	3 or more
1.5	4	1.9	4 or more
0.3	5	0.4	5 or more
0.05	6	0.06	6 or more
0.007	7	0.008	7 or more

This means that some customers will have one or more failures within a small fraction of the quoted MTTF. To avoid customer distress, disk manufacturers may underrepresent their MTTFs. Alternatively, in building a disk system, the goal should be to make the probability of data loss within a reasonable interval as low as possible. Thus we should strive for MTTFs substantially higher than the expected useful product lifetime. However, the overall reliability of users' computation also depends on main memory and CPU failure rates and even more so on software quality [GRAY85], so we should not be too extravagant with the design of an IO subsystem.

4. RAID Reliability Revisited

The redundancy of RAID is an application of a fault tolerant technique to address the problem of data loss due to disk drive failures. Small disk drives are not standalone units, but require support hardware: power supplies, SCSI (Small Computer System Interface) Host Bus Adapters (HBAs), cooling equipment, and cabling. To get a more accurate picture of RAID reliability, all parts of the array system should be considered.

The Berkeley RAID is based on the concept of *parity group*, i.e., a group of disks sharing a common parity check disk. When this RAID is implemented with a shared interconnect such as SCSI, a second type of grouping emerges that is based on the *SCSI group*, i.e., a group of disks sharing a common SCSI cable and HBA. There is also the *power group*, a group of disks sharing a common power supply, and the *cooling group*, a group of disks sharing a common fan. The interaction of these groups has a major influence on overall RAID reliability.

Figure 1 shows the reliability of various components of a SCSI based RAID. Note that the RAID MTTF equation presented in Section 1 considered only the reliability of disks. To build a realistic system assembly would require eight SCSI HBAs with cables, eight 300 Watt power supplies (each with a power cable), and eight fans for cooling. The overall

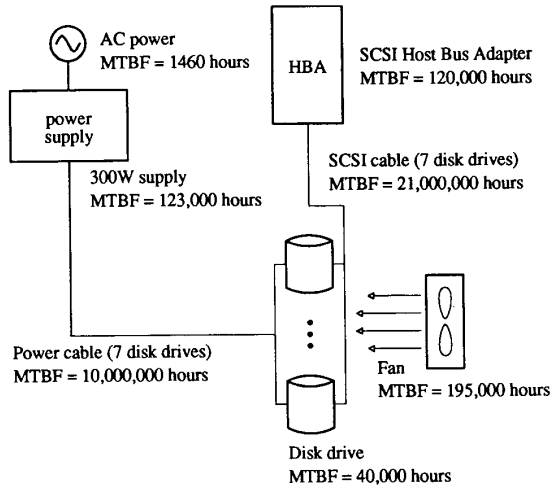

Figure 1: Reliability of RAID Components

A disk subsystem is more than just disks. This figure illustrates typical reliabilities for each part. The MTTFs for disk drives and Host Bus Adapters are estimates quoted from representative manufacturers [CDC88, Moren88]. MTTFs for fans are estimated from MIL-HDBK-217E part code 801 (electric motor, < 1 horsepower) with 4 solder connections. MTTFs for SCSI cables are estimated from MIL-HDBK-217E part code 1105 (printed wiring board connector) with 50 active pins and 50 milliamps per pin with 0.04 mate/unmate cycles per 1000 hours. Power supply MTTFs are from MIL-HDBK-217D [BARD86] and MTTFs for power cables are estimated from MIL-HDBK-217E part code 1103 (power connector) with 4 active pins and 2.5 amps per pin with 0.04 mate/unmate cycles per 1000 hours. The MTTF of the external power grid is taken from Gray [GRAY85].

RAID reliability should take account of these additional components and their independent rates of failure. Of greatest concern is the external power grid; without battery backup your system is at the mercy of the power company and can expect an overall MTTF no better than 2 months! Assuming that the power supply has battery backup but that any failure in the support components may cause data loss (a pessimistic assumption), $MTTF_{RAID}$ would be revised to 5734 hours. This is about 239 days and represents a factor of 356 decrease in MTTF from the simple estimate that considered only disk drive failures!

However, by judicious placement of parity, SCSI, power, and cooling "groups", we can do much better. If parity groups are mapped onto the disk array orthogonal to SCSI, power, and cooling groups, then no single hardware failure will cause data loss. We see this in Figure 2 because the loss of any complete column amounts to the loss of a single disk in each parity group and each of these is recoverable. This scheme does not have any explicit fault tolerance of the support hardware, but uses the redundancy of parity groups to protect against support hardware failures as well as disk failures. For this to work, support hardware failures should be repaired with little or no interruption of service. Since these failures involve a variety of types of equipment that may not be easily replaced, a separate, and probably longer, mean time to repair, $MTTR_{column}$, is used.

For the RAID of Figure 2, $MTTF_{RAID}$ can be estimated as

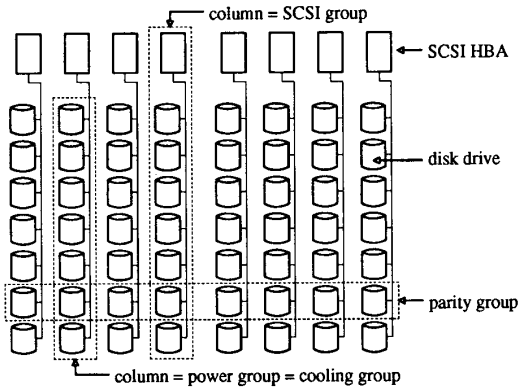

Figure 2: Groupings in a Disk Array

Data reliability groups are organized orthogonal to a SCSI string, while hardware reliability groups are organized around the string. This improves reliability by insuring that no single data or hardware failure will lead to data loss.

$$\frac{(MTTF_{disk})^2}{n_G\, G\, (G+1)\, MTTR_{disk}\left(1+\alpha_F \frac{1+\alpha_R}{\alpha_R}+\frac{\alpha_F{}^2}{n_G\, \alpha_R}\right)}$$

where $\alpha_F = \frac{MTTF_{disk}}{MTTF_{column}}$ and $\alpha_R = \frac{MTTR_{disk}}{MTTR_{column}}$.

Note that this formula reduces to the original as $MTTF_{column}$ goes towards infinity.

To illustrate this model, reconsider our earlier example. Assuming each column has one SCSI HBA, one 300 Watt power supply, one fan, one SCSI cable, and one power cable, then $MTTF_{column}$ is 46,000 hours (calculated by summing the failure rates of these support components). If the average time to repair support hardware failures in a column, $MTTR_{column}$, is 72 hours, then $MTTF_{RAID}$ is 55,000 hours. Although 55,000 hours is still a far cry from the simple estimate of 2,000,000 hours, it does meet our goal of exceeding the reliability of an individual conventional disk. In fact, conventional disk reliabilities also exclude interconnect and host bus adapter support hardware, so, using a SCSI interconnect and HBA, a conventional single disk subsystem built around 50,000 hour MTTF disk will have an overall MTTF of 35,235 hours (large conventional disks have built in power and cooling).

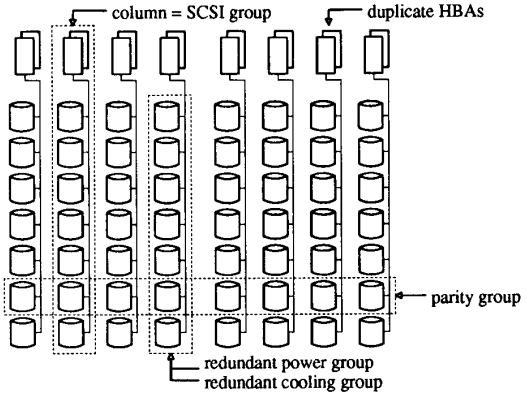

Figure 3: Redundant Groupings in a Disk Array

Adding redundant hardware to the column can increase its MTTF and thus the reliability of the whole array.

Just as data redundancy dramatically improves the reliability of disk data, additional hardware redundancy should improve the reliability of the support hardware. For example, power and cooling groups can be made redundant, or SCSI HBAs can be duplicated on each SCSI cable (see Figure 3). By adding redundant hardware components, $MTTF_{column}$ can be increased, thus further increasing $MTTF_{RAID}$. Table III shows the effects on RAID reliability of making various combinations of the support hardware redundant. We show both how far each combination is from the original $MTTF_{RAID}$ estimate, as a percent, and how this compares to a conventional single disk subsystem reliability of 35,235 hours, also as a percent. We should stress here that our model pessimistically assumes that fan failures render an entire column unavailable and two fan failures close together cause the loss of data.

5. Summary and Conclusions

RAID is a technique for increasing I/O bandwidth by spreading data across large numbers of small format disk drives organized as an array. Because a small number of conventional large format disks are replaced by much larger numbers of small disks, the reliability of the disk system is greatly reduced. To deal with the problem of substantially more frequent disk failures, RAIDs trade off some of the increased bandwidth and storage capacity for redundant data storage, so lost data can be reconstructed through parity calculations. An optimistic analysis, focusing on disk reliability alone, would indicate that RAID systems can be made very much more reliable than conventional systems at a very modest cost of extra "check" disks.

This paper has presented a more careful reliability analysis that clearly shows that the rest of the system components, such as the SCSI HBAs, power supplies, and fans cannot be ignored. We have presented a scheme in which the system support components are organized into groups orthogonal to the data redundancy groups, thus guaranteeing that no single disk OR component failure will permanently lose data. This approach yields disk arrays with reliability about 50% greater than conventional disk subsystems. If further reliability improvements are sought, various parts of the support hardware can be made redundant. We have shown how these affect the design of a 49 data disk RAID. From this example we see that redundancy in the relatively inexpensive fans and

Table III -- RAID Data Reliability Considering Support Hardware			
Configuration	$MTTF_{RAID}$	Percent of Maximum	Percent of Conventional
No Redundancy	816 hrs	0.04%	2.1%
Simple RAID	5,734 hrs	0.3%	16%
(Orthogonal) RAID	55,000 hrs	3%	156%
RAID + Redundant Fans	73,000 hrs	4%	208%
RAID + Redundant Power Supplies	90,000 hrs	4%	255%
RAID + Redundant HBAs	91,000 hrs	5%	259%
RAID + Redundant Power and Fans	144,000 hrs	7%	409%
RAID + Redundant HBAs and Fans	148,000 hrs	7%	418%
RAID + Redundant Power and HBAs	225,000 hrs	11%	639%
RAID + Redundant Power, HBAs, Fans	1,650,000 hrs	81%	4680%

power supply can yield overall MTTFs of about 4 times conventional disk subsystems.

6. References

[BALA88] Balanson, R., "GPD Products and Technology Directions," IBM Fellowship Conference Presentation, San Jose, CA, (November 1988).

[BARD86] Bardos, P., "The Reliability of Switch Mode Power Supplies," *Electronic Engineering*, V. 58, N 715, (July 1986), pp. 37-44.

[CDC 88] Control Data Corporation, "Product Specification for WREN IV SCSI Model 94171-344," Control Data OEM Product Sales, Minneapolis, MN, (January 1988).

[GRAY85] Gray, J., "Why Do Computers Stop and What Can Be Done About It?," Tandem Technical Report 85.7, (June 1985).

[MAXI88] Maxion, R.A., D.P. Siewiorek, "Symptom-directed diagnosis of distributed computing systems," *1986/1987 Research Review*, Computer Science Department, Carnegie Mellon, 1988.

[MIL 86] U. S. Department of Defense, *Military Handbook: Reliability Prediction of Electronic Equipment, MIL-HDBK-217E*, (October 1986).

[MORE88] Moren, W. D., Ciprico, Inc., private communication, (July 1988).

[PATT88] Patterson, D. A., G. Gibson, R. H. Katz, "A Case for Redundant Arrays of Inexpensive Disks (RAID)," ACM SIGMOD Conference, Chicago, IL, (June 1988).

[SCHU88] Schulze, M. E., "Considerations in the Design of a RAID Prototype," M.S. Report, U. C. Berkeley Computer Science Division, (August 1988).

[QUAN87] Quantum Corporation, "OEM/Programmers Manual for Q200 Series Disk Drives," Milpitas, CA, (May 1987).

[VASU88] Vasudeva, A., "A Case for Disk Array Storage Systems," Proc. Systems Design and Networks Conference, Santa Clara, CA, (April 1988).